Gene regulatory mechanisms guiding bifurcation of inhibitory and excitatory neuron lineages in the anterior brainstem

Sami Kilpinen*, Lassi Virtanen, Silvana Bodington-Celma, Amos Bonsdorff, Heidi Heliölä, Kaia Achim*, Juha Partanen*

Molecular and Integrative Biosciences Research Programme, Faculty of Biological and Environmental Sciences, University of Helsinki, Finland

* Corresponding author

Abstract

Selector transcription factors control choices of alternative cellular fates during development. The ventral rhombomere 1 of the embryonic brainstem produces neuronal precursors that can differentiate into either inhibitory GABAergic or excitatory glutamatergic neurons important for the control of behaviour. Transcription factors (TFs) *Tal1*, *Gata2* and *Gata3* are required for adopting the GABAergic neuronal identity and inhibiting the glutamatergic identity. Here, we asked how these selector TFs are activated and how they control the identity of the developing brainstem neurons. We addressed these questions by analysing chromatin accessibility at putative gene regulatory elements active during GABAergic and glutamatergic neuron lineage bifurcation, combined with studies of transcription factor expression and DNA-binding. Our results show that the *Tal1*, *Gata2* and *Gata3* genes are activated by highly similar mechanisms, with connections to regional patterning, neurogenic cell cycle exit and general course of neuronal differentiation. After activation, *Tal1*, *Gata2* and *Gata3* are linked by auto- and cross-regulation as well as regulatory interactions with transcription factors of the glutamatergic branch. Predicted targets of these selector TFs include genes expressed in GABAergic neurons, glutamatergic neurons, or both. Unlike genes specific to the glutamatergic branch, the genes expressed in GABAergic neurons appear to be under combinatorial control of *Tal1*, *Gata2* and *Gata3*. Understanding gene regulatory interactions affecting the anterior brainstem GABAergic and glutamatergic neuron differentiation may give genetic and mechanistic insights into neurodevelopmental traits and disorders.

Keywords: Neuron, GABAergic, glutamatergic, brainstem, pons, selector gene, transcription factor, Tal1, Gata, enhancer, transcriptome, chromatin accessibility, gene regulatory network

Introduction

The anterior brainstem is a complex neuronal mosaic controlling mood, motivation and movement. These neurons include diverse inhibitory GABAergic and excitatory glutamatergic neurons, some of which are associated with the dopaminergic and serotonergic systems.

An important source of these neurons is the ventrolateral rhombomere 1 (r1), where GABA- and glutamatergic neuron precursors are generated from a neuroepithelial region expressing a homeodomain TF *Nkx6-1*(Lahti et al., 2016; Waite et al., 2012). We here refer to these precursors as rhombencephalic V2 domain (rV2), reflecting its molecular similarity to the V2 domain in the developing spinal cord. The rV2 derived precursors give rise to GABAergic neurons in the posterior Substantia Nigra pars reticulata, Ventral Tegmental Area, Rostromedial Tegmental Nucleus and Ventral Tegmental Nucleus, all important for various aspects of behavioural control (Lahti et al., 2016; Morello et al., 2020a). A failure in the differentiation of the rV2 GABAergic neurons in the embryonic mouse brain, caused by mutation of the *Tal1* gene, results in robust behavioral changes in postnatal mice, including hyperactivity, impulsivity, hypersensitivity to reward, and learning deficits (Morello et al., 2020b).

The *Nkx6-1* positive proliferative progenitors of rV2 give rise to postmitotic V2b-like GABAergic precursors expressing selector genes *Tal1*, *Gata2* and *Gata3*, and V2a-like glutamatergic precursors expressing *Vsx1* and *Vsx2* (Achim et al., 2014; Joshi et al., 2009; Karunaratne et al., 2002a; Muroyama et al., 2005). In the absence of *Tal1*, *Gata2* and *Gata3* function, differentiation of all the rV2 precursors is directed to glutamatergic neuron types (Achim et al., 2012; Lahti et al., 2016). *Tal1*, *Gata2* and *Gata3* also retain their expression in the mature GABAergic neurons. Thus, *Tal1*, *Gata2* and *Gata3* TFs operate as fate-selectors, responsible for the induction and maintenance of the GABAergic identity in the anterior ventrolateral brainstem.

It has remained unclear how the selector TFs are activated during development and how they guide neuronal differentiation in the anterior brainstem. To address these questions, we have characterized the accessibility and TF binding in putative gene regulatory elements and correlated these events with changes in gene expression during differentiation of GABAergic and glutamatergic neurons in the anterior brainstem. Our results suggest that the GABAergic selector TF genes are activated by interconnected mechanisms reflecting developmental regionalization and timing, and that their functions converge to promote GABAergic and suppress glutamatergic neuron differentiation.

Materials and Methods

Data and code availability

The E12.5 mouse rhombomere 1 scRNAseq and E12.5 scATACseq reads generated in our study are deposited in SRA. The BioProject IDs of the E12.5 R1 scRNAseq data are: GEO GSE157963; BioProject PRJNA663556 E12.5 R1 samples SAMN16134208, SAMN16134206, SAMN16134205, SAMN16134207; and E12.5 R1 scATACseq: BioProject PRJNA929317, samples SAMN32957134, SAMN32957135.

All original code with used parameter values will be deposited in GitHub at (https://github.com/ComputationalNeurogenetics/rV2_bifurcation)
Any additional information required to reanalyze the data reported in this paper is available upon request.

The exact cell number, read yields and read quality for each sample are found in (Supplementary Methods 1 and https://github.com/ComputationalNeurogenetics/rV2_bifurcation/seq_quality_reports). Batch effect was evaluated not to be meaningful based on the distribution of cells from biological replicates in the UMAPs of both r1 E12.5 scRNA-seq and scATAC-seq (Supplementary Methods 1).

scRNAseq samples

scRNAseq sample preparation is described in detail in (Morello et al., 2020a). The GEO and BioProject IDs of the data are listed above.

Processing of scRNAseq data

scRNAseq data processing per brain region is described in detail in (https://github.com/ComputationalNeurogenetics/rV2_bifurcation) with main steps outlined below.

scRNAseq data was read into *Seurat* objects from 10x *filtered_feature_bc* matrix by using Ensembl gene IDs as features. Replicates per each brain region, or all replicates in the later analysis steps were merged as *Seurat* objects. After merging, quality control (QC) was performed by calculating *mt-percentage and filtering based on nFeature_RNA* and *percent.mt*. Data was then log-normalized, top 3000 most variable features were detected, finally data was scaled and *mt-percentage*, *nFeature_RNA*, *nCount_RNA* were set to be regressed out.

Data was then run through PCA using n variable features, and *Seurat RunUMAP* with selected top PCA dimensions was used to generate Uniform Manifold Approximation and Projection (UMAP) for further analysis. *Seurat* functions *FindNeighbors* and *FindClusters* were applied for neighborhood and community detection-based clustering.

scATACseq samples

Embryonic brains were isolated from E12.5 wildtype mouse embryos (n=2), and ventral R1 pieces (E12.5) were separated. Nuclei were isolated using the 10X Genomics Assay for Transposase Accessible Chromatin Nuclei Extraction from fresh tissues protocol CG000212. Chromium Single Cell Next GEM ATAC Solution (10X Genomics, 1000162) was used for the capture of nuclei and the accessible chromatin capture. In chip loading, the nuclei yield was set at 5000 nuclei per sample. Chromium Single Cell Next GEM ATAC Reagent chemistry v1.1 was used for the DNA library preparation. Sample libraries were sequenced on Illumina NovaSeq 6000 system using read lengths: 50bp (Read 1N), 8bp (i7 Index), 16bp

(i5 Index) and 50bp (Read 2N). The reads were quality filtered and mapped against the *Mus musculus* genome *mm10* using Cell Ranger ATAC v1.2.0 pipeline.

Processing scATACseq data

In this study we defined feature space as described previously (Kilpinen et al., 2023), except that we used both E14.5 (Kilpinen et al., 2023) and E12.5 scATACseq samples to further increase resolution to detect features from rare cell population and cellular clade cut height was defined based on smallest clade so that none of the clades contained less than 100 cells (h=8).

These jointly defined features were then used to process the E12.5 rhombomere 1 scATACseq data. Data was subjected to QC, binarization and LSA, steps described in more detail in (https://github.com/ComputationalNeurogenetics/rV2_bifurcation) largely following steps previously described (Stuart et al., 2021, 2019). Similarly, scATACseq data was integrated with scRNAseq using an approach already described (Stuart et al., 2019).

The UMAP projection based on scATACseq data was done with *Seurat* using SVD components 2-30 for cells with a reliability score from scRNAseq integration >0.5 (Figure 2A). The first SVD component was omitted due to the association with read depth. Neighborhood and community detection were made with Seurat functions *FindNeighbors*(dims=2:30) and *FindClusters*(resolution=1, algorithm=4).

Subsetting and subclustering the rV2 lineage

The rV2 lineage was subset from clustered E12.5 R1 scRNAseq data (Supplementary Figure 1A) as well as from E12.5 R1 scATACseq data (Supplementary Figure 2A). Subsetting was based on the expression levels of known marker genes (Ccnb1, Nkx6-1, Vsx2, Slc17a6, Tal1, Gad1). scATACseq clusters 1, 6, 8, 12, 13, 19, 20, 24 and 26 were considered to consist of rV2 lineage. Clusters 13 and 19 were considered to be progenitors and were named PRO1 and PRO2, respectively. Cluster 26 was considered to consist of common precursors and was named CO. Clusters tentatively belonging to common precursors, the GABAergic (branches 1, 6, 20 and 24), and the glutamatergic branches (8 and 12) were further subjected to Seurat *FindSubClusters* to reveal higher resolution cell types. This resulted in the splitting of CO into clusters CO1 and CO2, the GABAergic branch into six clusters named GA1 – GA6, and the glutamatergic branch into five clusters GL1 – GL5.

rV2 cluster correspondence between scATACseq and scRNAseq

To evaluate the overlap of marker genes between scRNAseq and scATACseq clusters (Supplementary Table 1, 2), we performed a hypergeometric test and visualised the results using a heatmap (Figure 1G).

The scRNAseq and scATACseq cluster markers were filtered with adj. p-value < 0.05 and top 25 based on log2FC. An intersection of unique gene names was identified between

scRNAseq and scATACseq objects and its size, N, was used as the total gene pool for the hypergeometric test.

VIA pseudotime analysis

For VIA (Stassen et al., 2021) based pseudotime analysis rV2-lineage embeddings coordinates from dimensions 2:30, as well as cluster identities as colors, were passed from Seurat object to Python based VIA. Calculations in both for scRNAseq and for scATACseq, gene expression levels were used to calculate the pseudotime value. In the case of scATACseq, these were the interpolated values from integration with the scRNAseq (Stuart et al., 2019). In the case of scATACseq PRO1 was chosen as the root cell population, and for scRNAseq cluster 22 was chosen as root cell population.

Sequence conservation

Conservation data is derived from UCSC database at (https://hgdownload.cse.ucsc.edu/goldenPath/mm10/phastCons60way/), which contains compressed phastCons scores for multiple alignments of 59 vertebrate genomes to the mouse mm10 genome (Siepel et al., 2005). Nucleotide was considered conserved if phast-score was at least 0.5 (range 0-1).

The scATACseq features and other selected regions of study were examined for conservation level in comparison to the whole genome, using Fisher's exact test. Conservation was also calculated for exon and intron regions within all chromosomes to demonstrate the accuracy of this method.

For sequence conservation at the TFBS-motif locations, positional weight matrix (PWM) of the motif was used to define weight of each nucleotide based on max probability at that nucleotide location in PWM and that weight was used in calculation of the weighted mean of conservation score over the nucleotides at TFBS locations.

TAD

Topologically associated domains (TADs) of the chromatin (Singh and Berger, 2021) were used to limit the search for cCREs. We used *tadmap* (Singh and Berger, 2021) for TAD definition.

Feature accessibility over pseudotime

Features within TAD of the selector gene were considered to be candidate cis-regulatory element (cCRE) if they had linkage correlation with expression of the gene with abs(zscore)>2 and p-value <0.05. Accessibility changes of cCREs were visualized as spline smoothed heatmaps over the pseudotime. Spline smoothing was calculated over entire rV2 cell groups in order of (PRO1-2, CO1-2, GA1-2, GA3-4, GA5-6, GL1-2, GL3-4, GL5).

TF-footprinting from scATACseq reads

TF-footprinting of scATACseq data was performed by using TOBIAS-framework (Bentsen et al., 2020) with Snakemake pipeline. As motif collection we used Hocomoco version 12 (Vorontsov et al., 2023). The motif for INSM1 was added to the Hocomoco v12 collection from Hocomoco v11 collection as it was seen relevant in earlier research. Suitability of mixing motifs from the two versions of the Hocomoco was checked in correspondence with Hocomoco team.

For the purpose of TF-footprinting we combined cell groups as follows to increase signal to noise ratio: PRO1 and PRO2 (PRO1-2), CO1 and CO2 (CO1-2), GA1 and GA2 (GA1-2), GA3 and GA4 (GA3-4), GA5 and GA6 (GA5-6), GL1 and GL2 (GL1-2), GL3 and GL4 (GL3-4). GL5 was kept as is. E12R1 BAM files were merged, and PCR duplicates removed with Picard (https://broadinstitute.github.io/picard/). BAM file was then split into BAM file per combined cell group with Sinto (https://timoast.github.io/sinto/).

TOBIAS was run with parameters given in (https://github.com/ComputationalNeurogenetics/rV2_bifurcation) and the cell group and genomic location specific footprint results were imported into an SQLite database (total of 49,390,084 records).

Footprinting data was visualized through R ggplot based dotplots per selected features, with filtering for TFBS-motifs having >0.5 weighted average conservation value, accessibility of the feature > 0.06 and expression of the TF in question > 1.2 (log1p) in any of the PRO1-2, CO1-2, GA1-2 or GL1-2 groups.

RNAscope® *in situ* hybridization

E12.5 wildtype mouse embryos were dissected and fixed in 4% paraformaldehyde (PFA; Sigma-Aldrich, Cat#P6148) in 1× PBS for 36 hours at room temperature. The fixed brains were embedded in paraffin (Histosec™ Pastilles (without DMSO, Cat#115161)), and sectioned on Leica microtome (Leica RM2255) at 5 µm. RNAscope® Multiplex Fluorescent Reagent Kit v2 (ACD, Cat#323270) was used to detect RNA transcripts. The paraffin sections were deparaffinized in xylene and ethanol series, treated with RNAscope® Hydrogen Peroxide (ACD, Cat#322381) for 10 minutes, and subjected to antigen retrieval in 1× Target Retrieval Solution (ACD, Cat#322000) for 15 minutes at 99–102°C. After rinsing and dehydration, hydrophobic barriers were drawn using an ImmEdge Pen (ACD, Cat#310018), and slides were dried for 30 minutes or overnight. Protease Plus (ACD, Cat#322381) was applied for 30 minutes at 40°C, followed by hybridization of probes at 40°C for 2 hours. Probes included mouse Ebf1 (ACD, Cat#433411, C2), E2f1 (ACD, Cat#431971, C1), Insm1 (ACD, Cat#43062, C1), Sox4 (ACD, Cat#471381, C2), Tal1 (ACD, Cat#428221, C4), and Tead2 (ACD, Cat#42028, C3). Positive (ACD, Cat#320881) and negative (ACD, Cat#320871) control probes were included. C1 probes were used undiluted; C2, C3, and C4 probes were diluted 1:50 in C1 probe or Probe Diluent (ACD, Cat#300041). Slides were washed in Wash Buffer (ACD, Cat#310091) and stored in 5× saline-sodium citrate (SSC) buffer overnight. Amplification was conducted using AMP1, AMP2, and AMP3

reagents (30 min, 30 min, 15 min) at 40°C. Signal detection was performed with HRP channels (C1, C2, C3) and TSA Vivid Fluorophores: TSA 520 (ACD, Cat#323271),TSA 570 (ACD, Cat#323272), and TSA 650 (ACD, Cat#323273), diluted 1:1500 in RNAscope® Multiplex TSA Buffer (ACD, Cat#322809). Each HRP step included blocking and washing. Slides were counterstained with DAPI (30 seconds), mounted with ProLong Gold Antifade Reagent (Thermo Fisher, Cat#P36934), and imaged using a Zeiss AxioImager 2 microscope with Apotome 2 structured illumination. Images were processed in ImageJ, and figures prepared in Adobe Illustrator CC 2020. Slides were stored at 4°C and imaged within two weeks.

RNAscope image segmentation and downstream analysis

We trained a Stardist (Schmidt et al., 2018; Weigert et al., 2020) model to segment the RNAScope images, for this purpose we manually painted cell nucleus onto 10 representative microscopy images of E12.5 rV2 area based on DAPI channel.

The trained model was then used to segment the analysis images and to return ROIs as ImageJ ROI coordinates. Regions of interest (ROIs) were imported into R for analysis. Fluorescence signals were quantified for three channels in each image. The boundary of the ventricular zone (VZ) was defined manually using three points, and calculating circle boundary passing these points (Supplementary Figure 3), and the distances of individual cells from the VZ were calculated in units of average cell diameter (avg. across all images). Fluorescence signals were filtered to remove background noise using a quantile threshold of 0.25. Raw fluorescence intensity was measured as the number of fluorescence dots per cell. To analyze patterns across biological replicates, fluorescence data from three independent experiments were included in the same aggregate analysis. Fluorescence intensities were normalized across replicates for each channel to account for variations in signal strength and aligned to ensure uniform spatial scaling. Aggregated data were visualized by calculating a smoothed sliding mean (window size=12) fluorescence trend for each channel, highlighting consistent spatial patterns of *Tal1*, *Tead2*, and *E2f1* fluorescence intensities across biological replicates.

CUT&Tag

Embryonic brains were isolated from E12.5 wildtype (ICR) mouse embryos, and ventral R1 pieces (E12.5) were separated. The ventral R1 pieces from the embryos of the same litter were pooled and dissociated. The cells were then divided between the samples. Embryos from different E12.5 mouse litters were used as biological replicates in CUT&Tag. The samples were processed following the CUT&Tag-It Assay Kit protocol (Active Motif, cat # 53170). The following primary antibodies were used: rabbit anti-Gata2 (Abcam, ab109241), 1:80 rabbit anti-Gata3 (Boster, M00593), rabbit anti-Tal1 (Abcam, ab75739), rabbit anti-Vsx2 (BioSite, 25825-1-AP-20), rabbit anti-Ebf1 (Boller et al., 2016) rabbit anti-Insm1 (BioSite, ASJ-IO4DE3-50), rabbit anti-Tead2 (Biorbyt, orb382464), rabbit anti-IgG (Cell Signalling, 86652/66362) and rabbit anti-H3K4me3 (Cell Signalling, 86652/9751). All primary antibodies were used at 1:80 dilution. Sample and sequencing library preparation was done following the recommended protocols of the assay. The libraries were multiplexed

at 10 nM equimolar ratios and sequenced on NextSeq 500 or Aviti. 75 bp paired-end sequences were obtained.

The demultiplexed reads were quality filtered and aligned against the *Mus musculus* genome *mm10*, using the nf-core (Ewels et al., 2020) pipeline nf-core/cutandrun.

Analysis of the TFs interacting with the selector gene cCREs

To search for the regulators of selector genes, we first searched for TFs per each selector gene per cell group with conditions as follows:

1) The TF has a footprint in any of the cCREs linked to the selector gene with LinkPeaks abs(z-score)>2 and p-value<0.05, and
2) The weighted mean of the conservation of the motif location is >0.5, and
3) The TF is expressed > 1.2 (log1p) on average in the cell group in question.

Defining the common regulators between the selector genes was done by comparison of the regulators found for individual selectors based on their intersects. Statistical significance levels reported for individual regulators are defined as minimum p-value for cCRE to gene linkage for each regulator per each selector gene and maximum p-value over all selectors in the intersect. To assess whether the observed overlap in regulators among the three selector genes was greater than expected by chance, we performed a permutation-based test. First, we determined the actual set of regulators for each of the three selector genes and calculated the size of their intersection. Next, to construct a null model, we repeatedly and randomly selected three sets of regulators (of the same sizes as the original sets) from the complete universe of candidate regulators, recalculating the intersection size each time. With 1 000 000 iterations we obtained a distribution of intersection sizes expected under random selection. The empirical p-value was then estimated as the proportion of these randomized trials in which the intersection size was equal to or larger than the observed intersection. If no randomized trial matched or exceeded the observed intersection size, we reported the p-value as less than an upper bound (e.g., <1/(N+1)), reflecting the rarity of the observed overlap under the assumed null hypothesis.

Selector gene target analysis

To uncover general functional roles of the selector genes at the bifurcation point we conducted a gene set enrichment (GSEA) (Mootha et al., 2003) analysis of the putative target genes of *Tal1*, *Gata2*, *Gata3* and *Vsx2* over the axis of avg_log2FC of mouse genes in between GA1-2 and GL1-2. For the GSEA axis only genes having an adjusted p-value <0.01 in the DEG test (Seurat FindMarkers with Wilcoxon) and an expression proportion of at least 0.1 in either group were taken into account.

Target genes were defined according to following conditions:

1) The selector gene (TF) motif is associated with a footprint in either GA1-2 or GL1-2, and
2) the motif situates in a scATACseq feature which is linked to the target gene with LinkPeaks abs(z-score)>2 and p-value<0.01, and
3) the weighted mean of the conservation of the motif location is >0.5, and
4) there is a CUT&Tag peak of the selector gene overlapping the feature.

As our GSEA axis should be considered relevant from both ends, one end signifying targets overexpressed in GA1-2 and another in GL1-2, we additionally looked for leading edge (traditionally called leading and trailing edges in GSEA analysis) genes from both ends. We defined GA1-2 leading edge genes to be the ones before the GSEA enrichment score gains its maximum, and GL1-2 leading edge genes to be ones after the GSEA enrichment score gains its minimum.

Comparison of selector gene targets and cell type markers

The selector gene target list was subset, selecting the genes with exp_avg_GL1-2>0.5 (log1p) OR exp_avg_GA1-2>0.5 (log1p). The resulting gene list was used as input in online Enrichr tool (Xie et al., 2021). The results against CellMarker database (Hu et al., 2023) were extracted.

Overlaps of the cell-type specific target genes of selectors.

Cell-type specific genes were defined as log2FC>0.5; exp_avg_GL1-2<0.5 (log1p); exp_avg_GA1-2>0.5 (log1p) for GABAergic neuron specific genes and log2FC<-0.5; exp_avg_GL1-2>0.5 (log1p); exp_avg_GA1-2<0.5 (log1p) for glutamatergic neuron specific genes. The Tal1, Gata2 and Gata3 target gene lists were subset using these criteria. Overlaps of the resulting gene lists were found using venn diagram app as shown on Figure 6.

ChromVAR analysis with Hocomoco motifs

ChromVAR analysis (Schep et al., 2017) was performed using Hocomoco (Vorontsov et al., 2023) v12 motifs and an INSM1 motif lifted from Hocomoco v11 by using functions built into Signac package. The motifs whose chromatin accessibility correlated with the RNA expression of their respective genes (Spearman correlation > 0.5 and p-value < 0.01) were identified, and their ChromVAR z-scores across rV2 lineage cell groups were aggregated by mean to a level of TF and then visualised using ComplexHeatmap R package applying clustering method “complete” and the clustering distance as “euclidean".

Results

Transcriptomic dynamics and chromatin accessibility in differentiating rV2 GABAergic and glutamatergic neurons

To understand the sequence of gene expression changes that correlate with the bifurcation of the rV2 GABAergic and glutamatergic cell lineages, we first studied the transcriptome and the genome accessibility in developing rV2 neurons. We used data from the scRNAseq of the ventral r1 region of E12.5 embryos (Morello et al., 2020a), where the rV2 lineages were identified based on the expression of markers of common proliferative progenitors (*Ccnb1*, *Nkx6-1*) and post-mitotic precursor branches of the rV2 by markers of GABAergic (*Tal1*, *Gad1*) and glutamatergic neurons (*Nkx6-1*, *Vsx2*, *Slc17a6*) (Figure 1A, C, Supplementary Table S1). Using pseudotime method VIA (Stassen et al., 2021), we found that the rV2 cells were arranged on a pseudotemporal axes representing differentiating GABAergic or glutamatergic lineages (Figure 1B). The GABAergic and glutamatergic selector TF genes

(*Tal1*, *Gata2*, *Gata3*, *Vsx1*, *Vsx2*, *Lhx4*) were strongly expressed in the post-mitotic precursors at lineage bifurcation point and at the start of the lineage branches (Supplementary Figure 1B).

To find the gene regulatory elements in the rV2 cell types, we applied scATACseq in E12.5 ventral r1 cells. We then clustered the r1 scATACseq cells based on their open chromatin regions (called features below; Supplementary Figure 2A), followed by integration with the E12.5 scRNAseq data (Stuart et al., 2019) that provides an estimate of the transcript levels for scATACseq measured cells. From the scATACseq data, we readily identified the cell clusters of the rV2 lineage, expressing TFs specific for GABAergic and glutamatergic cell types (Supplementary Figure 2). We identified two progenitor cell groups (PRO1, PRO2), rV2 common precursors (CO), GABAergic neuron branch consisting of 4 cell groups and glutamatergic branch of 2 cell groups (Supplementary Figure 2A). Further clustering of the isolated rV2 cell groups yielded two groups of common precursors (CO1, CO2), six cell groups in the GABAergic branch (GA1-GA6), and five cell groups in the glutamatergic branch (GL1-GL5) (Supplementary figure 2B, Figure 1D, F,). The placement of the scATACseq r1 cells in the pseudotemporal axis (Figure 1E) shows that, as expected, the early state in pseudotime is assigned to cells expressing markers of proliferative progenitors and early post-mitotic common precursors, while late pseudotime state cells express markers of differentiating GABAergic and glutamatergic precursors. Cell clusters have consistently increasing average pseudotime from PRO1 to the end of both GABAergic and glutamatergic branches (Figure 1E). We next compared the scRNAseq and scATACseq cell cluster markers across the rV2 lineages (top 25 marker genes per cluster, similarity testing using hypergeometric distribution) (Supplementary Table 1, Supplementary Table 2, Figure 1G). The similarity scores indicate discrete matches between the scATACseq and scRNAseq clusters, with the exception of PRO1 and PRO2 largely missing a counterpart in the scRNAseq clusters, as the rV2 progenitor cells did not clearly segregate in the scRNAseq data (Supplementary Figure 1A, progenitor markers *Nes* and *Phgdh*) and were therefore excluded from the rV2 lineage analyses. In conclusion, open chromatin features largely reflected the gene expression in the ventral r1 and allowed placing the developing rV2 neurons in cell groups and in a pseudotemporal order.

Chromatin accessibility profiles in genomic regions around selector genes in rV2-lineage subclusters

Tal1, Gata2 and Gata3 are required for GABAergic fate selection in rV2 and their expression is activated in the GABAergic post-mitotic precursors, whereas Vsx2 is expressed in the glutamatergic precursors soon after rV2 lineage bifurcation (Supplementary Figure S1B, Supplementary Figure S2B) (Achim et al., 2013; Lahti et al., 2016; Morello et al., 2020a). To understand the mechanisms behind activation of the expression of these lineage-specific TFs, we analysed the accessibility of chromatin features in *Tal1*, *Gata2*, *Gata3* and *Vsx2* loci in the rV2 cell groups. To find selector gene-associated chromatin features, we tested all

features within the topologically associating domains (TAD) around the transcription start sites (TSS) of the selector TF genes for the correlation between the feature accessibility and the target gene expression. We defined the features linked to the target gene with LinkPeaks z-score below -2 or above 2 and p-value <0.05 as candidate cis-regulatory elements (cCRE) of the gene (Supplementary Table 3) (Stuart et al., 2021)).

The *Tal1* locus, +/-50 kb of the TSS, contains five genes (*Cmpk1, Stil, Tal1, Pdzk1ip1* and *Cyp4x1*) and 20 scATAC features (Figure 2A). *Tal1* TAD contains 107 features. Some of the features at the *Tal1* locus were accessible in all the cell groups, including the progenitors (PRO1, PRO2, Figure 2A-B). These include features at the promoter regions of *Cmpk1* and *Stil* (Supplementary Table 3). Other features become accessible in both GABAergic and glutamatergic precursors after the bifurcation of the lineages, like the *Tal1* +77.4 kb cCRE located at the promoter region of the neuronally expressed *Cyp4x1* gene (Bylund et al., 2002). Accessibility of 15 out of the 107 features in the TAD containing the *Tal1* gene were linked to the *Tal1* expression (p-value < 0.05) (Supplementary Table 3, *Tal1*). Notably, the accessibility of *Tal1* cCREs at +0.7 and +40 kb correlated closely with the expression level of *Tal1* mRNA (Figure 2B), strongly increasing in the earliest GABAergic precursors (GA1) and maintained at a lower level in the more mature GABAergic precursor groups (GA2-GA6), as well as *Pdzk1ip1*, which is activated transiently in the earliest GABAergic precursors (GA1) (Figure 2A, RNA expression). Other features within the *Tal1* gene (+7 kb) as well as in more distal intergenic area (+15.1, +16, +16.4 and +23 kb cCREs) were also found to correlate with the activation of the *Tal1* mRNA expression (Figure 2B). Remarkably, the identified *Tal1* promoter feature (*Tal1* +0.7kb cCRE) contains *Tal1* midbrain and hindbrain-spinal cord enhancers characterized earlier (Bockamp et al., 1997; Sinclair et al., 1999). The *Tal1* +40 kb cCRE corresponds to the previously characterized enhancer region driving expression in the midbrain and hindbrain (Ogilvy et al., 2007). Enhancers at -3.8 kb and +15.1/16/16.4 kb of *Tal1* have been shown to drive *Tal1* expression in hematopoietic stem cells and erythrocytes (Göttgens et al., 2004) (Supplementary Table 4, Supplementary Figure 4A).

Similarly to *Tal1* locus, the TADs of *Gata2* and *Gata3* genes contained features where accessibility increased specifically in the GABAergic precursors, correlating with the activation of *Gata2* and *Gata3* mRNA expression (Supplementary Table 3, Supplementary Figure 5). Two *Gata2* cCREs, at -3.7 kb and +0 kb, are located in the proximal and promoter regions, which have been earlier shown to function as *Gata2* mid- and hindbrain expression enhancers (Nozawa et al., 2009), as well as critical *Gata2* autoregulation elements in hematopoietic cells (Kobayashi-Osaki et al., 2005) (Supplementary Table 4, Supplementary Figure 4B). We also identified an intronic *Gata2* cCRE at +11.6 kb that may mediate the regulation of *Gata2* expression in the spinal cord V2 interneurons by Gata2:Tal1 co-binding (Joshi et al., 2009). Four additional previously uncharacterized *Gata2* cCREs were identified downstream of the *Gata2* gene (+22, +25.3, +32.8 and 47.5 kb cCREs, Supplementary Figure 5A-B) and five more further downstream within the TAD (Supplementary Figure 5B, Supplementary Figure 4B, Supplementary Table 3).

At the *Gata3* locus, the GABAergic precursor –specific features included several intragenic features (*Gata3* +11.1 kb, +17.2 kb, +18.8 kb cCREs), and five features upstream the gene at -2.7 kb to -7.7 kb of *Gata3* TSS (Supplementary Table 3; Supplementary Figure 5C-D). The Gata3 +0.6 kb, +2.2 kb and the five upstream cCREs locate within the previously identified Gata3 CNS enhancer region (Lakshmanan et al., 1999), and two of those also match Gata3 lens expression enhancers (Martynova et al., 2018) (Supplementary Table 4, Supplementary Figure 4C). The *Gata3* +11.1 kb cCRE corresponds to the *Gata3* proximal enhancer driving expression in PNS regions (Lieuw et al., 1997). The *Gata3* +17.2 kb cCRE is located in the intron 3-4 of *Gata3* gene and may have enhancer activity in spinal cord V2b interneurons (Joshi et al., 2009) (Supplementary Table 4, Supplementary Figure 4C).

Importantly, accessibility of the cCREs of the GABAergic selector TF genes changed with different kinetics during neuronal differentiation. For example, the accessibility of *Tal1* +23 kb and *Tal1* +40 kb cCREs was robustly detected before and at the onset of GABAergic differentiation and declined thereafter (Figure 2B). The accessibility pattern of the *Tal1* +23 kb cCRE is particularly interesting. The feature is accessible at 3' position early, and gains accessibility at 5' positions concomitant with GABAergic differentiation (Figure 2A, accessibility). Detailed feature analysis later indicated the binding of Nkx6-1 and Ascl1 that are expressed in the rV2 neuronal progenitors, at 3' positions, and binding of Insm1 and Tal1 TFs that are activated in early precursors, at 5' positions (Described below, see Figure 3C). The accessibility of the Tal1 +0.7 kb cCRE largely reflected the *Tal1* mRNA expression (Figure 2B). Similarly, the *Gata2* +22 kb, +25.3 kb and +32.8 kb cCREs were accessible in the early precursors, before the opening of the *Gata2* -3.7 kb cCRE in GA1 and *Gata2* +0 kb cCRE in GA3-GA6 (Supplementary Figure 5B). Thus, distinct regulatory elements of the selector TFs may be responsible for the activation and for the maintenance of the lineage-specific expression of selector TF gene.

Vsx2 gene expression was detected in the common precursors CO1 and increased in CO2 and the glutamatergic precursors (GL1-GL5) after bifurcation of the rV2 lineages (Figure 2C, RNA expression). At the *Vsx2* locus, an intronic feature (*Vsx2* +20.5 kb) and the distal features at -61 kb and -21.3 kb of *Vsx2* TSS were robustly opened concomitant with its expression activation, whereas accessibility of the *Vsx2* promoter region (*Vsx2* +1.1 kb cCRE) and +27 kb cCRE increased later in the glutamatergic precursors (Figure 2D). The information about the enhancers of Vsx2 brain expression is scarce, but we found that our *Vsx2* r1 CREs at -18.7 kb, -21.3 kb, +1.1 kb and +20.5 kb match previously characterized enhancers of *Vsx2* expression in photoreceptor cells and retinal bipolar neurons, and contain putative autoregulatory Vsx2 biding sites occupied in developing spinal cord motor neuron and interneuron lineages (Supplementary Table 4, Supplementary Figure 4D).

Developmentally important gene regulatory elements are often conserved. We therefore asked if features detected by scATACseq could present more evolutionary conserved sequences than random sequences of mouse genome. To that aim, we analyzed nucleotide conservation in scATACseq features, using phastCons study data over 60 vertebrates from UCSC database (Siepel et al., 2005). We calculated statistical enrichment of conserved

nucleotides within exons, introns and our scATACseq features (Supplementary Table 5). Systematically, odds ratios per chromosome were highest for exons and lowest for introns, odds ratios for features were between these two. All of the odds ratios showed a significant difference from what one would expect if there were no relationship between the nucleotides within features and conservation (p=< 2.43e-33, Fisher's Exact test, Supplementary Table 5). The conservation of chromatin features is consistent with a putative role as regulatory sites.

In summary, we identified the genomic features that are linked to GABAergic and glutamatergic fate selector genes, and where chromatin accessibility changes specifically during rV2 neuron differentiation. Those features could represent regulatory elements of the selector genes. Furthermore, literature research revealed that the previously characterized enhancers, and especially the nervous system expression enhancers of *Tal1*, *Gata2*, *Gata3* and *Vsx2* mostly are found within the ATAC-features identified here (63.6 - 100% of gene enhancers overlapping ATAC-features, Supplementary Figure 6).

Transcription factor activity at the putative regulatory elements of the rV2 cell fate selector genes

Next, we wanted to identify the TFs associated with the putative regulatory elements of *Tal1*, *Gata2*, *Gata3* and *Vsx2* genes. TF binding to chromatin affects its accessibility, leaving a footprint which can be detected in scATACseq data (Bentsen et al., 2020). We applied the footprint analysis across the rV2 cell groups using the known TF binding sites (TFBS) from HOCOMOCO collection (Vorontsov et al., 2023). As validation, we analyzed chromatin interactions of selected candidate TFs in E12.5 r1 using CUT&Tag-sequencing (CUT&Tag) (Kaya-Okur et al., 2019). As a complementary method, we used ChromVAR (Schep et al., 2017) to calculate genome-wide accessibility score for each TFBS we used in the study. While ChromVAR is based on scATACseq data, this analysis is independent of the footprinting and thus provides a robust genome-wide validation of the most active TFBS.

Our TF footprinting analysis identified global TF activity patterns correlating with GABA- and glutamatergic neuron differentiation (Supplementary Figure 7). To find the TFs that may regulate the expression of *Tal1*, *Gata2*, *Gata3* and *Vsx2*, we identified the TFBSs at the conserved genomic locations (weighted average conservation at TFBS > 0.5) within each selector gene cCRE. The results of the footprint analysis are shown for selected features of *Tal1* (Figure 3A-D) and *Vsx2* (Supplementary Figure 8A,C-E), showing the TFs expressed (average RNA expression > 1.2 (log1p)) in any of the four cell groups from PRO1-2, CO1-2, GA1-2 and GL1-2. We observed both common and distinct patterns of TF binding in the cCREs of *Tal1* and *Vsx2* in the rV2 common precursors and early GABA- and glutamatergic cells. Both *Tal1* and *Vsx2* cCREs contain conserved GC-rich areas, E-boxes, TCF class binding sites of TCF and Ebf TFs, regions of Sox TF binding elements 5'TTGT, NF I binding elements TGGCA and binding sites for various homeodomain (HD) TF-s (Figure 3B-D, Supplementary Figure 8C-E). The analysis of TFBSs at the footprint regions suggested that each region may have affinity to several different TFs belonging to a TF class. For example, four Sox TFs are expressed in early rV2 lineage cells (*Sox2*, *Sox3*, *Sox4* and *Sox11*), which all

may compete for the same binding sites in *Tal1* +0.7 kb cCRE or *Vsx2* -61 kb cCRE (Figure 3C, Supplementary Figure 8D).

To further understand the mechanisms of the activation of lineage-specific selector expression, we asked which TFs commonly and specifically regulate the three GABAergic selector TFs: *Tal1*, *Gata2*, and *Gata3*. Search for the common regulators of GABAergic fate selectors revealed 21 TFs in CO1-2 and 22 TFs in the GA1-2 cell groups (Figure 3E). Interestingly, the common regulators between the stages are overlapping partially, with the TFs associated with the neurogenic cell cycle exit predominant in the common precursors (CO) and the postmitotic differentiation and the autoregulation network TFs showing activity later, in the GABAergic precursors.

We next asked which upstream regulators are common to both rV2 GABAergic (*Tal1*, *Gata2*, *Gata3*) and glutamatergic (*Vsx1* and *Vsx2)* fate selector genes. The analysis of common regulators revealed 12 TFs, including *Ascl1*, *E2f1*, *Nkx6-1* and *Vsx1* may regulate both GABA- and glutamatergic selectors in the CO1-2 cell group (Figure 3E and Supplementary Figure 8F). The distinct regulators of GABA-and glutamatergic fate selectors, aside the lineage-specific autoregulatory functions of the selectors themselves, include *Lhx1* and *En2* in GABAergic sublineage (*Lhx1* binding *Tal1* +0.7 kb feature, Figure 3B), and *Lhx4*, *Shox2* and *Nhlh2* regulating *Vsx2* expression in the glutamatergic lineage (Supplementary Figure 8C-E). Footprints at Nkx6-1 TFBS were detected in the cCREs of all selectors in common precursors, but after onset of differentiation become restricted to the cCREs of *Vsx2* and *Vsx1* in the glutamatergic cells. Interestingly, the *Vsx2* -68.8 kb cCRE contains conserved GATA TF binding sites (Supplementary Figure 8E). Conversely, *Vsx1* and *Vsx2* binding sites were found in the cCREs of *Gata2* and *Gata3*, suggesting a mutual cross-regulation between the GABA-and glutamatergic selector genes. Putative *Vsx1* binding sites were detected in all three GABAergic fate selector genes, *Tal1*, *Gata2* and *Gata3*, in the rV2 common precursors (Figure 3E, CO, CO1-2).

The proneural bHLH TFs appeared to contribute to the activation of the selector genes. Interestingly, *Ascl1* and *Neurog2* TFBS are found at the cCREs of both *Tal1* (Figure 3C, D) and *Vsx2* (Supplementary Figure 8C, D). In addition, *Vsx2* is regulated by *Neurog1* in the rV2 common precursors (Supplementary Figure 8C, D).

Interestingly, several cell cycle-associated TFs may contribute to the regulation of selector gene expression. Those include *Sox2*, *Sox21*, *Insm1*, *Ebf1*, *Ebf3, and E2f1*. *E2f1* and *Ebf1-3* binding sites are located within the GC-rich areas (Figure 3B-D, Supplementary Figure 8C-E). The expression of *E2f1* declines after differentiation, while *Ebf1-3* are expressed in common precursors and maintained in differentiating neurons of both GABA- and glutamatergic sublineages (Figure 3B-D, Supplementary Figure 8C-E).

In addition to the footprint analysis, we analyzed the chromatin interactions of *Tal1*, *Gata2*, *Gata3*, *Vsx2*, *Ebf1* and *Insm1* in the E12.5 r1 by CUT&Tag (Supplementary Figure 9, Figure 3A, Supplementary Figure 10, Supplementary Figure 11). The binding of *Ebf1* and *Insm1*, as well as *Tal1*, *Gata2*, and *Gata3* at the *Tal1*-linked features was confirmed by CUT&Tag (Figure 3A, Supplementary Figures 9-11), consistent with the autoregulatory activity of the GABAergic fate selector genes. Interestingly, in addition to the *Tal1* +0.7 kb and +40 kb

cCREs, we detected *Tal1*, *Gata2* and *Gata3* binding at the *Tal1* +23 kb cCRE, where scATACseq-footprinting did not indicate binding at conserved sites (Figure 3A, C).

In summary, the regulatory feature analysis revealed several mechanisms that may contribute to the timing of activation and the maintenance of selector TF expression. Firstly, we found the binding sites of cell cycle progression or progenitor identity-associated TFs *E2f1* and Insm1, proneural TFs *Ascl1*, *Neurog1*, *Neurog2*, *Neurod1* and *Neurod2*, as well as TFs expressed after cell cycle exit, such as *Sox4* and *Ebf3* (*Coe3*) at the putative distal regulatory elements. Conserved binding sites for other *Ebf* TFs, *Ebf1* and *Ebf2* are also found in the cCREs of *Tal1* and *Gata2*, but not *Gata3*. In addition, the proximal elements and the promoter regions of *Gata2* and *Tal1* genes contained footprints at conserved GATA::TAL motif in the GABAergic neurons, indicating autoregulation of the selector TFs. Similarly, conserved *Vsx1* and *Vsx2* motifs in *Vsx2* cCREs were associated with ATAC-footprints in the glutamatergic neurons.

Pattern of expression of the putative regulators of the selector TF expression

Our analysis of TFBSs in the selector gene cCREs indicated that GABAergic selector TF expression may be regulated by E2f1 and Ebf1, TFs involved in cell cycle progression and neuronal differentiation. In addition, we found mediators of Hippo and Notch signalling pathways, *Tead2* and *Insm1*, among the possible regulators of *Tal1*.
To test whether the *E2f1*, *Ebf1*, *Sox4*, *Insm1* and *Tead2* genes are expressed in the *Tal1+* lineage precursors, we analysed their pattern of expression in the E12.5 mouse r1 by RNA *in situ* hybridisation (RNAscope). In the embryonic brain, the progression of neuronal differentiation follows the apical-basal axis, as the newly born precursors exit the ventricular zone and migrate to intermediate and mantle zones. *Tal1* is expressed shortly after the cell cycle exit. The expression of *Insm1*, *E2f1*, *Ebf1* and *Tead2* RNA overlapped with *Tal1* RNA expression in the intermediate zone of the ventrolateral r1 at E12.5 (Figure 4A-B, E-F, I-J). We quantified the fluorescence signal intensity along the apical-to-basal axis of the neuroepithelium, showing that the maximum intensity of *Insm1*, *E2f1*, *Ebf1* and *Tead2* probe signal is reached slightly closer to the apical surface than that of *Tal1* (Figure 4C, G, K). The expression of *Sox4* RNA (Figure 4A-B) was detected in the neuronal cells at more basal position in the neuroepithelium, where the differentiated neurons reside. The expression patterns observed along the apical-basal axis in the E12.5 r1 neuroepithelium matched with the expression levels detected by scRNAseq in the respective cell groups (Figure 4C-D, G-H, K-L).

In conclusion, *Insm1*, *E2f1*, *Ebf1* and *Tead2* are expressed in the early rV2 precursors, including the *Tal1* expressing GABAergic cell lineage. The expression of these TFs precedes the expression of *Tal1* on the differentiation axis, consistent with a putative role in the initiation of *Tal1* gene expression.

Regulatory signature of developing rV2 GABAergic neurons

To shed light on the genome-wide activity of TFs in the differentiating rV2 neurons, we applied ChromVAR (Schep et al., 2017). Consistent with the footprint analysis in cCREs, ChromVAR analysis indicated that accessible genomic regions in rV2 progenitor and precursor cells are enriched in *Sox*, *NeuroD1* and *Nfia/b/x* TF binding motifs (Figure 5A). In addition, *Nhlh1/2*, *NeuroD2*, *Bhlhe22*, *Ebf1-3*, *Pou3f4* and *Lhx5* motifs are found accessible in common precursors, shortly preceding the gain of accessibility in *Gata2/3* and *Tal1* motifs in GABAergic neurons (Figure 5A, square). We also analyzed the activity of *Tal1*, *Gata2* and *Gata3* in the rV2 neurons more closely. For those TFs, the accessibility at most of their TFBSs closely follows the TF RNA expression (Figure 5B-C). Interestingly, there are three TFBS motifs of *Tal1* in the HOCOMOCO v12 resource. Of these *Tal1* TFBSs, the activity pattern of the dual GATA:TAL binding motif (Figure 5C, TAL1.H12CORE.0.P.B) was most specific to the GABAergic lineage, similar to the GATA2 and GATA3 TFBSs (Figure 5C). The other TFBSs of *Tal1* showed lower activity (TAL1.H12CORE.1.P.B) or an earlier peak of activity in common precursors with a decline after differentiation (TAL1.H12CORE.2.P.B) (Figure 5C).

Targets of the GABAergic fate selectors in the developing rhombomere 1

We next aimed to describe the roles of GABAergic fate selector genes in the determination of the GABA- vs glutamatergic phenotype in rV2 cell lineages in more detail. Using the combined evidence of TF footprints, TFBS conservation and *in vivo* CUT&Tag experiments (Figure 6A), we identified 3948 target genes of *Tal1*, 780 target genes of Gata2, and 721 target genes of *Gata3* (Figure 6B). About half of the identified target genes are differentially expressed in GABA- or glutamatergic neurons (statistically significant differential expression between in GA1-2 and GL1-2 groups with p-value <0.01 in Wilcoxon's test). However, most of the identified target genes of *Tal1*, *Gata2*, and *Gata3* are lowly expressed in the GABA- and glutamatergic neurons, with only 30% of the genes being expressed above RNA avg(exp)>0.5 (log1p) in GA1-2 or GL1-2 cell groups (Supplementary Table 6). GSEA (Subramanian et al., 2005) showed that *Tal1* target genes are expressed in both GABA- and glutamatergic neurons (Figure 6C), while *Gata2* and *Gata3* target genes are overexpressed in GABAergic over glutamatergic precursors (Figure 6 H, M). To further characterize the identified targets, we compared the numbers of target features associated with the target gene by a positive or negative feature-to-gene expression link (Figure 6 D, I, N). The positive links outnumbered negative links by roughly factor of 2, which may reflect transcriptional activator function of the studied TFs, but also the complexity of gene regulation as there could be multiple, positive and negative links to each gene. We also calculated the amount of target genes considering their closest linked feature with distance bins of 0-5kb, 5-50kb and >50kb (Figure 6 E, J, O). The results showed largely expected distribution where category of longest links contained most target genes. This is likely due to the category being limited by TAD which are often large, yielding considerable amounts of distal regulatory elements for many genes. We next calculated the expression variability of the target genes, stratified by the distance to closest selector-bound linked feature. The target genes with proximal selector-

bound features (0-5 kb) show the most cell type-specific patterns (Figure 6 F, K, P). For example, the variability of proximal feature-linked *Tal1* target genes in rV2 cell groups was 0.509 - 0.423, and 0.282 - 0.191 for distal feature-linked genes. Thus, regardless of the large amounts of potential distal regulatory elements for many targets, the targets having a proximal selector-bound regulatory element show most drastic changes in expression over the rV2 lineage cell states. However, a few cell type-specific genes also appeared to be regulated by distal features at >50 kb from TSS (*St18*, *Nkx6-1*, *Nhlh1*) or 5-50 kb from TSS (*Zfpm1*, *Lmo1*) in the case of Tal1 targets (Supplementary Table 6). Finally, we also investigated cell type associations of the target genes (exp>0.5 (log1p) in GA1-2 or GL1-2 cell groups) by calculating scores against CellMarker 2024 database (Chen et al., 2013; Hu et al., 2023), revealing clear enrichment of inhibitory interneuron markers among the Tal1, Gata2 and Gata3 target genes (Figure 6 G, L, Q).

Finally, we analyzed the overlap of the cell type-specific target gene lists, aiming to find genes regulated by all the GABAergic fate selectors (common target genes) or by any selector TF specifically. No common target genes were found among the glutamatergic neuron specific genes (Figure 6R). The genes commonly regulated by Tal1, Gata2 and Gata3 are exclusively overexpressed in the GABAergic lineage and contain genes comprising the autoregulatory and GATA co-factor networks (*Gata2*, *Tal1*, *Zfpm1*, *Lmo1*). The GABAergic phenotype (*Gad1*, *Gad2*, *Slc32a1*) is regulated by Gata2 and Gata3 (Figure 6S). Tal1 seems to be involved in the activation or modulation of Notch signaling pathway genes (*Pdzk1ip1*, *Hes5*). Interestingly, Tal1 also interacts with genes specific to rV2 glutamatergic cells (*Lhx4*, *Nkx6-1*), likely repressing their expression in the developing GABAergic precursors. However, as *Tal1* is expressed already in the common precursors of rV2 lineage, the possibility of the activation of *Lhx4* by Tal1 is not excluded. We also suggest the cross-regulation of GABA-and glutamatergic selector genes in the developing rV2 neuronal precursors: Gata3 may repress the *Vsx2* expression in the GABAergic cells, while Vsx2 may repress *Lmo1* in glutamatergic cells, establishing the initiated cell type-specific gene expression programs (Figure 6T).

Discussion

Extensive studies during the past decades have identified several TFs instrumental for neuronal fate decisions in the embryonic brain. How these TFs are regulated and how they guide neuronal differentiation remains less understood. Differentiation of GABAergic and glutamatergic neurons in the ventrolateral r1 is a good model system for TF guided cell differentiation, as in this region the activation of Tal and Gata TFs is both spatially and temporally controlled, and these TFs guide a binary selection between the two alternative cell fates. In this work, we integrated analyses of gene expression, chromatin accessibility and TF occupancy during bifurcation of GABAergic and glutamatergic lineages in the ventrolateral r1. Our results suggest shared mechanisms connecting selector TF activation to

developmental processes and demonstrate combinatory functions of *Tal1*, *Gata2* and *Gata3* TFs in the GABAergic fate selection.

Mechanisms behind selector TF activation. The homeodomain TF code, established by intercellular communication in the neuronal progenitors, defines the competence of differentiation into specific lineages. In the ventrolateral r1, the *Tal1*, *Gata2*, *Gata3* and *Vsx2* TFs are activated in the region where the proliferative progenitor cells express the homeodomain TF *Nkx6-1* (Achim et al., 2012; Lahti et al., 2016). Our result suggests that *Nkx6-1* directly regulates the spatial expression pattern of the selector TFs important for acquisition of GABAergic and glutamatergic neuron phenotypes. *Nkx6-1* is not sufficient for selector TF expression, which is activated only when the proliferative progenitor cells exit the cell cycle and give rise to postmitotic precursors. Our results suggest that TFs *E2f1*, *Ebf1* and *Insm1* temporally control the selector TFs and connect their expression to cell-cycle exit. The expression of *E2f1*, *Ebf1* and *Insm1* is not spatially restricted, but is upregulated in early postmitotic precursors across the whole r1 neuroepithelium. Our results show that the expression of these TFs precedes activation of *Tal1*. *E2f1* is best known for its function as a positive regulator of cell cycle progression. However, there also is evidence suggesting a more complex role for *E2f1* in cell cycle regulation. It has been suggested that the effect of *E2f1* is dependent on its level of expression, low levels promoting cell cycle progression but elevated levels cell cycle exit (Radhakrishnan et al., 2004; Shats et al., 2017; Wang et al., 2007). Our results are consistent with this observation. *E2f1* is expressed at a low level in the proliferative progenitors, and its expression peaks concomitant to cell cycle exit. Also, *Ebf1* and *Insm1* have earlier been associated with cell cycle exit and delamination of neuronal progenitors, as well as differentiation and migration of post-mitotic precursors (Faedo et al., 2017; Garcia-Dominguez et al., 2003; Monaghan et al., 2017; Tavano et al., 2018). Thus, the mechanisms that guide cell cycle exit and other cell biological events of neurogenesis may also contribute to activation of neuronal selector TF expression and adoption of the neuronal phenotype.

Selector TFs directly autoregulate themselves and cross-regulate each other. Positive feedback loops are commonly found in the gene regulatory networks involving developmental selector TFs. The positive feedback loops can ensure stable cell fate decisions. Tal and Gata TFs have been shown to autoregulate their own gene expression in hematopoietic cells. Our results suggest that autoregulation also is a prominent feature of the control of *Tal1*, *Gata2* and *Gata3* expression in the developing brain. At maintaining their expression by positive feedback, these TFs appear to work together, potentially as a physical complex involving Lmo1 (Ono et al., 1998), promoting the expression of its components. In addition to the positive cross-regulation of *Tal1*, *Gata2* and *Gata3*, these TFs also interact with the genes of glutamatergic lineage-specific TFs, such as *Vsx2* and *Lhx4*, expressed in the glutamatergic precursors. We envision negative feedback allows cross-repression between the Tal1, Gata2, Gata3, Lmo1 network and Vsx2 in the cells where both alternative fate selectors are expressed. Auto- and cross-regulation would ensure robust patterns of selector TF expression, also after the initial inputs for their activation have ceased.

TFs often have context-dependent activating or repressive effects on their targets, and our data on chromatin occupancy cannot distinguish between the two possible outcomes. However, according to our scRNAseq and earlier studies in the developing spinal cord V2 domain (Francius et al., 2016; Karunaratne et al., 2002b), the time of co-expression of *Gata2*/*3* and *Vsx1*/*2* is short. In the rV2, *Vsx1* is expressed in the rV2 common precursors, where *Tal1*, *Gata2* and *Gata3* expression is still low. Parsimonious explanations would be the negative cross-regulation and differential regulation of *Lmo1*. Furthermore, Vsx1 has been shown to repress the expression of *Tal1* in the zebrafish spinal cord V2a interneurons (Zhang et al., 2020).

Targets of the selector TFs. In addition to establishing cell type-specific patterns of selector TF expression, the selector TFs are thought to be responsible for realization of neuronal identities. Our results show that a combinatory function of the selector TFs is important for adopting a GABAergic identity. This is supported by similar fate transformations observed in *Tal1* and *Gata2*/*Gata3* mutant embryos. Importantly, our results here highlight differences between *Tal1*, *Gata2* and *Gata3* target genes. Most prominently, the activity of *Tal1* is not as clearly biased towards genes expressed in the GABAergic branch as are the genes regulated by *Gata2* and *Gata3*. In general, the *Tal1*, *Gata2* and *Gata3* selector TFs appear to regulate not only GABAergic subtype–specific genes, but also genes expressed more broadly. This suggests that the function of the fate selectors is not restricted to regulation of the neuronal subtype properties such as the neurotransmitter phenotype, but they also participate in enabling the overall cellular function.

Concluding remarks. Our work provides insights into the complex gene regulatory networks regulating spatially and temporally controlled cell fate choices in the developing mouse brainstem. In addition to addressing principles of cell differentiation, our work may also contribute to understanding of neuropsychiatric traits and disorders. The anterior brainstem is an important center for behavioral control, and many of the behavioral disorders have a developmental basis. We have shown that, in mice, a failure in development of *Tal1*-dependent neurons results in severe ADHD-like behavioral changes (Morello et al., 2020a). It is therefore possible that variants in the genomic regions important for gene regulation during brainstem neuron differentiation can lead to altered brain development and predisposition to neuropsychiatric disorders.

Caveats of the study. Partly due to technical limitations, some of the TF – chromatin interactions are inferred only based on ATAC-footprint. Some of the footprints may signify the binding sites of several different TFs, although mostly of the same TF family, and the exact interacting TF may not be identified. In addition, chromatin interaction is only suggestive of a regulatory function.

Acknowledgements

We thank Outi Kostia for expert technical assistance and Iiris Chrons for help with the bioinformatic analyses. We thank Arto Viitanen for his guidance with methods of image

segmentation. We acknowledge the services of FIMM Single Cell Analysis Unit, University of Helsinki. This work was supported by grants from the Academy of Finland, the Sigrid Juselius Foundation, and the Magnus Ehrnrooth Foundation.

Figures

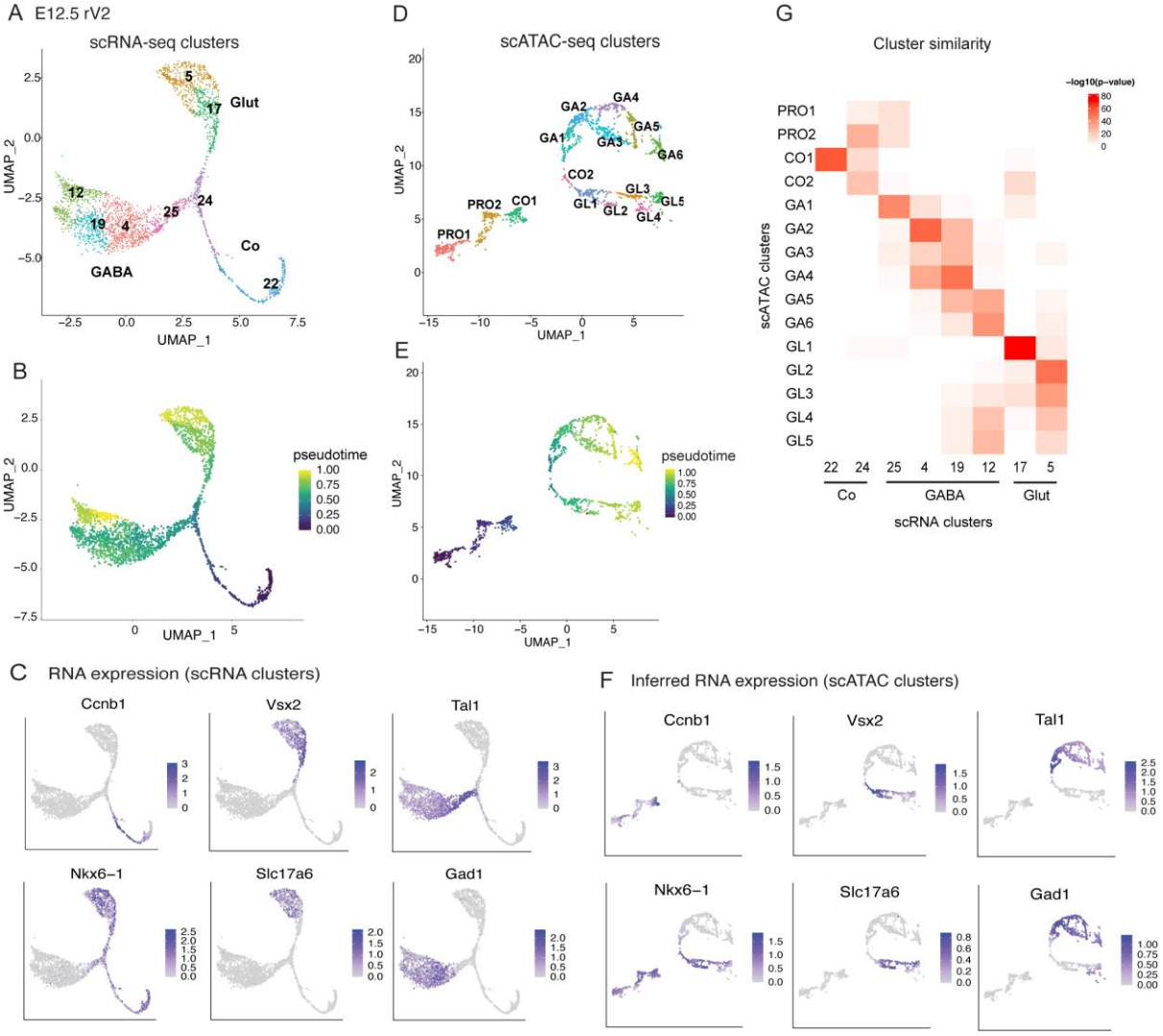


Figure 1. Differentiation of the E12.5 mouse rV2 GABAergic and glutamatergic neurons based on transcriptome and chromatin accessibility

A. UMAP projection of rV2 lineage cells based on scRNA profiles.
B. Pseudotime (VIA) scores shown on the scRNA UMAP projection.
C. Expression of marker genes of progenitors and postmitotic precursors of GABAergic and glutamatergic neurons shown on the scRNA UMAP projection.
D. UMAP projection of rV2 lineage cells based on scATAC profiles.
E. Pseudotime (VIA) scores shown on the scATAC UMAP projection.
F. Expression of the marker genes of progenitors and postmitotic precursors of GABAergic and glutamatergic neurons shown on the scATAC UMAP projection (interpolated expression values).
G. Heatmap of scATAC and scRNA cluster similarity scores.

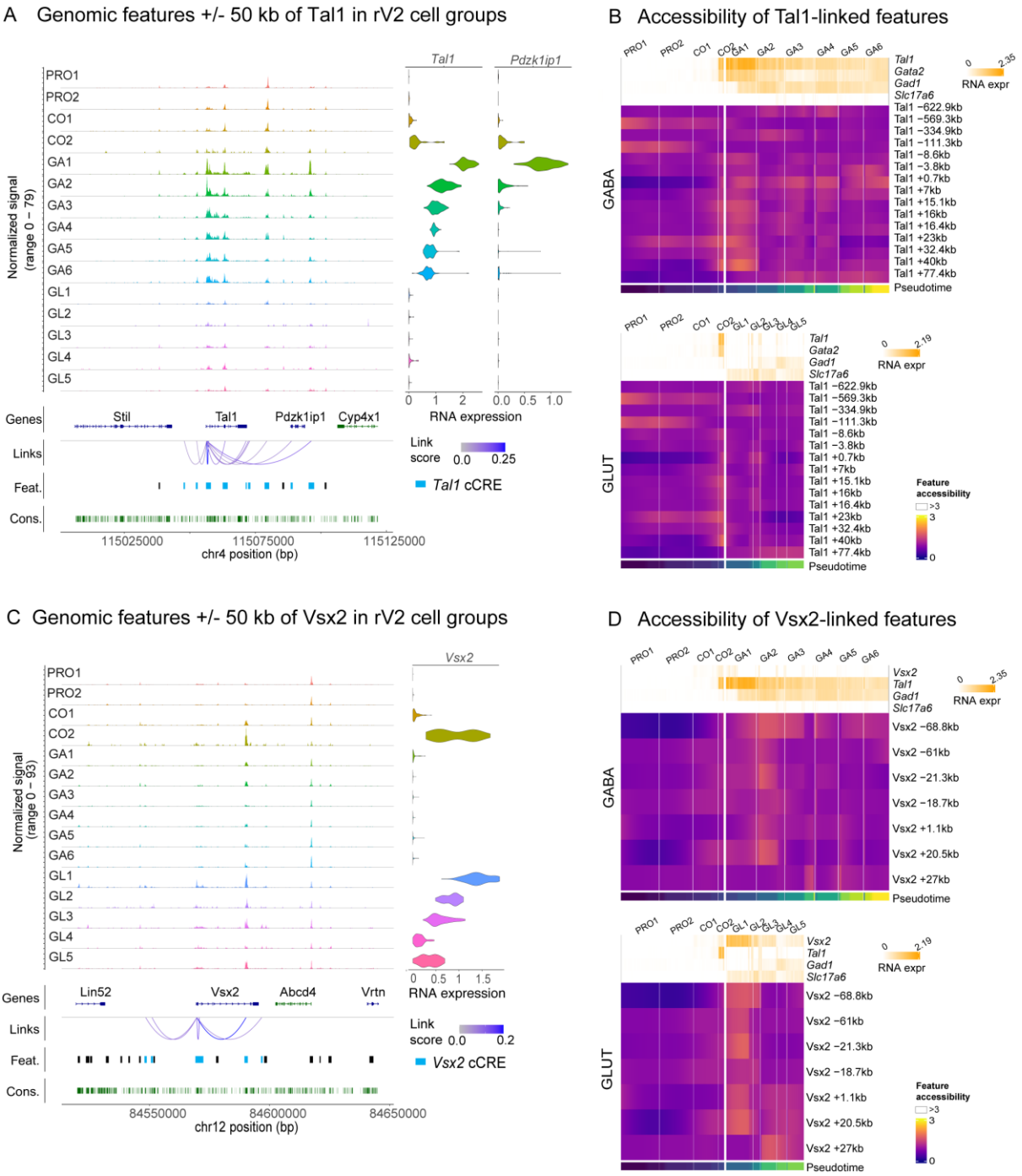


Figure 2. Accessibility of chromatin features in the *Tal1* and *Vsx2* loci during GABAergic and glutamatergic development.

A. Chromatin accessibility per cell group (normalized signal), Ensembl gene track (Genes), scATACseq features (Feat.), feature linkage to gene (Links with the LinkPeaks abs(zscore) >2) and nucleotide conservation (Cons.) within +/-50 kbp region around *Tal1* TSS. Violin plots on the right show the expression levels of *Tal1* and *Pdzk1ip1* per cell group.

B. Spline-smoothed z-score transformed heatmaps of chromatin accessibility at *Tal1*-linked scATACseq features (*Tal1* cCREs) in the single cells of rV2 GABAergic and glutamatergic lineages with RNA expression (sliding window mean(width=6) smoothed) of *Tal1*, *Gata2*,

Gad1 and *Slc17a6* as column covariable (top). Cells on the x-axis are first grouped per cell group and then ordered by the pseudotime (bottom) within each group.

C. Same as in A, but for the *Vsx2* locus.

D. Accessibility of *Vsx2* cCREs in the rV2 GABAergic and glutamatergic lineages, shown as in (B). The expression levels of *Vsx2*, *Tal1*, *Gad1*, and *Slc17a6* are shown above the heatmaps.

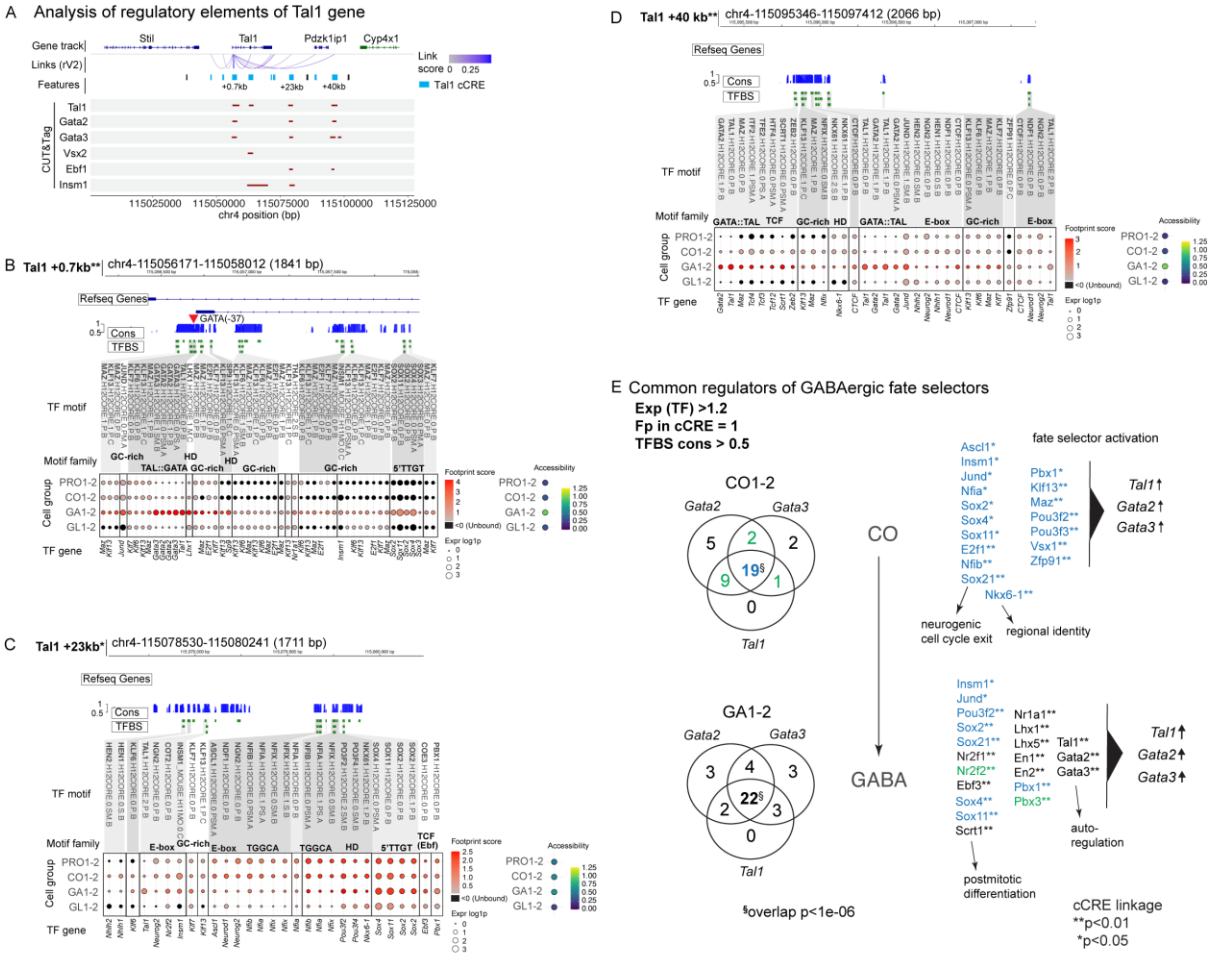


Figure 3. TF binding in putative regulatory elements of *Tal1* and the overlap between TFs interacting with other GABAergic fate selector genes.

A. ATAC features within +/-50 kb of the *Tal1* gene. *Tal1* cCREs are shown in blue. CUT&Tag, CUT&Tag consensus peaks indicating Tal1, Gata2, Gata3, Vsx2, Ebf1, Insm1, and Tead2 binding in the E12.5 mouse r1.

B-D. Footprinting based TF binding activity in the features at +1 kb, +23 kb and +40 kb of the *Tal1* TSS. Feature linkage p to *Tal1* is indicated by asterisks. Footprint scores at conserved TFBSs are shown for progenitors (PRO1-2), common precursors (CO1-2), GABAergic precursors (GA1-2) and glutamatergic precursors (GL1-2). In each dotplot, the strength of footprint at TFBS in the feature is shown as colour (Footprint score, average of cell group) and the expression of the TF gene in dot size (log1p). Average feature accessibility in cell groups is shown at the right. TFBS names (Hocomoco v12) are shown at the top and the TF gene names (mouse) are shown at the bottom of the dotplot.

The red arrowhead in (B) indicates the conserved Gata2 TFBS at –37 bp position required for the neural expression of Tal1 (see also Supplementary Table 4).

E. Overlap of the TFs with footprints in the cCREs of *Tal1*, *Gata2*, and *Gata3* in the common precursors of rV2 lineages (CO1-2) and in the rV2 GABAergic precursors (GA1-2). Venn diagrams show the number of TFs with a footprint detected in *Tal1*, *Gata2*, or *Gata3* -linked features (Fp in cCRE = 1) and with their gene expression > 1.2 (log1p) in the analysed cell group. The TFs associated with the cCREs of all three selector genes in common precursors (CO1-2) and GABAergic precursors (GA1-2) are listed. Blue text: 19 TFs interact with cCREs of *Tal1*, *Gata2* and *Gata3* in the CO1-2 cells. Some of these TFs continue to be expressed and interact with the *Tal1*, *Gata2* and *Gata3* genes in the GA1-2 cells. 22 TFs interact with *Tal1*, *Gata2* and *Gata3* in GA1-2 cells. Green text, TFs found associated with two selector genes in CO1-2 (green in the Venn diagram of CO1-2) and associated with all three selector genes in the GA1-2. Black text, TFs co-regulating the selector TFs in GA1-2 and not expressed in the CO1-2 cells.

[§] The probability of finding n overlapping genes considering all mouse genes equally is p<1e-6.

*,** The collective minimum statistical significance of feature to gene links for selector genes *Tal1*, *Gata2* and *Gata3* cCREs for the given TF is shown as: *p-value<0.05; **p-value<0.01 (with LinkPeaks z-score above 2 or below -2).

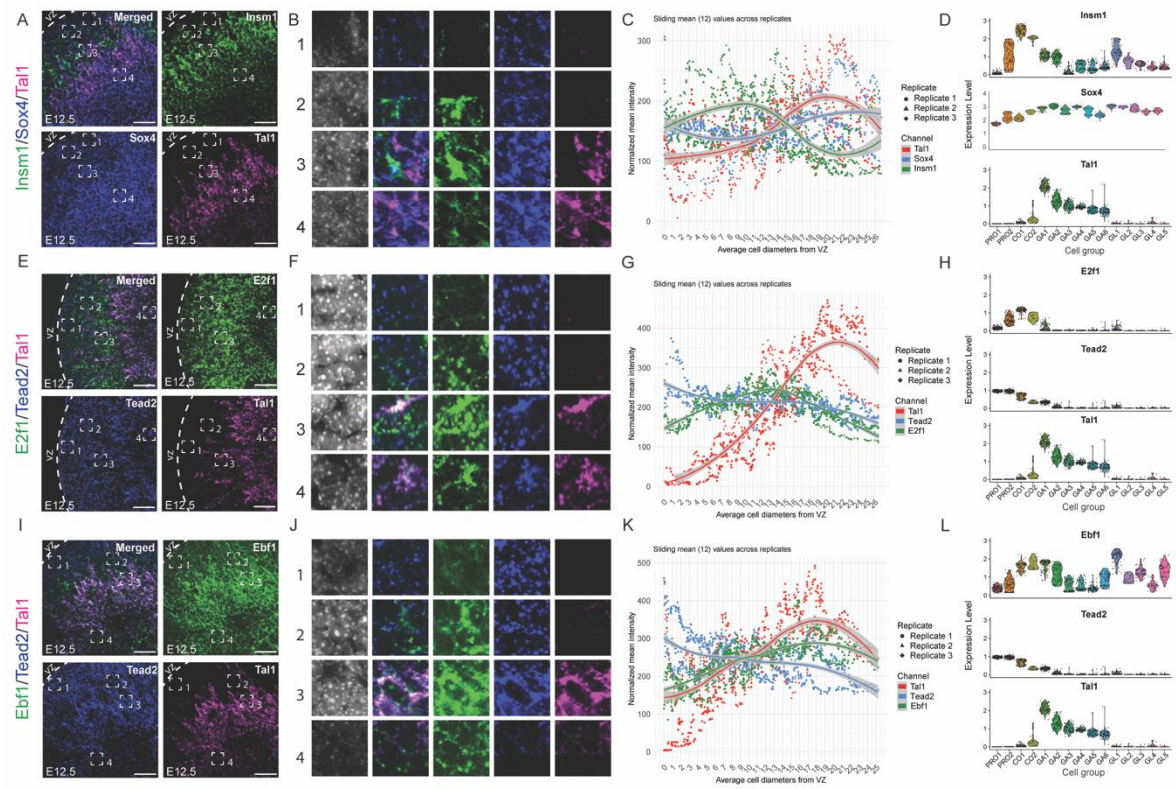


Figure 4. Expression patterns of TFs associated with Tal1 cCREs in the developing anterior brainstem.

A. RNAscope® *In Situ* Hybridization with *Insm1*, *Sox4*, and *Tal1* RNA probes on paraffin sections of E12.5 wild-type mouse embryos. Transversal sections of r1 are shown. Dashed

lines indicate the ventricular zone (VZ) border. Numbered boxes indicate areas shown on the close-up images in (B), numbering order is from apical to basal. Scale bar: 50 μm.

B. Close-up images showing co-expression of *Insm1*, *Sox4*, and *Tal1* with DAPI.

C. Signal intensity of *Insm1*, *Sox4*, and *Tal1* probes along the apical-basal axis. Average intensity line is based on three replicates. The 95% confidence interval of fluorescence intensity is highlighted in gray. The y-axis shows average fluorescence intensity per cell, and the x-axis shows the number of average cell diameters (μm) from the VZ.

D. Violin plots showing *Insm1*, *Sox4*, and *Tal1* mRNA expression detected by single-cell RNA-seq in the rV2 lineage. Cell clusters are aligned by pseudotime order on x-axis, showing first the differentiated cell clusters of GABAergic lineage, and then glutamatergic lineage.

E. RNAscope® *In Situ* Hybridization with *E2f1*, *Tead2*, and *Tal1* RNA probes, images shown and labelled as in (A). Transversal sections of E12.5 r1 are shown. Scale bar: 50 μm.

F. Close-up images showing co-expression of *E2f1*, *Tead2*, and *Tal1* with DAPI in areas indicated in (E).

G. Signal intensity of *E2f1*, *Tead2*, and *Tal1* probes along the apical-basal axis, with the lines showing the average of three replicates.

H. Violin plots showing *E2f1*, *Tead2*, and *Tal1* mRNA expression in rV2 neuronal populations.

I. RNAscope® *In Situ* Hybridization with *Ebf1*, *Tead2*, and *Tal1* RNA probes, images shown and labelled as in (A) and (E). Transversal sections of E12.5 r1 are shown. Scale bar: 50 μm.

J. Close-up images showing co-expression of *Ebf1*, *Tead2*, and *Tal1* with DAPI in areas indicated in (I).

K. Signal intensity of *Ebf1*, *Tead2*, and *Tal1* probes along the apical-basal axis, with the lines showing the average of three replicates.

L. Violin plots showing *Ebf1*, *Tead2*, and *Tal1* mRNA expression in rV2 neuronal populations.

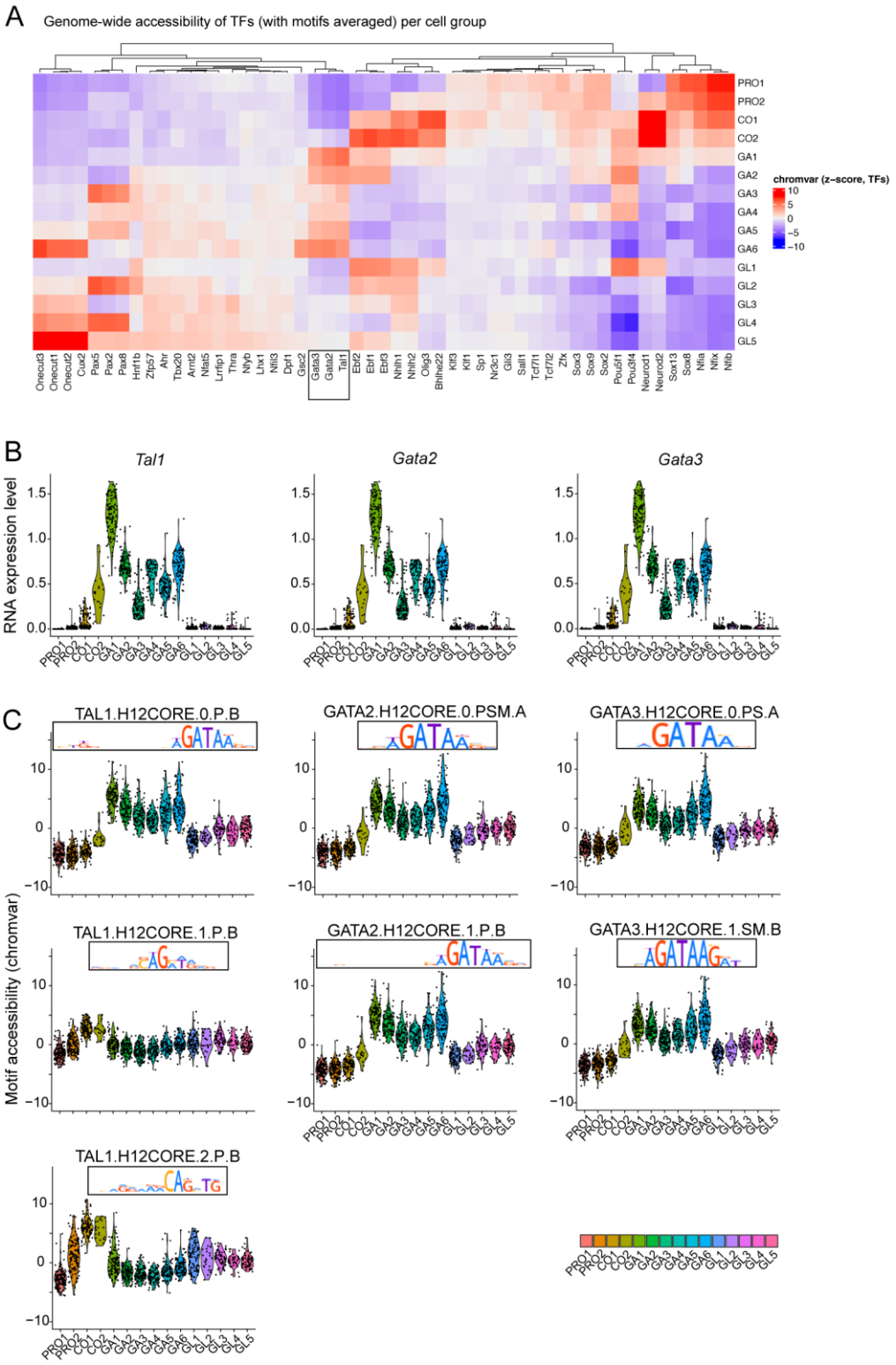
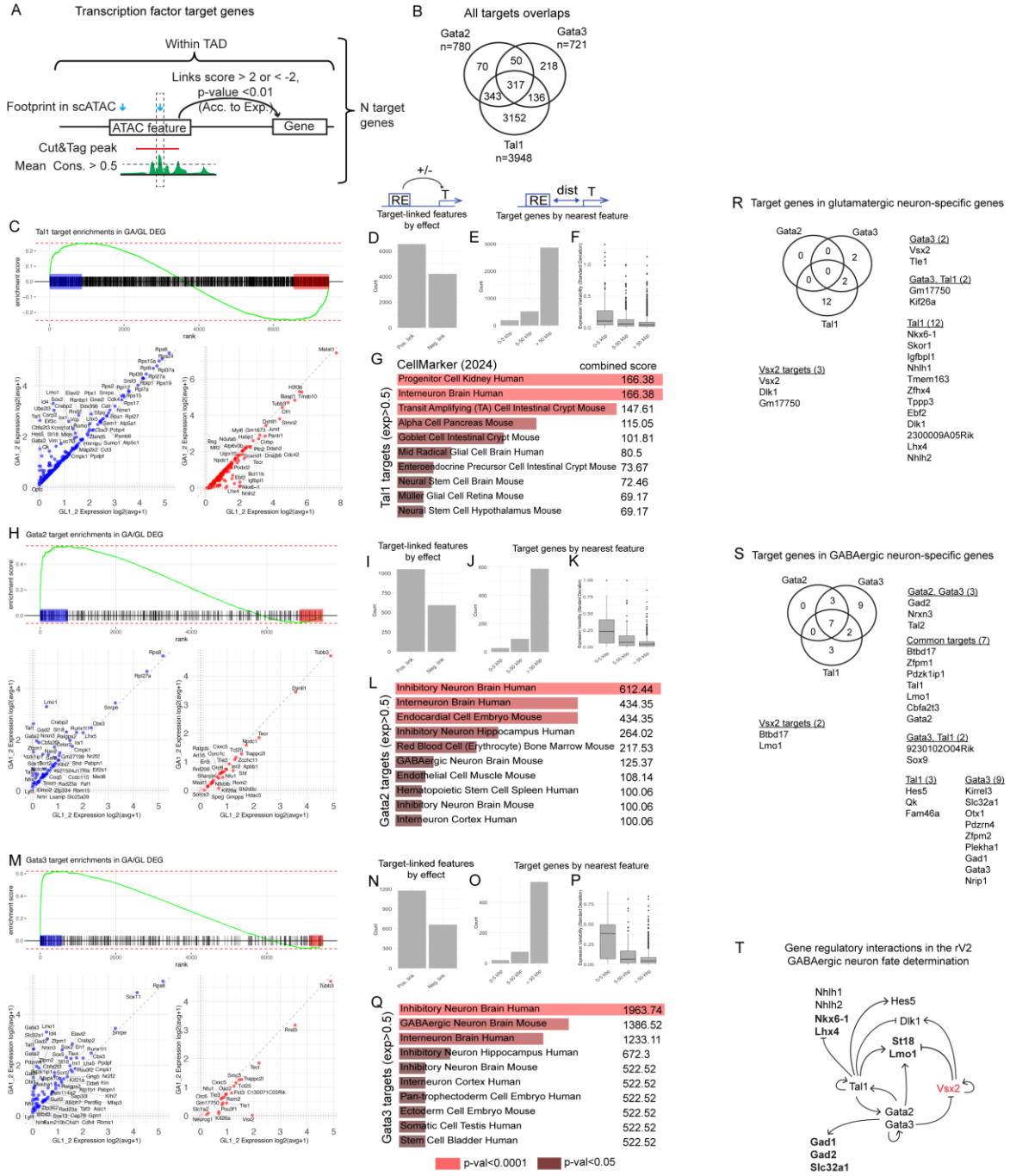


Figure 5: Dynamic accessibility of TF binding sites suggests a genome-wide function for Gata2, Tal1 and Gata3 in the selection of the GABAergic fate.

A. Genome-wide enrichment of TF binding sites in the accessible regions in GABAergic and glutamatergic lineages. Heatmap of chromVAR z-scores (avg(TFBS motifs) per TF) for the expressed TFs in rV2 cell groups. All TFs of which a TFBS-motif chromVAR score was among top 10 scores in any cell group are shown.

B. Violin plots of the TF gene expression (average scRNA expression) for *Tal1*, *Gata2*, and *Gata3;* and motif accessibility (chromVAR score) of *Tal1*, *Gata2*, and *Gata3* TF-motifs in the rV2 cell clusters. HOCOMOCO v12 contains three *Tal1* motifs, two *Gata2* motifs and two *Gata3* motifs. Ordering of cell groups is the same in all violin plots.


A
Transcription factor target genes
Within TAD
Links score > 2 or < -2,
p-value <0.01
(Acc. to Exp.)
Footprint in scATAC
ATAC feature
Gene
Cut&Tag peak
Mean Cons. > 0.5
N target genes
B
All targets overlaps
Gata2
n=780
Gata3
n=721
Tal1
n=3948
50
70
218
317
343
136
3152
+/-
RE
T
Target-linked features by effect
dist
Target genes by nearest feature
C
Tal1 target enrichments in GA/GL DEG
D
E
F
G
CellMarker (2024)
combined score
Progenitor Cell Kidney Human 166.38
Interneuron Brain Human 166.38
Transit Amplifying (TA) Cell Intestinal Crypt Mouse 147.61
Alpha Cell Pancreas Mouse 115.05
Goblet Cell Intestinal Crypt Mouse 101.81
Mid Radial Glial Cell Brain Human 80.5
Enteroendocrine Precursor Cell Intestinal Crypt Mouse 73.67
Neural Stem Cell Brain Mouse 72.46
Müller Glial Cell Retina Mouse 69.17
Neural Stem Cell Hypothalamus Mouse 69.17
Tal1 targets (exp>0.5)
R
Target genes in glutamatergic neuron-specific genes
Gata3 (2)
Vsx2
Tle1
Gata3, Tal1 (2)
Gm17750
Kif26a
Tal1 (12)
Nkx6-1
Skor1
Igfbpl1
Nhlh1
Tmem163
Zfhx4
Tppp3
Ebf2
Dlk1
2300009A05Rik
Lhx4
Nhlh2
Vsx2 targets (3)
Vsx2
Dlk1
Gm17750
H
Gata2 target enrichments in GA/GL DEG
I
J
K
L
Inhibitory Neuron Brain Human 612.44
Interneuron Brain Human 434.35
Endocardial Cell Embryo Mouse 434.35
Inhibitory Neuron Hippocampus Human 264.02
Red Blood Cell (Erythrocyte) Bone Marrow Mouse 217.53
GABAergic Neuron Brain Mouse 125.37
Endothelial Cell Muscle Mouse 108.14
Hematopoietic Stem Cell Spleen Human 100.06
Inhibitory Neuron Brain Mouse 100.06
Interneuron Cortex Human 100.06
Gata2 targets (exp>0.5)
S
Target genes in GABAergic neuron-specific genes
Gata2, Gata3 (3)
Gad2
Nrxn3
Tal2
Common targets (7)
Btbd17
Zfpm1
Pdzk1ip1
Tal1
Lmo1
Cbfa2t3
Gata2
Gata3, Tal1 (2)
9230102O04Rik
Sox9
Tal1 (3)
Hes5
Qk
Fam46a
Gata3 (9)
Kirrel3
Slc32a1
Obx1
Pdzrm4
Zfpm2
Plekha1
Gad1
Gata3
Nrip1
Vsx2 targets (2)
Btbd17
Lmo1
M
Gata3 target enrichments in GA/GL DEG
N
O
P
Q
Inhibitory Neuron Brain Human 1963.74
GABAergic Neuron Brain Mouse 1386.52
Interneuron Brain Human 1233.11
Inhibitory Neuron Hippocampus Human 672.3
Inhibitory Neuron Brain Mouse 522.52
Interneuron Cortex Human 522.52
Pan-trophectoderm Cell Embryo Human 522.52
Ectoderm Cell Embryo Mouse 522.52
Somatic Cell Testis Human 522.52
Stem Cell Bladder Human 522.52
Gata3 targets (exp>0.5)
p-val<0.0001
p-val<0.05
T
Gene regulatory interactions in the rV2 GABAergic neuron fate determination
Nhlh1
Nhlh2
Nkx6-1
Lhx4
Hes5
Dlk1
St18
Lmo1
Tal1
Vsx2
Gata2
Gata3
Gad1
Gad2
Slc32a1

936

Figure 6. Genomic targets of the GABAergic selector TFs.

A. Schematic explaining the strategy of identifying the targets of *Tal1*, *Gata2* and *Gata3* selector TFs. Within the TAD containing a gene, a feature overlapping a CUT&Tag peak for the selector TF, and a footprint for the TF at a position with weighted mean conservation score > 0.5 are found. Genes linked to features fulfilling these conditions are considered target genes. For linkage, the Spearman correlation-based LinkPeaks score between the feature targeted by the selector TF and expression of the gene is required to be >2 (positive effect link) or <-2 (negative effect link) with a p-value <0.01.

B. Number of target genes and the overlap between *Gata2*, *Gata3* and *Tal1* target genes.

C. GSEA of *Tal1* targets. Mouse genes are ranked by the difference in the expression in GA1-2 vs GL1-2 cell groups (log2 avg FC). Tal1 target genes are indicated with black lines. In leading edges, the GA1-2-enriched target genes are highlighted in blue and GL1-2-enriched genes with red. Scatterplots show the expression of the target genes in both edges.

D-F. Characterization of *Tal1* target genes.

D. Count of target features by the positive and negative effect link.

E. Count of target genes by the nearest linked feature, stratified by the nearest feature distance bins as indicated.

F. The variability of target gene expression in rV2 lineage cell clusters, stratified by the nearest feature distance bins.

G. Top terms in CellMarker gene set database using the list of *Tal1* target genes with exp>0.5 (log1p) in GA1-2 or GL1-2 cell groups or both.

H. GSEA of Gata2 targets, as in (C).

I-L. Characterization of Gata2 target genes, as in (D-G).

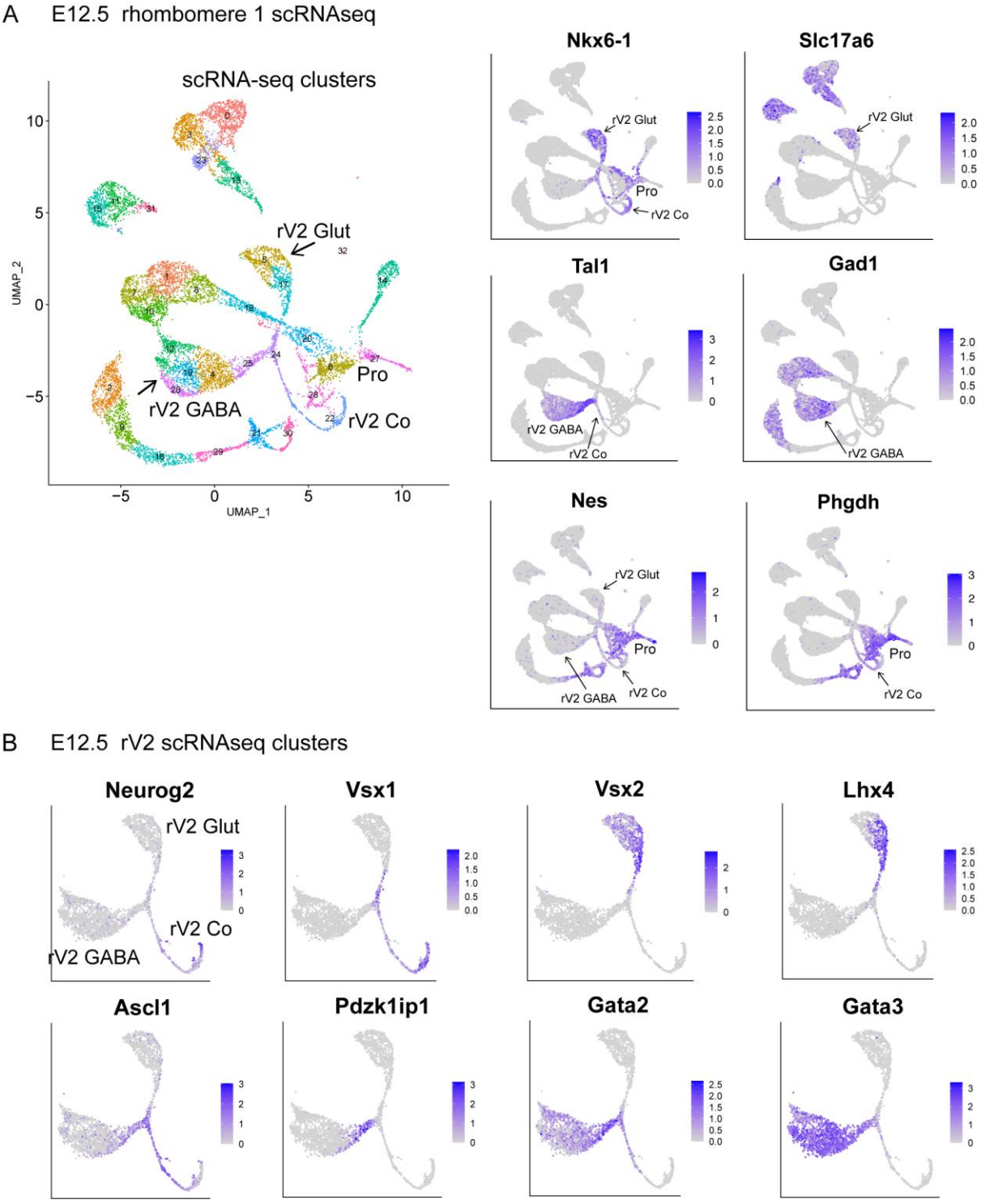
M. GSEA of Gata3 targets, as in (C).

N-Q. Characterization of Gata3 target genes, as in (D-G).

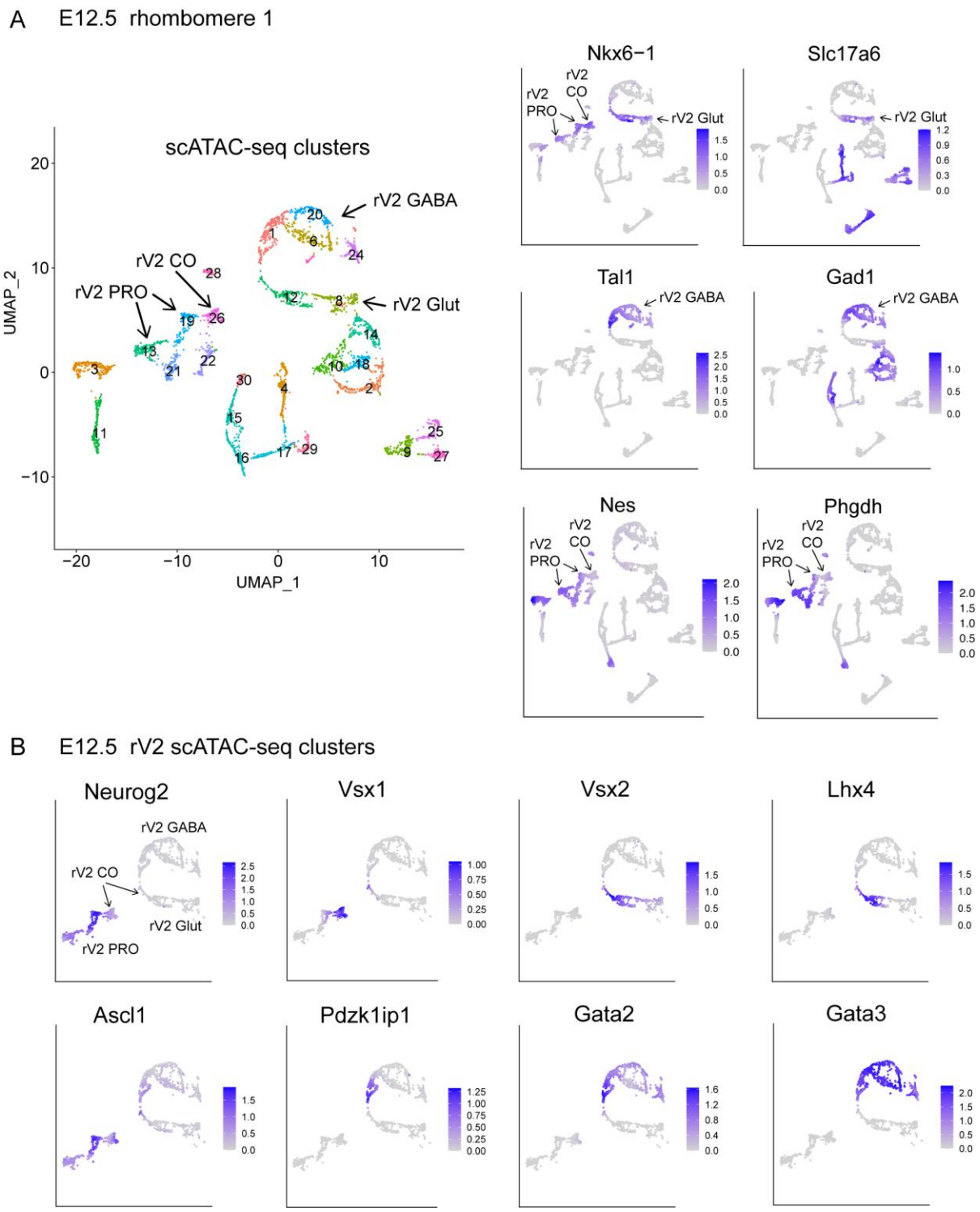
R. Venn diagram of selector TF target genes that show specificity to rV2 glutamatergic neurons (defined as avg(exp in GA)<0.5 (log1p) AND avg(exp in GL)>0.5 (log1p), and see Methods). Genes in every overlap category are listed. Vsx2 target genes that show specificity to rV2 glutamatergic neurons (n=3) are separately listed.

S. Venn diagram of selector TF target genes that show specificity to rV2 GABAergic neurons (defined as avg(exp in GL)<0.5 (log1p) AND avg(exp in GA)>0.5 (log1p)). Genes in every overlap category are listed. Vsx2 target genes that show specificity to rV2 GABAergic neurons (n=2) are separately listed.

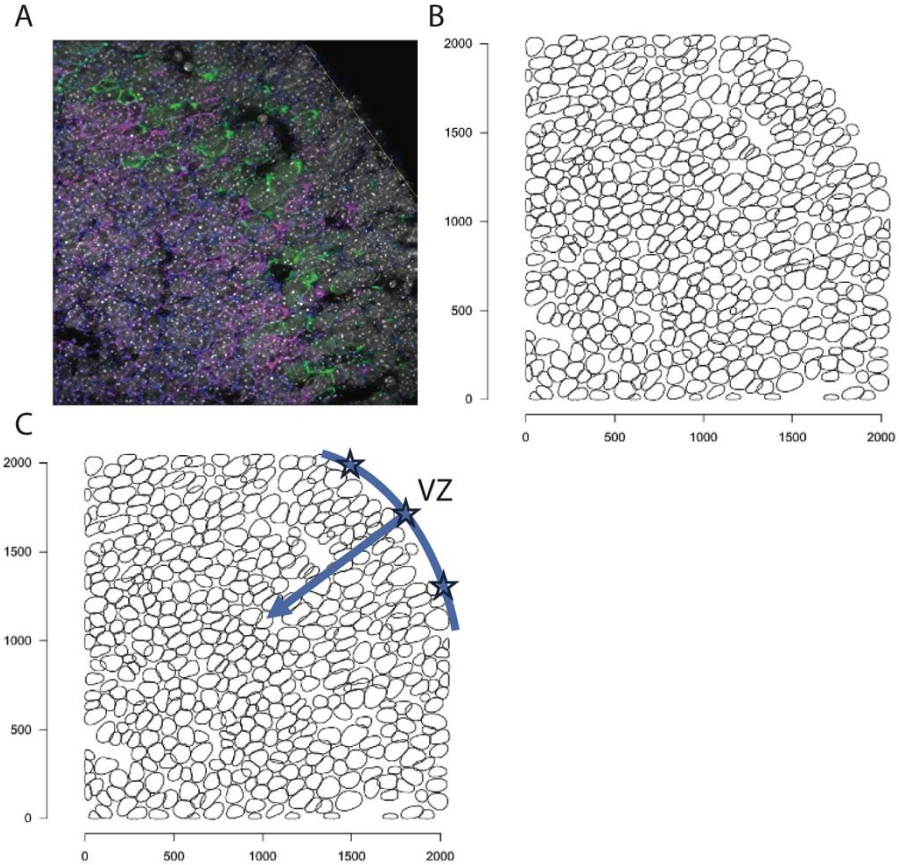
T. Proposed gene regulatory network guiding the GABA- vs glutamatergic fate selection in the rV2. Arrows represent positive regulation and were drawn when the regulator and the target expression were co-expressed in the same cell group. Blunt arrows represent negative regulation and were drawn when the regulator and the target expression were found in different cell groups (GA1-2 vs GL1-2) and not in the same cell group together.



Supplementary Figure 1. **E12.5 mouse rhombomere 1 single-cell RNAseq**.
A. UMAP plot of scRNAseq clusters (left), and the expression level of *Nkx6-1*, *Tal1*, *Slc17a6* and *Gad1* RNA in the clusters (right).
B. The expression of GABA- and glutamatergic neuron markers in the UMAP of the ventrolateral r1 lineage after lineage subsetting (E12.5 rV2 scRNAseq clusters). rV2 progenitors (rV2 pro), GABAergic neuron (rV2 GABA) and glutamatergic neuron branches (rV2 Glut) are labelled and indicated with arrows.



Supplementary Figure 2. E12.5 mouse rhombomere 1 single-cell ATACseq.
A. UMAP plot of scATACseq clusters (left), and the inferred expression level of *Nkx6-1*, *Tal1*, *Slc17a6* and *Gad1* RNA in the clusters (right).
B. The expression of GABA- and glutamatergic neuron markers in the UMAP of the ventrolateral R1 lineage clusters after lineage subsetting (E12.5 rV2 scATACseq clusters). Arrows indicate the clusters of rV2 lineage progenitors (rV2 PRO), common precursors (rV2 CO), GABAergic neuron (rV2 GABA) and glutamatergic neuron branches (rV2 Glut).

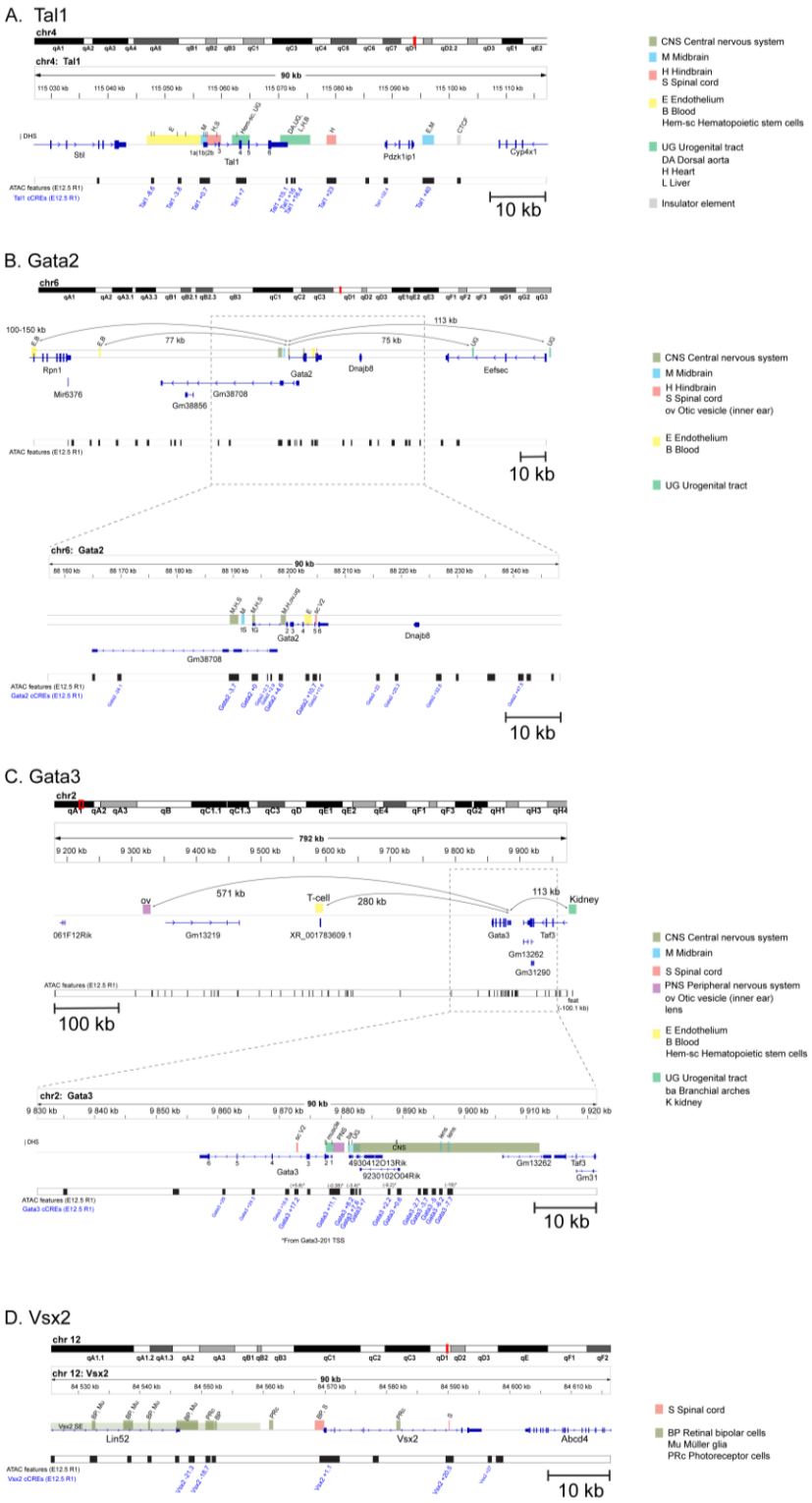


Supplementary Figure 3. RNAscope image segmentation and VZ distance calculation schematics

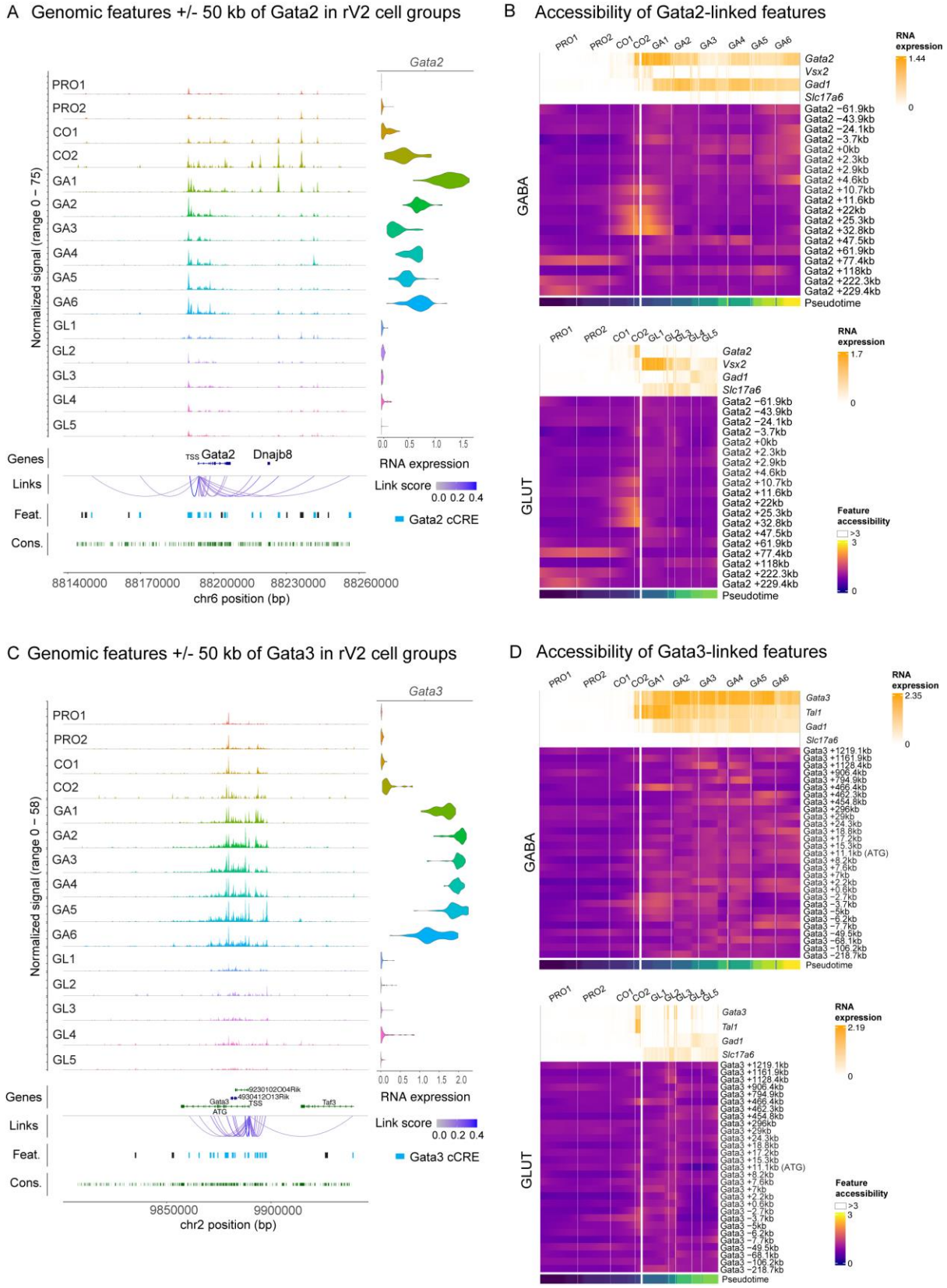
A. Typical dissection image with VZ visible on the upper right corner.

B. Same image with cells segmented with StarDist.

C. Schematics of VZ boundary distance -based ordering of the cells.



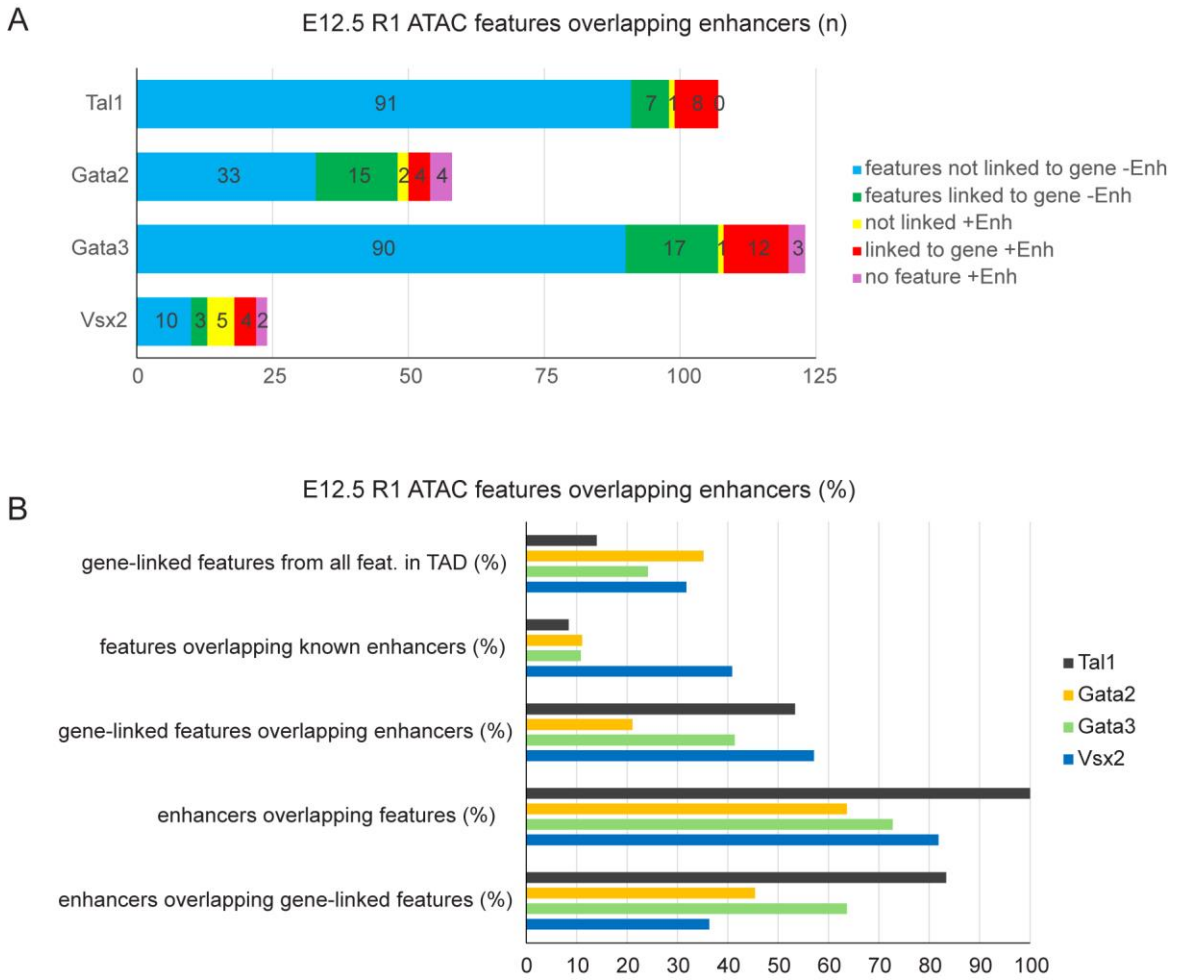
Supplementary Figure 4. Comparison of the previously characterized enhancers of *Tal1*, *Gata2*, *Gata3* and *Vsx2* with the scATAC features identified in this study. On the genomic loci of the genes, the previously characterised enhancers (details and references in the Supplementary Table 5) are indicated in colour. The features identified in this study are shown on the ATAC features (E12.5 R1) track. The gene-linked features are labelled in blue text.



Supplementary Figure 5. Genomic features and feature accessibility at *Gata2* and *Gata3* gene loci.

A. Normalized scATACseq signal in +/-50 kb region of Gata2 (A) TSS, in the rV2 lineage single-cell clusters at top. Below, the Ensembl gene models (Genes), the linkage of features to *Gata2* or *Gata3* (Links), the scATAC features (Feat.; cCREs are shown in blue), and by-

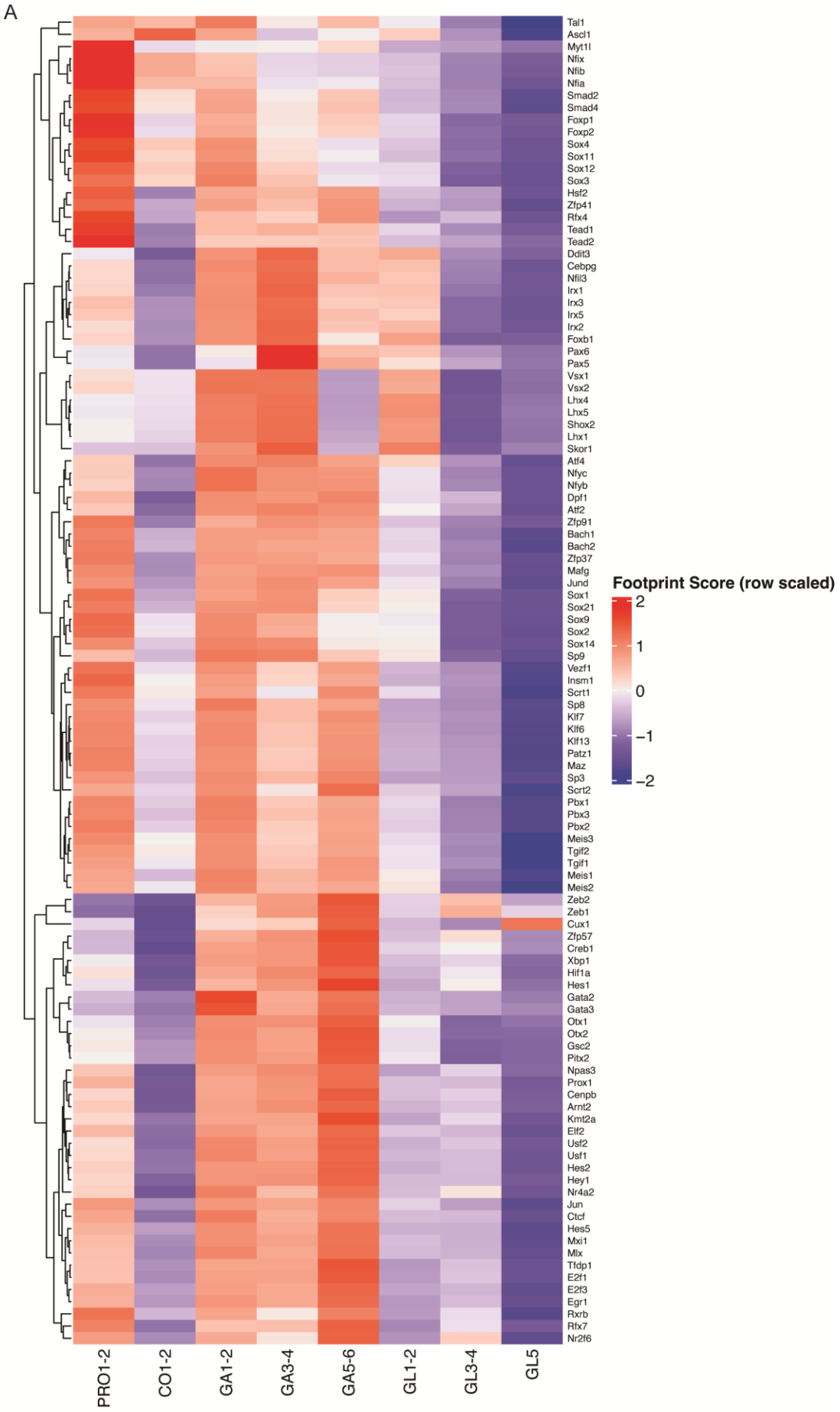
nucleotide conservation of DNA across vertebrate species (Cons.) is shown. Violin plots show the distribution of *Gata2* (A) RNA expression (log1p) in single cell clusters.
B. Smoothed heatmaps of the accessibility of the *Gata2* cCREs in the rV2 GABAergic and glutamatergic cell lineages (GABA, GLUT). Cluster identities are shown on top of heatmaps. Cells are ordered by the pseudotime value (Pseudotime). cCREs within the TAD bounds are shown on feature heatmaps. RNA expression (log1p) of *Gata2*, *Vsx2*, *Gad1* and *Slc17a6* is shown above the heatmaps.
C. scATACseq signal in the rV2 lineage single-cell clusters and genomic features in +/-50 kb region of *Gata3* TSS, similar to (A). Violin plots show the *Gata3* RNA expression (log1p) in single cells of rV2 cell clusters. *Gata3* gene ATG is located +11 kb from the TSS, the position is indicated in Genes view.
D. Smoothed heatmaps of the accessibility of the *Gata3* cCREs in the rV2 GABAergic and glutamatergic cell lineages (GABA, GLUT). RNA expression (log1p) of *Gata3*, *Tal1*, *Gad1* and *Slc17a6* is shown above the heatmaps.



Supplementary Figure 6. Comparison of the previously characterized enhancers of *Tal1*, *Gata2*, *Gata3* and *Vsx2* with the scATAC features identified in this study.
A. Stacked bar chart showing the count (n) of scATAC features overlapping enhancers and enhancers overlapping scATAC features, separated by the linkage to selector gene.

B. Bar chart showing the percentage of gene-linked features from all features in gene TAD, the percentage of scATAC features overlapping enhancers and enhancers overlapping all scATAC features and gene-linked features.

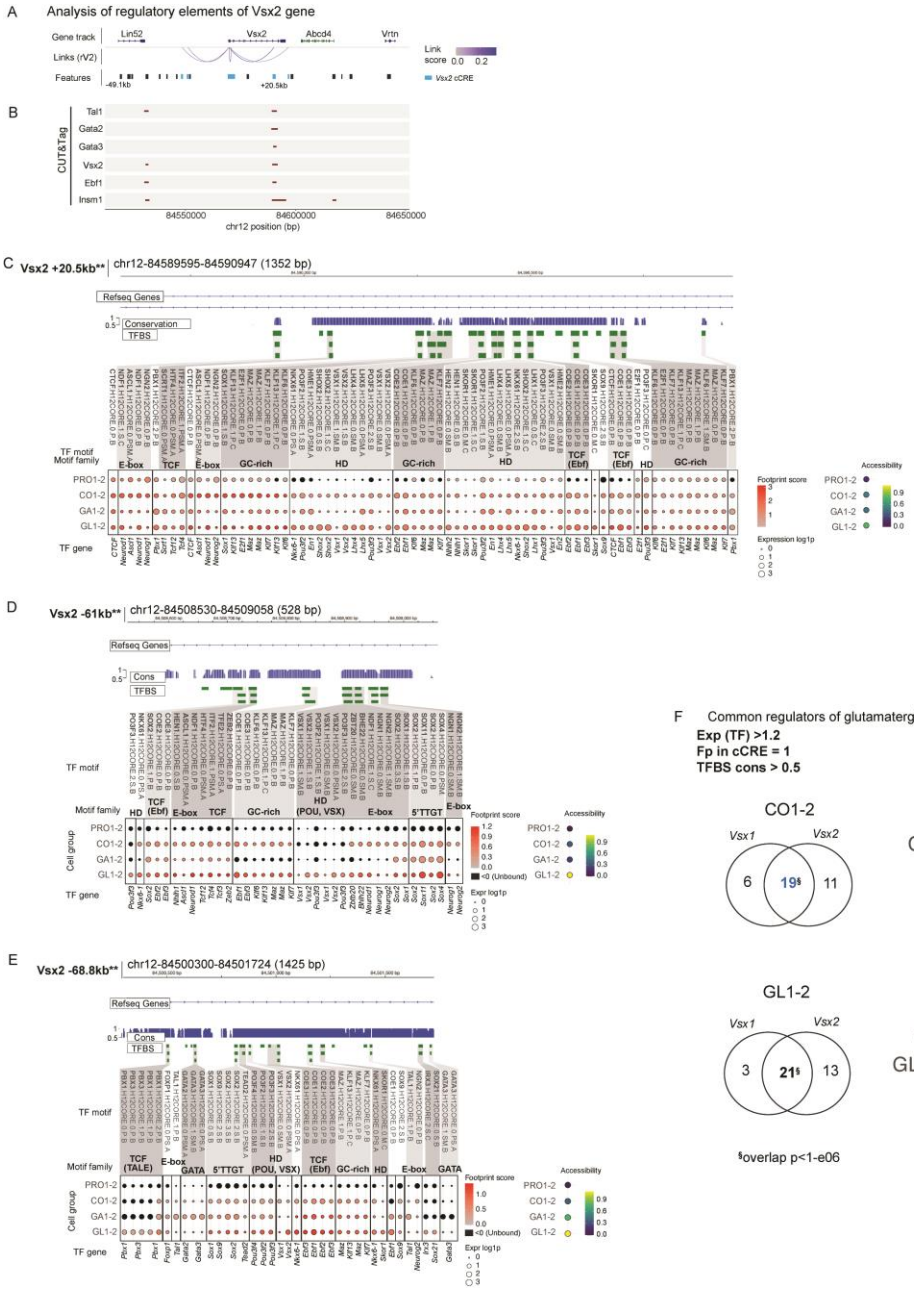
A


Tal1
Ascl1
Myt1l
Nfix
Nfib
Nfia
Smad2
Smad4
Foxp1
Foxp2
Sox4
Sox11
Sox12
Sox3
Hsf2
Zfp41
Rfx4
Tead1
Tead2
Ddit3
Cebpg
Nfil3
Irx1
Irx3
Irx5
Irx2
Foxb1
Pax6
Pax5
Vsx1
Vsx2
Lhx4
Lhx5
Shox2
Lhx1
Skor1
Atf4
Nfyc
Nfyb
Dpf1
Atf2
Zfp91
Bach1
Bach2
Zfp37
Mafg
Jund
Sox1
Sox21
Sox9
Sox2
Sox14
Sp9
Vezf1
Insm1
Scrt1
Sp8
Klf7
Klf6
Klf13
Patz1
Maz
Sp3
Scrt2
Pbx1
Pbx3
Pbx2
Meis3
Tgif2
Tgif1
Meis1
Meis2
Zeb2
Zeb1
Cux1
Zfp57
Creb1
Xbp1
Hif1a
Hes1
Gata2
Gata3
Otx1
Otx2
Gsc2
Pitx2
Npas3
Prox1
Cenpb
Arnt2
Kmt2a
Elf2
Usf2
Usf1
Hes2
Hey1
Nr4a2
Jun
Ctcf
Hes5
Mxi1
Mlx
Tfdp1
E2f1
E2f3
Egr1
Rxrb
Rfx7
Nr2f6
PRO1-2
CO1-2
GA1-2
GA3-4
GA5-6
GL1-2
GL3-4
GL5
Footprint Score (row scaled)
2
1
0
−1
−2

1041

Supplementary Figure 7. TF signatures found by scATACseq-footprinting.

A. Heatmap of the TF footprint scores (as avg(TFBS footprint scores) per TF) over the cell groups. Y-axis has been hierarchically clustered with Euclidean distance and complete linkage.



Supplementary Figure 8. Regulatory features of the glutamatergic fate selector *Vsx2* and common regulators of *Vsx2* and *Vsx1*.

A. *Vsx2* gene with its protein-coding regions and associated chromatin features.

B. CUT&Tag. Consensus peaks of CUT&Tag signal with Gata2, Gata3, Tal1, Vsx2, Ebf1, Insm1 and Tead2 antibodies in E12.5 mouse r1.

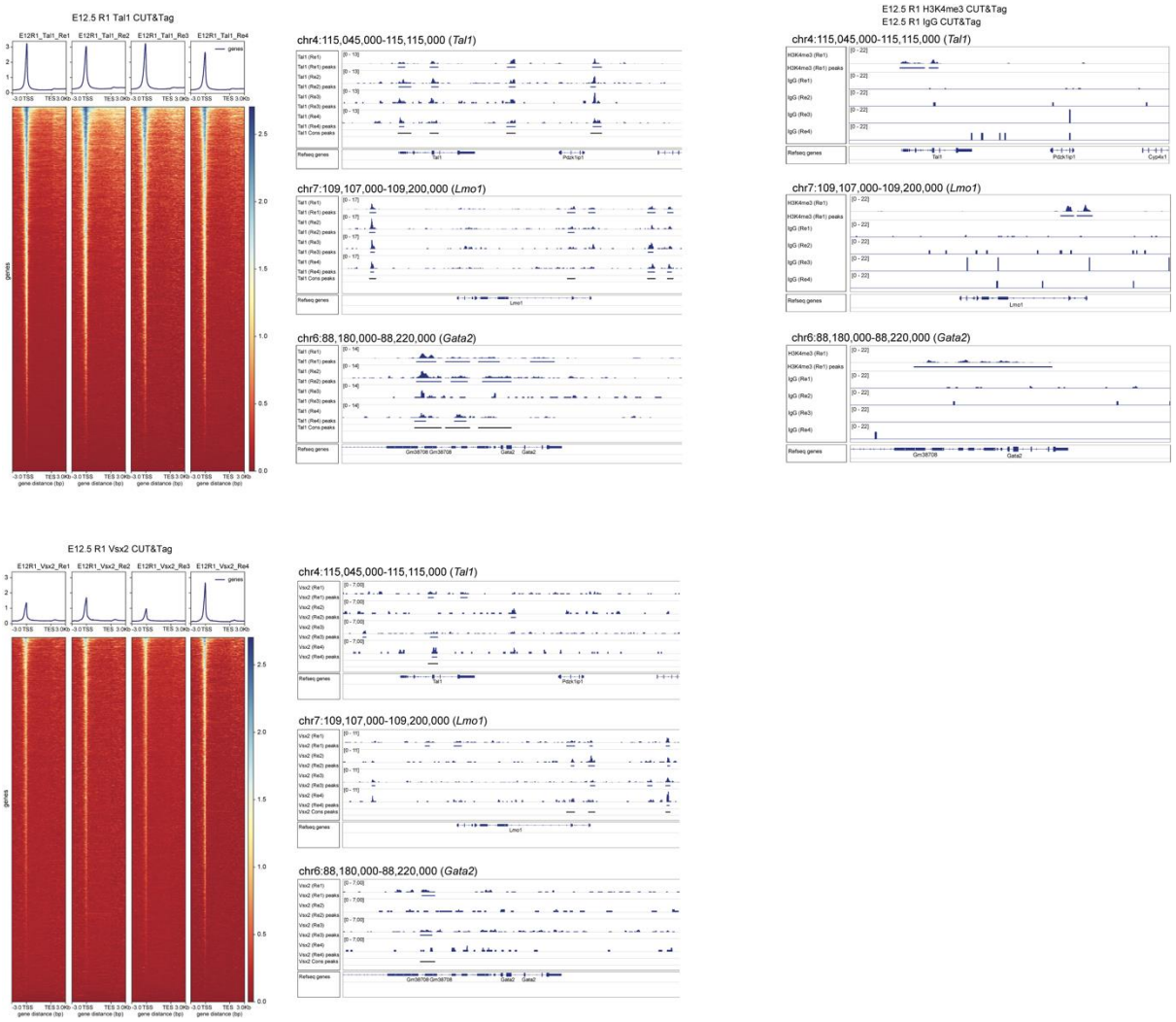
C-E. TF binding activity in the *Vsx2*-associated features at +20.5 kb, -61 kb and –68.8 kb of Vsx2 TSS according to ATAC-footprinting analysis. In each feature, the strength of footprint (averaged over the cell group shown at the side) at TFBS in the feature is shown as colour

and the expression of the TF gene in dot size. Feature accessibility in cell groups is shown at the right (Accessibility dotplot). TFBS names are shown at the top and the TF gene names (mouse genes) are shown under the dotplot.

F. Overlap of the TFs interacting with the cCREs of *Vsx1* and *Vsx2* in the common precursors of rV2 lineages (CO) and in the rV2 glutamatergic precursors (GL). Common regulators *Vsx1* and *Vsx2* in CO (n=18) and in GL (n=21) are listed next to venn diagrams. The 18 TFs regulating both *Vsx1* and *Vsx2* in CO cells are listed in blue text.

[§] The probability of finding n overlapping genes considering all mouse genes equally is p<1e-6.

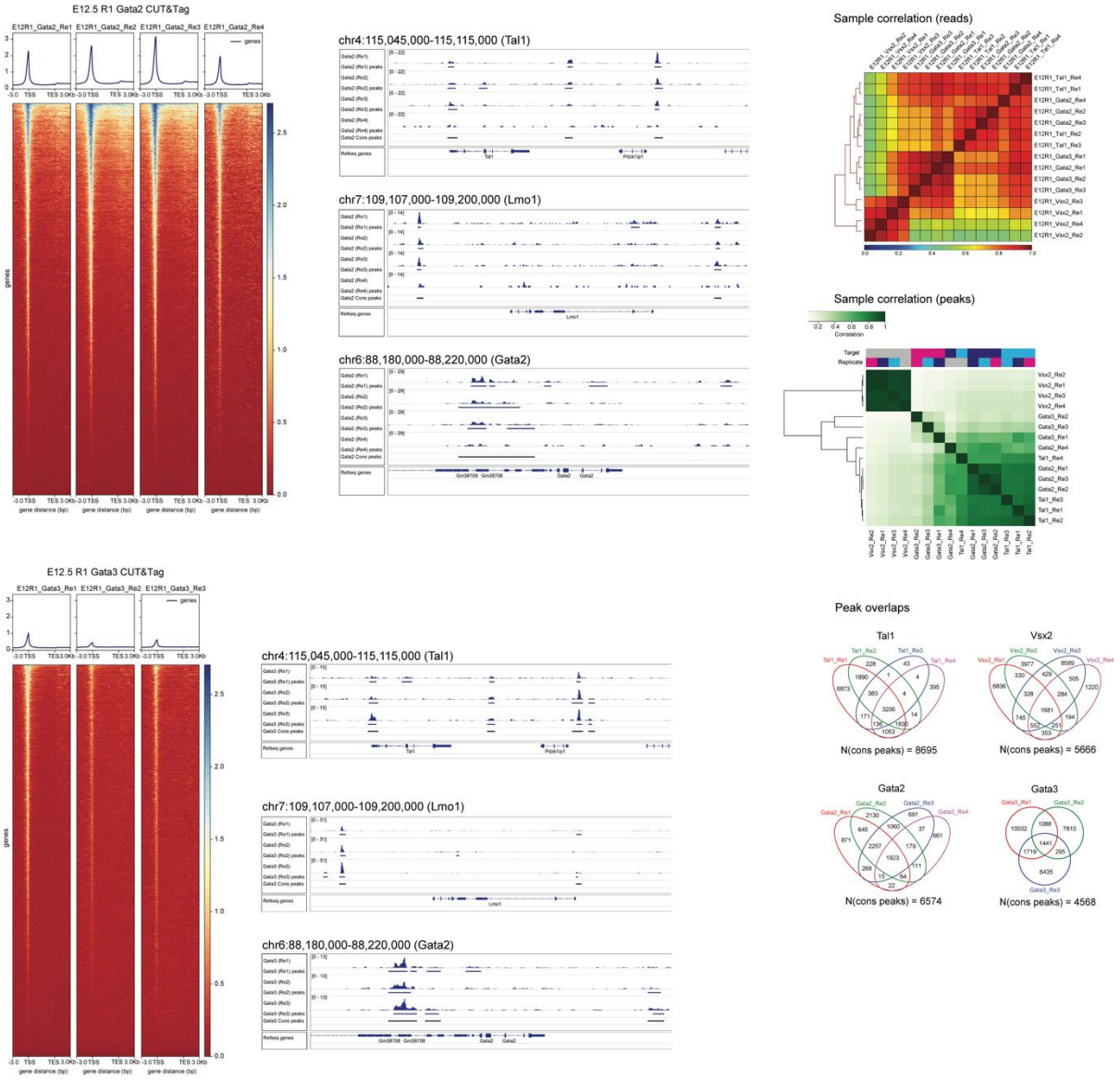
*,** The collective minimum statistical significance of feature-to-gene links for selector genes *Vsx1 and Vsx2* cCREs for the given TF is shown as: *p-value<0.05; **p-value<0.01 (with LinkPeaks z-score above 2 or below -2).



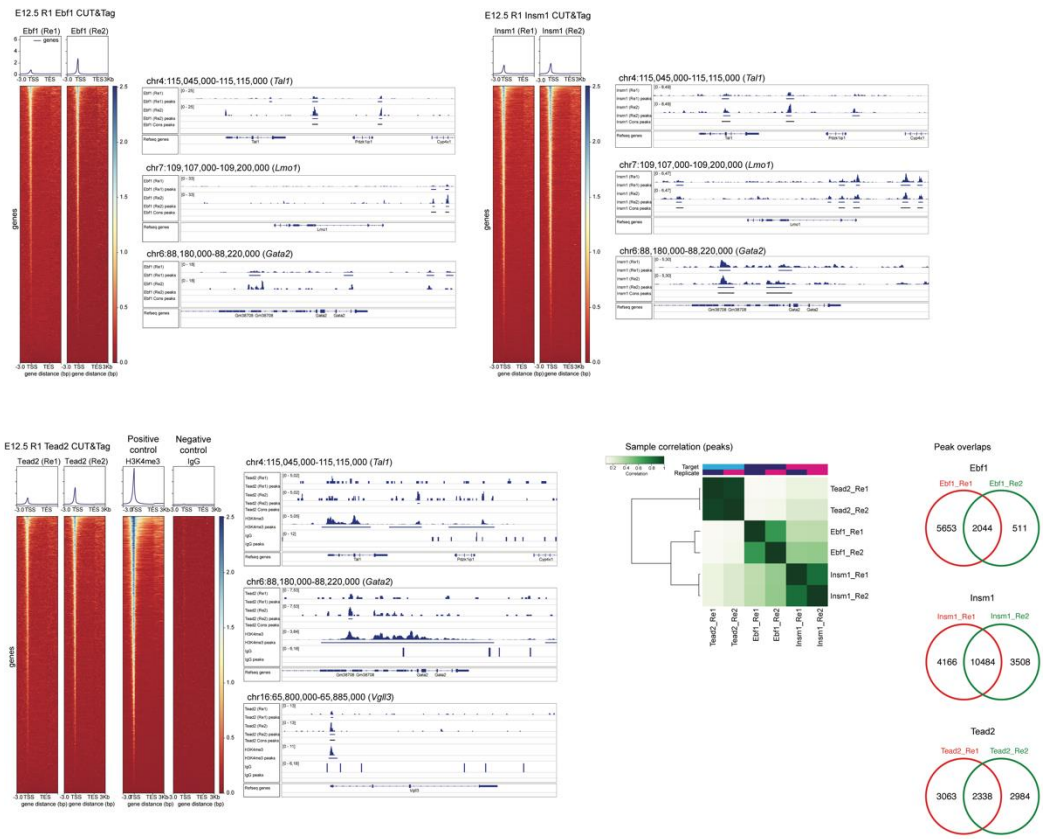
Supplementary Figure 9. CUT&Tag of *Tal1* and *Vsx2* -associated chromatin in E12.5 mouse R1, and positive and negative controls.

A. Left panel, Heatmaps and profile plots of the CUT&Tag signal intensity along the mouse genes in each replicate. Genes are scaled and the areas of 3 kb before transcription start site (TSS) and 3 kb after transcription end site (TES) are shown. Next to heatmaps, genome views

of *Tal1*, *Lmo1* and *Gata2* gene loci showing the CUT&Tag signal and detected peaks in all replicates (Re1 - Re4) and consensus peaks (Cons peaks).
B. Genome views of the CUT&Tag signal with the anti-H3K4me3 (positive control) and the anti-IgG antibody (negative control, Re1-Re4). Consensus peaks are not shown. Only two consensus peaks were detected between the Re1-Re4 of IgG treated samples.



Supplementary Figure 10. CUT&Tag of Gata2 and Gata3 -associated chromatin in E12.5 mouse R1.
Heatmaps and profile plots of the CUT&Tag signal intensity along the mouse genes (gene +/- 3kb) are shown for each replicate. Genome views of *Tal1*, *Lmo1* and *Gata2* gene loci show the CUT&Tag signal, the detected peaks in all replicates (Re1 - Re4) and consensus peaks (Cons peaks). Correlation heatmaps in the right panel show the CUT&Tag sample correlation by reads, and by peaks. Venn diagrams show the peak count and peak overlap between the replicates.



Supplementary Figure 11. Overview of the CUT&Tag for the candidate regulators of Tal1 genes in E12.5 mouse R1.

CUT&Tag of *Ebf1*, *Insm1* and *Tead2* -associated chromatin in E12.5 mouse R1, and the signal from the negative control (IgG) and positive control (H3K4me3) samples. Heatmaps and profile plots show the CUT&Tag signal intensity along the mouse gwenes (gene +/- 3kb) in each replicate. Genome views show the CUT&Tag signal and the detected peaks in *Tal1*, *Lmo1* and *Gata2* (*Ebf1*, *Insm1* CUT&Tag) or *Vgll3* (*Tead2* CUT&Tag) gene loci from all replicates (Re1 - Re2) and the defined consensus peaks (Cons peaks).
Correlation heatmaps show the CUT&Tag sample correlation by peaks. Venn diagrams show the peak count and peak overlap between the replicates.

References

Achim K, Peltopuro P, Lahti L, Li J, Salminen M, Partanen J. 2012. Distinct developmental origins and regulatory mechanisms for GABAergic neurons associated with dopaminergic nuclei in the ventral mesodiencephalic region. *Development* **139**:2360–2370. doi:10.1242/DEV.076380

Achim K, Peltopuro P, Lahti L, Tsai HH, Zachariah A, Åstrand M, Salminen M, Rowitch D, Partanen J. 2013. The role of Tal2 and Tal1 in the differentiation of midbrain GABAergic neuron precursors. *Biol Open* **2**:990–997. doi:10.1242/BIO.20135041

Achim K, Salminen M, Partanen J. 2014. Mechanisms regulating GABAergic neuron development. *Cell Mol Life Sci* **71**:1395–1415. doi:10.1007/s00018-013-1501-3

Bentsen M, Goymann P, Schultheis H, Klee K, Petrova A, Wiegandt R, Fust A, Preussner J, Kuenne C, Braun T, Kim J, Looso M. 2020. ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. *Nat Commun* **11**. doi:10.1038/S41467-020-18035-1

Bockamp EO, McLaughlin F, Göttgens B, Murrell AM, Elefanty AG, Green AR. 1997. Distinct mechanisms direct SCL/tal-1 expression in erythroid cells and CD34 positive primitive myeloid cells. *Journal of Biological Chemistry* **272**:8781–8790. doi:10.1074/jbc.272.13.8781

Boller S, Ramamoorthy S, Akbas D, Nechanitzky R, Burger L, Murr R, Schübeler D, Grosschedl R. 2016. Pioneering Activity of the C-Terminal Domain of EBF1 Shapes the Chromatin Landscape for B Cell Programming. *Immunity* **44**:527–541. doi:10.1016/J.IMMUNI.2016.02.021

Bylund J, Zhang C, Harder DR. 2002. Identification of a novel cytochrome P450, CYP4X1, with unique localization specific to the brain. *Biochem Biophys Res Commun* **296**. doi:10.1016/S0006-291X(02)00918-X

Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles G V., Clark NR, Ma'ayan A. 2013. Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**:1–14. doi:10.1186/1471-2105-14-128/FIGURES/3

Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, Garcia MU, Di Tommaso P, Nahnsen S. 2020. The nf-core framework for community-curated bioinformatics pipelines. *Nature Biotechnology 2020 38:3* **38**:276–278. doi:10.1038/s41587-020-0439-x

Faedo A, Laporta A, Segnali A, Galimberti M, Besusso D, Cesana E, Belloli S, Moresco RM, Tropiano M, Fucà E, Wild S, Bosio A, Vercelli AE, Biella G, Cattaneo E. 2017. Differentiation of human telencephalic progenitor cells into MSNs by inducible expression of Gsx2 and Ebf1. *Proc Natl Acad Sci U S A* **114**:E1234–E1242. doi:10.1073/PNAS.1611473114

Francius C, Hidalgo-Figueroa M, Debrulle S, Pelosi B, Rucchin V, Ronellenfitch K, Panayiotou E, Makrides N, Misra K, Harris A, Hassani H, Schakman O, Parras C, Xiang M, Malas S, Chow RL, Clotman F. 2016. Vsx1 transiently defines an early intermediate V2 interneuron precursor compartment in the mouse developing spinal cord. *Front Mol Neurosci* **9**:222932. doi:10.3389/FNMOL.2016.00145/BIBTEX

Garcia-Dominguez M, Poquet C, Garel S, Charnay P. 2003. Ebf gene function is required for coupling neuronal differentiation and cell cycle exit. *Development* **130**:6013–6025. doi:10.1242/DEV.00840

Göttgens B, Broccardo C, Sanchez M-J, Deveaux S, Murphy G, Göthert JR, Kotsopoulou E, Kinston S, Delaney L, Piltz S, Barton LM, Knezevic K, Erber WN, Begley CG, Frampton J, Green AR. 2004. The scl +18/19 stem cell enhancer is not required for hematopoiesis: identification of a 5' bifunctional hematopoietic-endothelial enhancer bound by Fli-1 and Elf-1. *Mol Cell Biol* **24**:1870–1883. doi:10.1128/MCB.24.5.1870-1883.2004

Hu C, Li T, Xu Y, Zhang X, Li F, Bai J, Chen J, Jiang W, Yang K, Ou Q, Li X, Wang P, Zhang Y. 2023. CellMarker 2.0: an updated database of manually curated cell

markers in human/mouse and web tools based on scRNA-seq data. *Nucleic Acids Res* **51**. doi:10.1093/nar/gkac947

Joshi K, Lee S, Lee B, Lee JW, Lee SK. 2009. LMO4 controls the balance between excitatory and inhibitory spinal V2 interneurons. *Neuron* **61**:839–851. doi:10.1016/J.NEURON.2009.02.011

Karunaratne A, Hargrave M, Poh A, Yamada T. 2002a. GATA proteins identify a novel ventral interneuron subclass in the developing chick spinal cord. *Dev Biol* **249**:30–43. doi:10.1006/dbio.2002.0754

Karunaratne A, Hargrave M, Poh A, Yamada T. 2002b. GATA Proteins Identify a Novel Ventral Interneuron Subclass in the Developing Chick Spinal Cord. doi:10.1006/dbio.2002.0754

Kaya-Okur HS, Wu SJ, Codomo CA, Pledger ES, Bryson TD, Henikoff JG, Ahmad K, Henikoff S. 2019. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nature Communications 2019 10:1* **10**:1–10. doi:10.1038/s41467-019-09982-5

Kilpinen S, Heliölä H, Achim K. 2023. Range of chromatin accessibility configurations are permissive of GABAergic fate acquisition in developing mouse brain. *BMC Genomics* **24**. doi:10.1186/s12864-023-09836-x

Kobayashi-Osaki M, Ohneda O, Suzuki N, Minegishi N, Yokomizo T, Takahashi S, Lim K-C, Engel JD, Yamamoto M. 2005. GATA motifs regulate early hematopoietic lineage-specific expression of the Gata2 gene. *Mol Cell Biol* **25**:7005–7020. doi:10.1128/MCB.25.16.7005-7020.2005

Lahti L, Haugas M, Tikker L, Airavaara M, Voutilainen MH, Anttila J, Kumar S, Inkinen C, Salminen M, Partanen J. 2016. Differentiation and molecular heterogeneity of inhibitory and excitatory neurons associated with midbrain dopaminergic nuclei. *Development (Cambridge)* **143**:516–529. doi:10.1242/DEV.129957/256855/AM/DIFFERENTIATION-AND-MOLECULAR-HETEROGENEITY-OF

Lakshmanan G, Lieuw KH, Lim K-C, Gu Y, Grosveld F, Engel JD, Karis A. 1999. Localization of Distant Urogenital System-, Central Nervous System-, and Endocardium-Specific Transcriptional Regulatory Elements in the GATA-3 Locus. *Mol Cell Biol* **19**:1558. doi:10.1128/MCB.19.2.1558

Lieuw KH, Li G long, Zhou Y, Grosveld F, Engel JD. 1997. Temporal and spatial control of murine GATA-3 transcription by promoter- proximal regulatory elements. *Dev Biol* **188**:1–16. doi:10.1006/dbio.1997.8575

Martynova E, Bouchard M, Musil LS, Cvekl A. 2018. Identification of Novel Gata3 Distal Enhancers Active in Mouse Embryonic Lens. *Dev Dyn* **247**:1186–1198. doi:10.1002/DVDY.24677

Monaghan CE, Nechiporuk T, Jeng S, McWeeney SK, Wang J, Rosenfeld MG, Mandel G. 2017. REST corepressors RCOR1 and RCOR2 and the repressor INSM1 regulate the proliferation-differentiation balance in the developing brain. *Proc Natl Acad Sci U S A* **114**:E406–E415. doi:10.1073/PNAS.1620230114

Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstråle M, Laurila E, Houstis N, Daly MJ, Patterson N, Mesirov JP, Golub TR, Tamayo P, Spiegelman B, Lander ES, Hirschhorn JN, Altshuler D, Groop LC. 2003. PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genetics 2003 34:3* **34**:267–273. doi:10.1038/ng1180

Morello F, Borshagovski D, Survila M, Tikker L, Sadik-Ogli S, Kirjavainen A, Estartús N, Knaapi L, Lahti L, Törönen P, Mazutis L, Delogu A, Salminen M, Achim K,

Partanen J. 2020a. Molecular Fingerprint and Developmental Regulation of the Tegmental GABAergic and Glutamatergic Neurons Derived from the Anterior Hindbrain. *Cell Rep* **33**:108268. doi:10.1016/J.CELREP.2020.108268

Morello F, Voikar V, Parkkinen P, Panhelainen A, Rosenholm M, Makkonen A, Rantamäki T, Piepponen P, Aitta-aho T, Partanen J. 2020b. ADHD-like behaviors caused by inactivation of a transcription factor controlling the balance of inhibitory and excitatory neuron development in the mouse anterior brainstem. *Transl Psychiatry* **10**. doi:10.1038/S41398-020-01033-8

Muroyama Y, Fujiwara Y, Orkin SH, Rowitch DH. 2005. Specification of astrocytes by bHLH protein SCL in a restricted region of the neural tube. *Nature* **438**. doi:10.1038/nature04139

Nozawa D, Suzuki N, Kobayashi-osaki M, Pan X, Engel JD, Yamamoto M. 2009. GATA2-dependent and region-specific regulation of Gata2 transcription in the mouse midbrain. *Genes Cells* **14**:569–582. doi:10.1111/J.1365-2443.2009.01289.X

Ogilvy S, Ferreira R, Piltz SG, Bowen JM, Göttgens B, Green AR. 2007. The SCL +40 enhancer targets the midbrain together with primitive and definitive hematopoiesis and is regulated by SCL and GATA proteins. *Mol Cell Biol* **27**:7206–7219. doi:10.1128/MCB.00931-07

Ono Y, Fukuhara N, Yoshie O. 1998. TAL1 and LIM-Only Proteins Synergistically Induce Retinaldehyde Dehydrogenase 2 Expression in T-Cell Acute Lymphoblastic Leukemia by Acting as Cofactors for GATA3. *Mol Cell Biol* **18**:6939–6950. doi:10.1128/MCB.18.12.6939

Radhakrishnan SK, Feliciano CS, Najmabadi F, Haegebarth A, Kandel ES, Tyner AL, Gartel AL. 2004. Constitutive expression of E2F-1 leads to p21-dependent cell cycle arrest in S phase of the cell cycle. *Oncogene 2004 23:23* **23**:4173–4176. doi:10.1038/sj.onc.1207571

Schep AN, Wu B, Buenrostro JD, Greenleaf WJ. 2017. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nature Methods 2017 14:10* **14**:975–978. doi:10.1038/nmeth.4401

Schmidt U, Weigert M, Broaddus C, Myers G. 2018. Cell detection with star-convex polygonsLecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). doi:10.1007/978-3-030-00934-2_30

Shats I, Deng M, Davidovich A, Zhang C, Kwon JS, Manandhar D, Gordân R, Yao G, You L. 2017. Expression level is a key determinant of E2F1-mediated cell fate. *Cell Death Differ* **24**:626–637. doi:10.1038/CDD.2017.12

Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LDW, Richards S, Weinstock GM, Wilson RK, Gibbs RA, Kent WJ, Miller W, Haussler D. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15**. doi:10.1101/gr.3715005

Sinclair AM, Göttgens B, Barton LM, Stanley ML, Pardanaud L, Klaine M, Gering M, Bahn S, Sanchez MJ, Bench AJ, Fordham JL, Bockamp EO, Green AR. 1999. Distinct 5’ SCL enhancers direct transcription to developing brain, spinal cord, and endothelium: Neural expression is mediated by GATA factor binding sites. *Dev Biol* **209**:128–142. doi:10.1006/dbio.1999.9236

Singh R, Berger B. 2021. Deciphering the species-level structure of topologically associating domains. *bioRxiv* 2021.10.28.466333. doi:10.1101/2021.10.28.466333

Stassen S V., Yip GGK, Wong KKY, Ho JWK, Tsia KK. 2021. Generalized and scalable trajectory inference in single-cell omics data with VIA. *Nat Commun* **12**. doi:10.1038/S41467-021-25773-3

Stuart T. n.d. Sinto. *https://github.com/timoast/sinto*.

Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, Hao Y, Stoeckius M, Smibert P, Satija R. 2019. Comprehensive Integration of Single-Cell Data. *Cell* **177**:1888-1902.e21. doi:10.1016/j.cell.2019.05.031

Stuart T, Srivastava A, Madad S, Lareau CA, Satija R. 2021. Single-cell chromatin state analysis with Signac. *Nat Methods* **18**. doi:10.1038/s41592-021-01282-5

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**:15545–15550. doi:10.1073/PNAS.0506580102

Tavano S, Taverna E, Kalebic N, Haffner C, Namba T, Dahl A, Wilsch-Bräuninger M, Paridaen JTML, Huttner WB. 2018. Insm1 Induces Neural Progenitor Delamination in Developing Neocortex via Downregulation of the Adherens Junction Belt-Specific Protein Plekha7. *Neuron* **97**:1299-1314.e8. doi:10.1016/J.NEURON.2018.01.052

Vorontsov IE, Eliseeva IA, Zinkevich A, Nikonov M, Abramov S, Boytsov A, Kamenets V, Kasianova A, Kolmykov S, Yevshin IS, Favorov A, Medvedeva YA, Jolma A, Kolpakov F, Makeev VJ, Kulakovskiy IV. 2023. HOCOMOCO in 2024: a rebuild of the curated collection of binding models for human and mouse transcription factors. *Nucleic Acids Res* gkad1077. doi:10.1093/nar/gkad1077

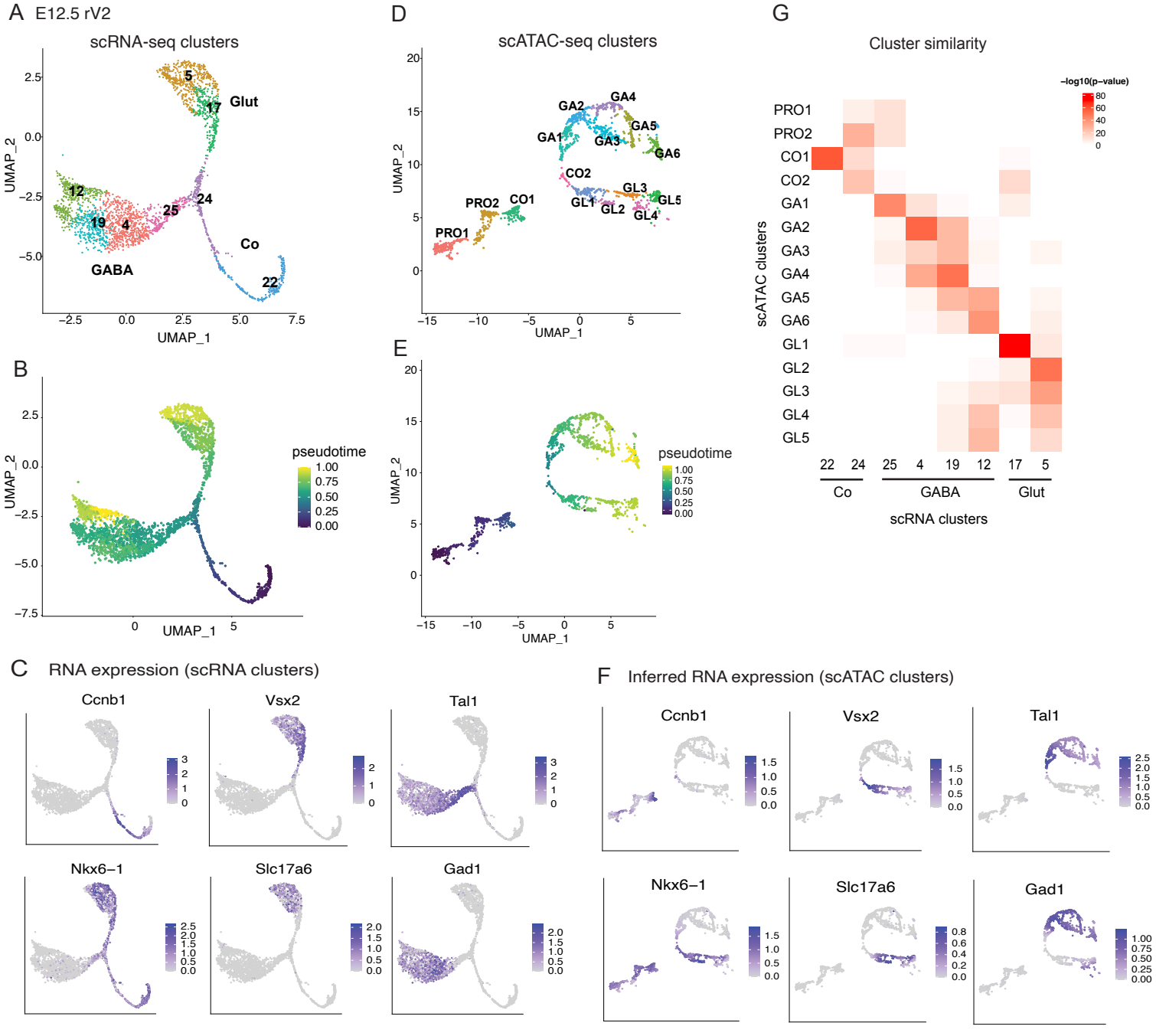
Waite MR, Skaggs K, Kaviany P, Skidmore JM, Causeret F, Martin JF, Martin DM. 2012. Distinct populations of GABAergic neurons in mouse rhombomere 1 express but do not require the homeodomain transcription factor PITX2. *Molecular and Cellular Neuroscience* **49**:32–43. doi:10.1016/J.MCN.2011.08.011

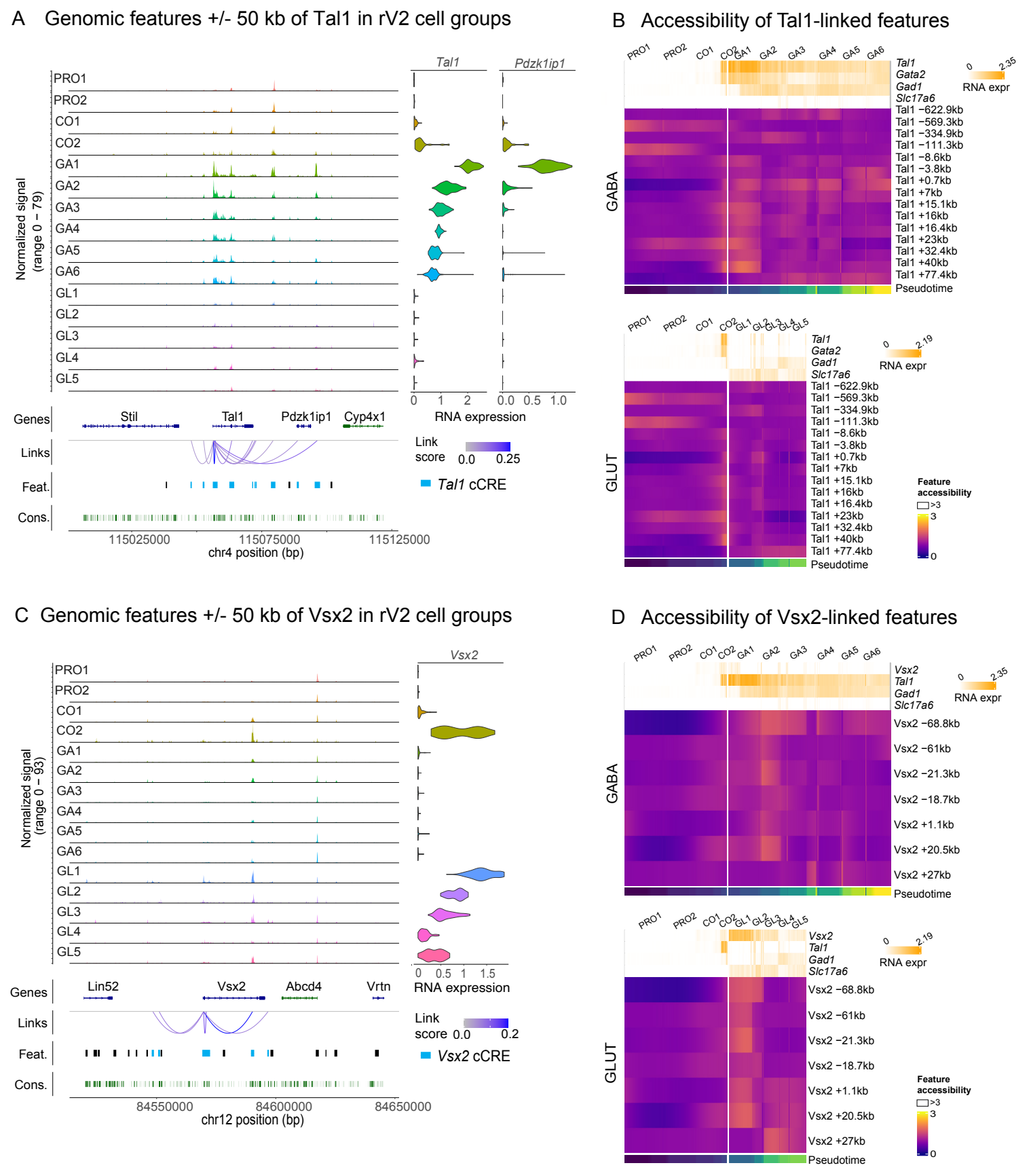
Wang L, Wang R, Herrup K. 2007. E2F1 works as a cell cycle suppressor in mature neurons. *J Neurosci* **27**:12555–12564. doi:10.1523/JNEUROSCI.3681-07.2007

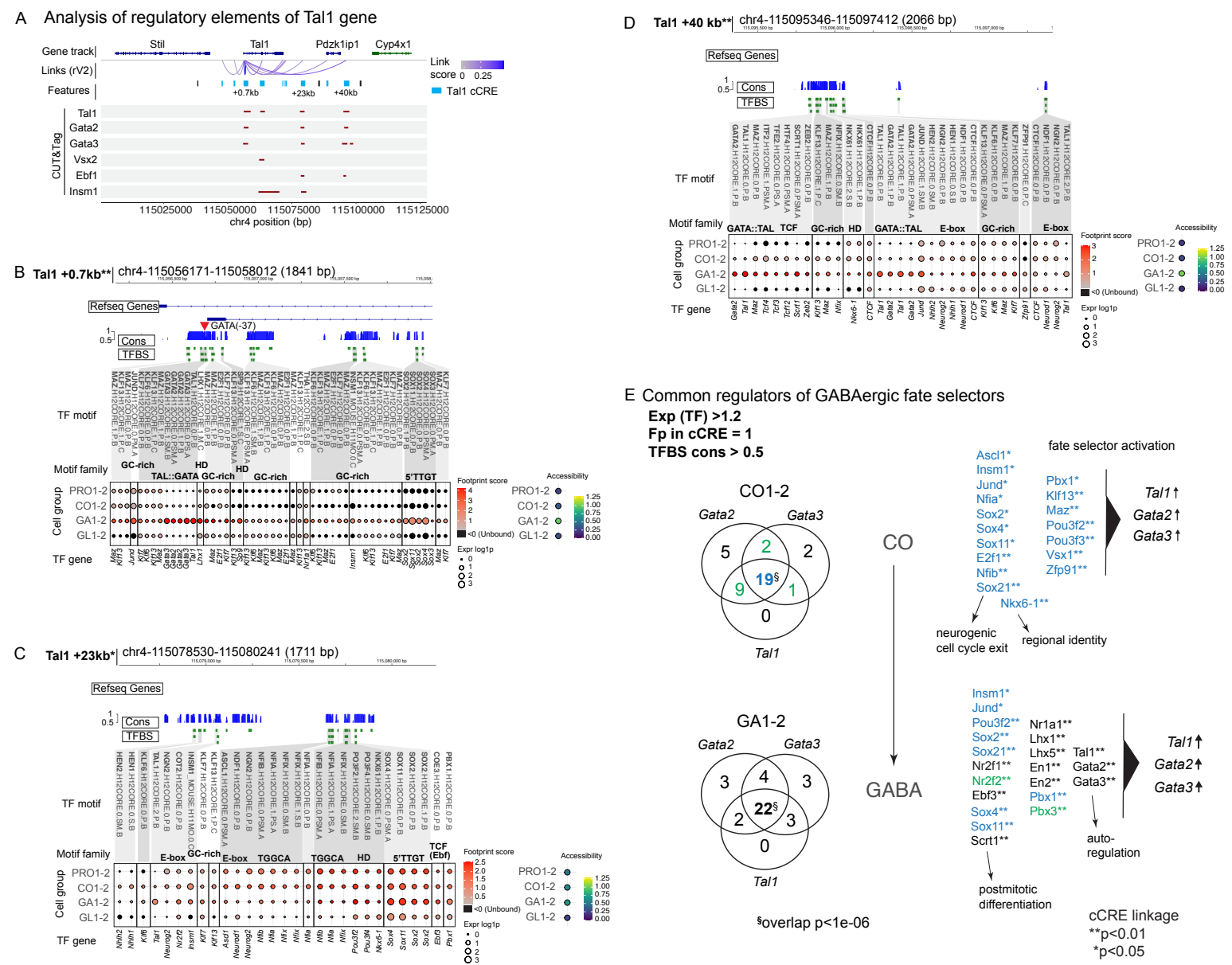
Weigert M, Schmidt U, Haase R, Sugawara K, Myers G. 2020. Star-convex polyhedra for 3D object detection and segmentation in microscopyProceedings - 2020 IEEE Winter Conference on Applications of Computer Vision, WACV 2020. doi:10.1109/WACV45572.2020.9093435

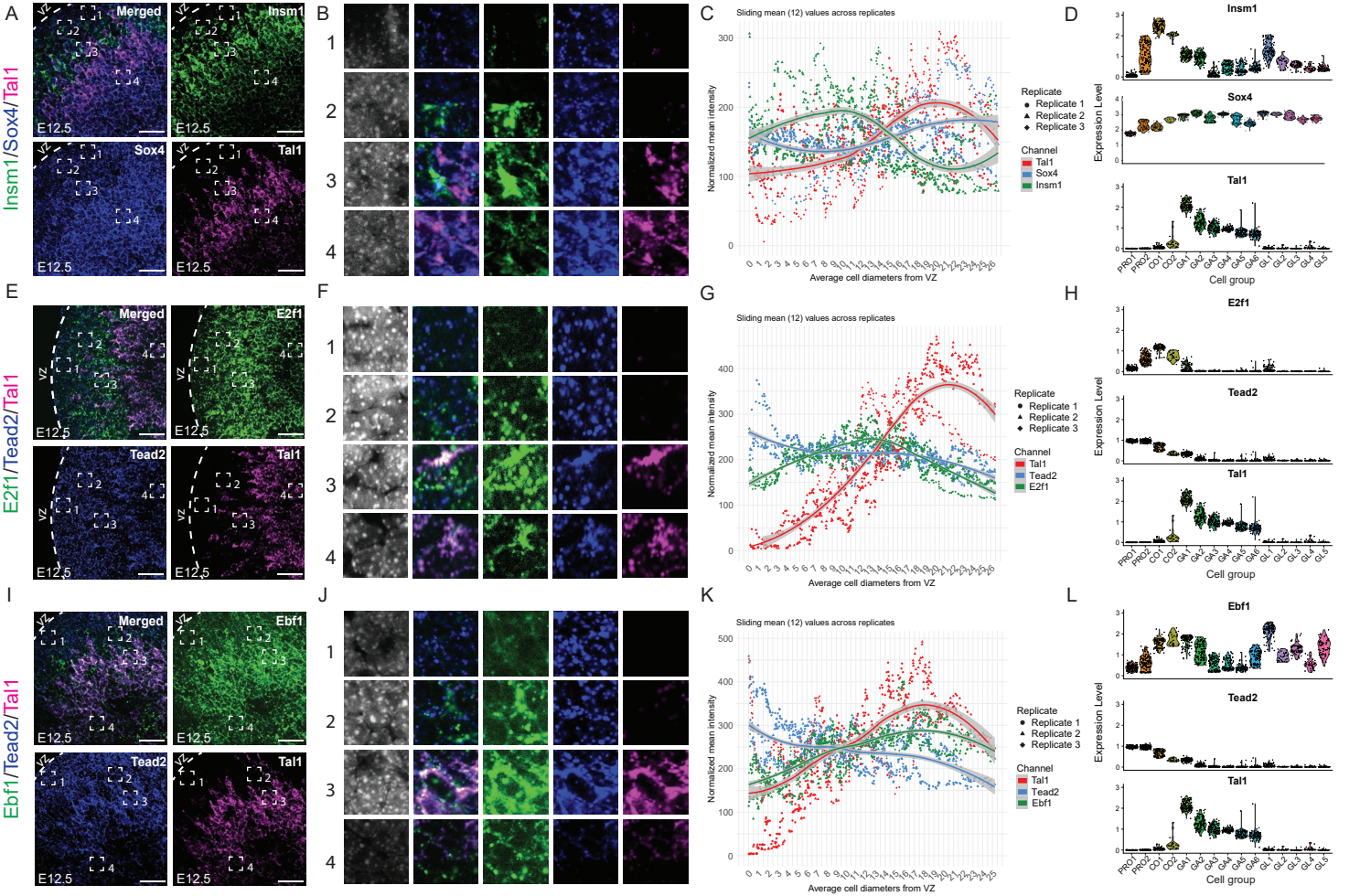
Xie Z, Bailey A, Kuleshov M V., Clarke DJB, Evangelista JE, Jenkins SL, Lachmann A, Wojciechowicz ML, Kropiwnicki E, Jagodnik KM, Jeon M, Ma'ayan A. 2021. Gene Set Knowledge Discovery with Enrichr. *Curr Protoc* **1**. doi:10.1002/CPZ1.90

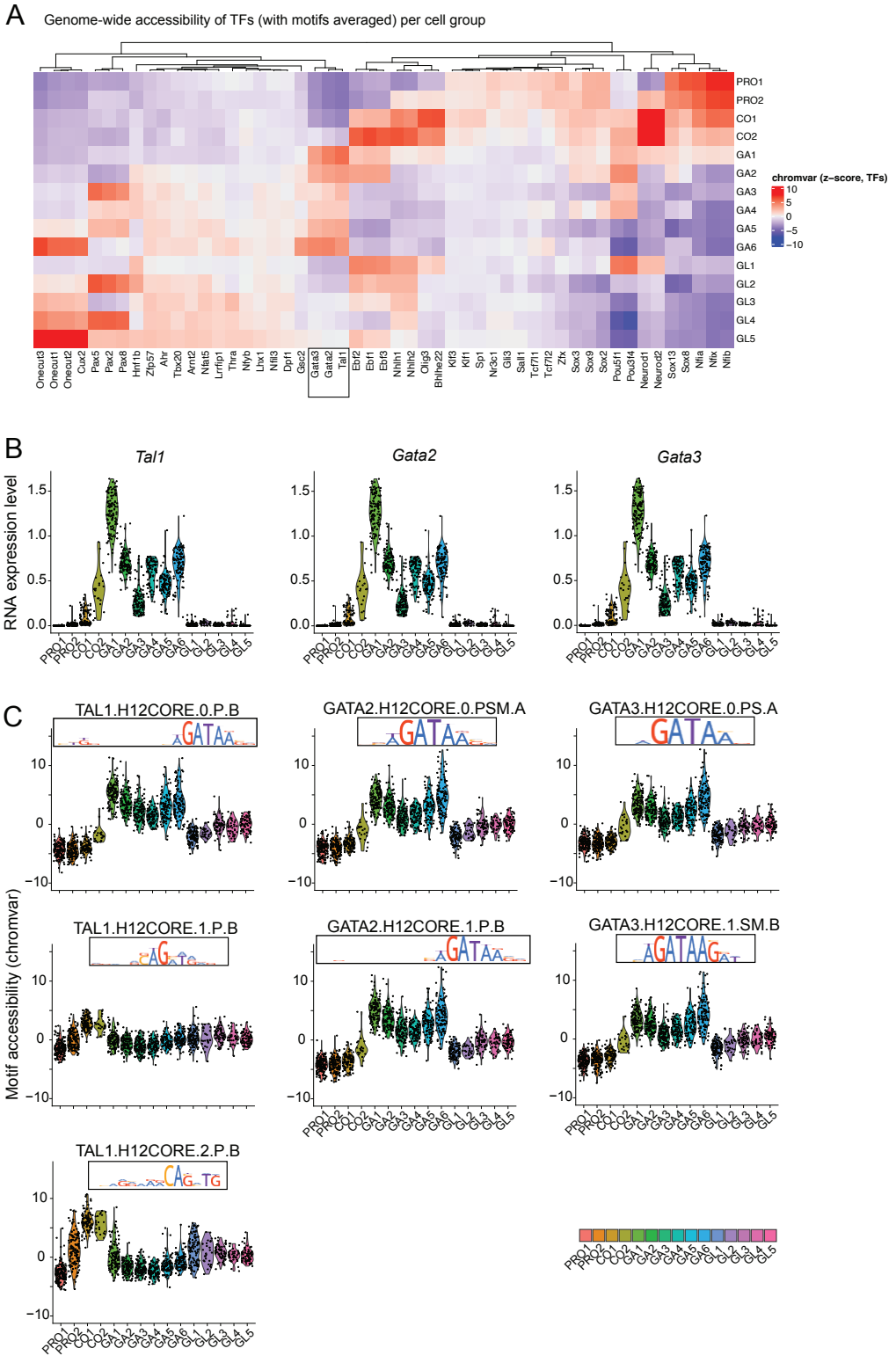
Zhang Q, Xu H, Zhao W, Zheng J, Sun L, Luo C. 2020. Zygotic Vsx1 Plays a Key Role in Defining V2a Interneuron Sub-Lineage by Directly Repressing tal1 Transcription in Zebrafish. *Int J Mol Sci* **21**. doi:10.3390/IJMS21103600

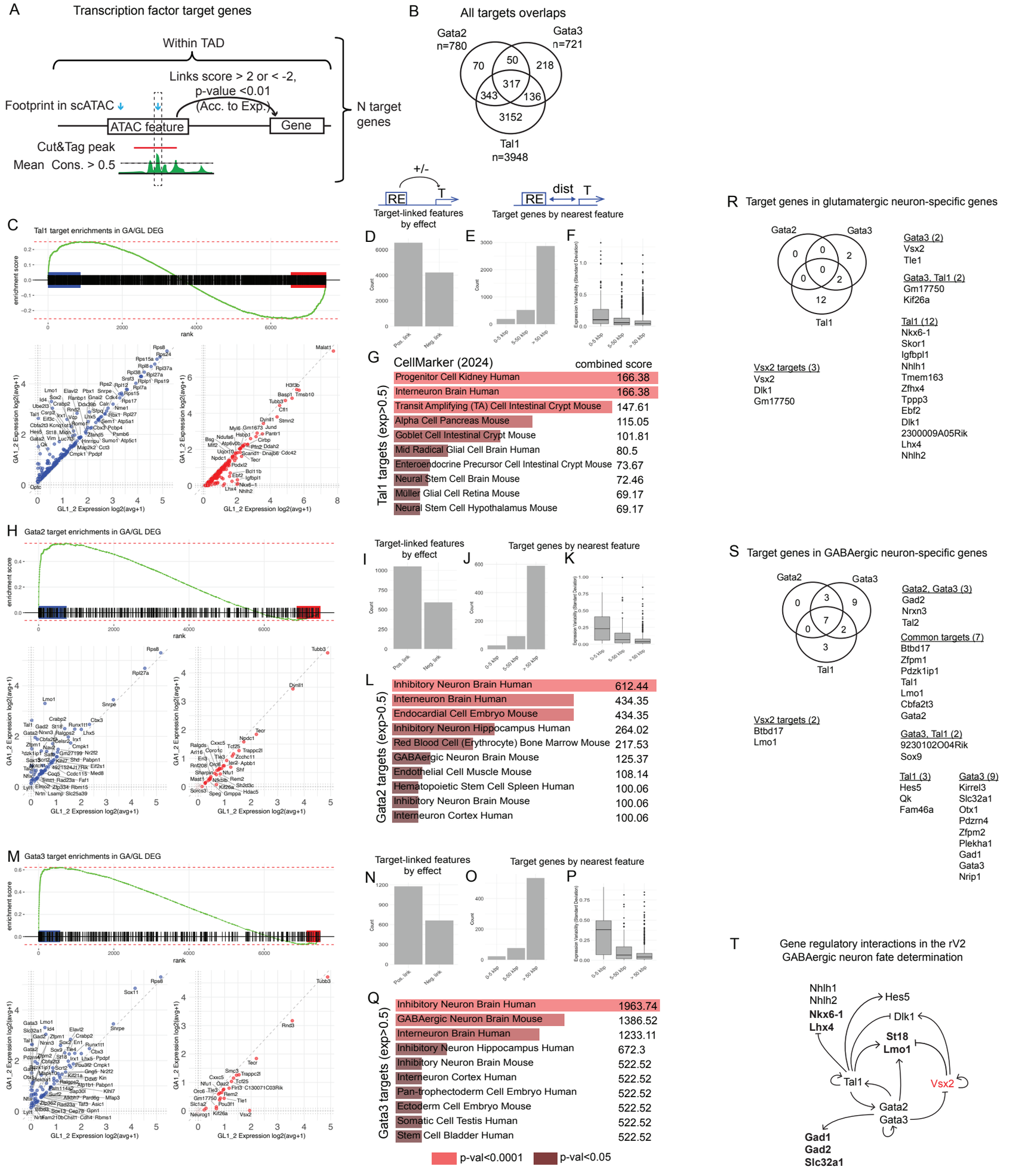

A
E12.5 rV2
scRNA-seq clusters
5
17
Glut
12
19
4
25
24
Co
GABA
22
UMAP_1
UMAP_2
B
pseudotime
1.00
0.75
0.50
0.25
0.00
D
scATAC-seq clusters
GA2
GA4
GA5
GA1
GA3
GA6
CO2
GL3
PRO2
CO1
GL1
GL2
GL5
GL4
PRO1
E
C
RNA expression (scRNA clusters)
Ccnb1
Vsx2
Tal1
Nkx6−1
Slc17a6
Gad1
F
Inferred RNA expression (scATAC clusters)
G
Cluster similarity
−log10(p−value)
scATAC clusters
scRNA clusters
Co
GABA
Glut


A Genomic features +/- 50 kb of Tal1 in rV2 cell groups
Normalized signal (range 0 – 79)
PRO1
PRO2
CO1
CO2
GA1
GA2
GA3
GA4
GA5
GA6
GL1
GL2
GL3
GL4
GL5
Tal1
Pdzk1ip1
RNA expression
Genes
Stil
Tal1
Pdzk1ip1
Cyp4x1
Links
Link score
Feat.
Tal1 cCRE
Cons.
115025000
115075000
115125000
chr4 position (bp)
B Accessibility of Tal1-linked features
GABA
GLUT
Tal1
Gata2
Gad1
Slc17a6
RNA expr
Tal1 −622.9kb
Tal1 −569.3kb
Tal1 −334.9kb
Tal1 −111.3kb
Tal1 −8.6kb
Tal1 −3.8kb
Tal1 +0.7kb
Tal1 +7kb
Tal1 +15.1kb
Tal1 +16kb
Tal1 +16.4kb
Tal1 +23kb
Tal1 +32.4kb
Tal1 +40kb
Tal1 +77.4kb
Pseudotime
Feature accessibility
C Genomic features +/- 50 kb of Vsx2 in rV2 cell groups
Normalized signal (range 0 – 93)
Vsx2
Lin52
Abcd4
Vrtn
Vsx2 cCRE
84550000
84600000
84650000
chr12 position (bp)
D Accessibility of Vsx2-linked features
Vsx2 −68.8kb
Vsx2 −61kb
Vsx2 −21.3kb
Vsx2 −18.7kb
Vsx2 +1.1kb
Vsx2 +20.5kb
Vsx2 +27kb


A Analysis of regulatory elements of Tal1 gene
Gene track
Stil
Tal1
Pdzk1ip1
Cyp4x1
Links (rV2)
Features
+0.7kb
+23kb
+40kb
Link score 0 0.25
Tal1 cCRE
CUT&Tag
Tal1
Gata2
Gata3
Vsx2
Ebf1
Insm1
115025000 115050000 115075000 115100000 115125000
chr4 position (bp)
B Tal1 +0.7kb** chr4-115056171-115058012 (1841 bp)
Refseq Genes
GATA(-37)
Cons
TFBS
TF motif
Motif family
GC-rich
HD
TAL::GATA
5'TTGT
Cell group
PRO1-2
CO1-2
GA1-2
GL1-2
TF gene
Footprint score
<0 (Unbound)
Accessibility
Expr log1p
C Tal1 +23kb* chr4-115078530-115080241 (1711 bp)
Refseq Genes
Cons
TFBS
TF motif
Motif family
E-box
GC-rich
TGGCA
HD
5'TTGT
TCF (Ebf)
Cell group
PRO1-2
CO1-2
GA1-2
GL1-2
TF gene
Footprint score
<0 (Unbound)
Accessibility
Expr log1p
D Tal1 +40 kb** chr4-115095346-115097412 (2066 bp)
Refseq Genes
Cons
TFBS
TF motif
Motif family
GATA::TAL
TCF
GC-rich
HD
GATA::TAL
E-box
GC-rich
E-box
Cell group
PRO1-2
CO1-2
GA1-2
GL1-2
TF gene
Footprint score
<0 (Unbound)
Accessibility
Expr log1p
E Common regulators of GABAergic fate selectors
Exp (TF) >1.2
Fp in cCRE = 1
TFBS cons > 0.5
CO1-2
Gata2
Gata3
Tal1
5
2
2
19§
9
1
0
CO
GA1-2
Gata2
Gata3
Tal1
3
4
3
22§
2
3
0
GABA
§overlap p<1e-06
fate selector activation
Ascl1*
Insm1*
Jund*
Nfia*
Sox2*
Sox4*
Sox11*
E2f1**
Nfib**
Sox21**
Nkx6-1**
Pbx1*
Klf13**
Maz**
Pou3f2**
Pou3f3**
Vsx1**
Zfp91**
Tal1↑
Gata2↑
Gata3↑
neurogenic cell cycle exit
regional identity
Insm1*
Jund*
Pou3f2**
Sox2**
Sox21**
Nr2f1**
Nr2f2**
Ebf3**
Sox4**
Sox11**
Scrt1**
Nr1a1**
Lhx1**
Lhx5**
En1**
En2**
Pbx1**
Pbx3**
Tal1**
Gata2**
Gata3**
Tal1↑
Gata2↑
Gata3↑
auto-regulation
postmitotic differentiation
cCRE linkage
**p<0.01
*p<0.05


A
Insm1/Sox4/Tal1
Merged
Insm1
Sox4
Tal1
E12.5
VZ
B
C
Sliding mean (12) values across replicates
Normalized mean intensity
Average cell diameters from VZ
Replicate
Replicate 1
Replicate 2
Replicate 3
Channel
Tal1
Sox4
Insm1
D
Insm1
Sox4
Tal1
Expression Level
Cell group
PRO1
PRO2
DO1
DO2
GA1
GA2
GA3
GA4
GA5
GA6
GL1
GL2
GL3
GL4
GL5
E
E2f1/Tead2/Tal1
Merged
E2f1
Tead2
Tal1
F
G
Sliding mean (12) values across replicates
Normalized mean intensity
Average cell diameters from VZ
Replicate
Replicate 1
Replicate 2
Replicate 3
Channel
Tal1
Tead2
E2f1
H
E2f1
Tead2
Tal1
Expression Level
Cell group
I
Ebf1/Tead2/Tal1
Merged
Ebf1
Tead2
Tal1
J
K
Sliding mean (12) values across replicates
Normalized mean intensity
Average cell diameters from VZ
Replicate
Replicate 1
Replicate 2
Replicate 3
Channel
Tal1
Tead2
Ebf1
L
Ebf1
Tead2
Tal1
Expression Level
Cell group


A
Genome-wide accessibility of TFs (with motifs averaged) per cell group
PRO1
PRO2
CO1
CO2
GA1
GA2
GA3
GA4
GA5
GA6
GL1
GL2
GL3
GL4
GL5
chromvar (z−score, TFs)
10
5
0
−5
−10
Onecut3
Onecut1
Onecut2
Cux2
Pax5
Pax2
Pax8
Hnf1b
Zfp57
Ahr
Tbx20
Arnt2
Nfat5
Lrrfip1
Thra
Nfyb
Lhx1
Nfil3
Dpf1
Gsc2
Gata3
Gata2
Tal1
Ebf2
Ebf1
Ebf3
Nhlh1
Nhlh2
Olig3
Bhlhe22
Klf3
Klf1
Sp1
Nr3c1
Gli3
Sall1
Tcf7l1
Tcf7l2
Zfx
Sox3
Sox9
Sox2
Pou5f1
Pou3f4
Neurod1
Neurod2
Sox13
Sox8
Nfia
Nfix
Nfib
B
Tal1
Gata2
Gata3
RNA expression level
1.5
1.0
0.5
0.0
C
TAL1.H12CORE.0.P.B
GATA2.H12CORE.0.PSM.A
GATA3.H12CORE.0.PS.A
TAL1.H12CORE.1.P.B
GATA2.H12CORE.1.P.B
GATA3.H12CORE.1.SM.B
TAL1.H12CORE.2.P.B
Motif accessibility (chromvar)
10
0
−10


A
Transcription factor target genes
Within TAD
Links score > 2 or < -2,
p-value <0.01
(Acc. to Exp.)
Footprint in scATAC
ATAC feature
Gene
Cut&Tag peak
Mean Cons. > 0.5
N target genes
B
All targets overlaps
Gata2 n=780
Gata3 n=721
Tal1 n=3948
70
50
218
317
343
136
3152
+/-
RE
T
Target-linked features by effect
dist
Target genes by nearest feature
C
Tal1 target enrichments in GA/GL DEG
enrichment score
rank
GA1_2 Expression log2(avg+1)
GL1_2 Expression log2(avg+1)
D
E
F
G
CellMarker (2024)
combined score
Tal1 targets (exp>0.5)
Progenitor Cell Kidney Human 166.38
Interneuron Brain Human 166.38
Transit Amplifying (TA) Cell Intestinal Crypt Mouse 147.61
Alpha Cell Pancreas Mouse 115.05
Goblet Cell Intestinal Crypt Mouse 101.81
Mid Radical Glial Cell Brain Human 80.5
Enteroendocrine Precursor Cell Intestinal Crypt Mouse 73.67
Neural Stem Cell Brain Mouse 72.46
Müller Glial Cell Retina Mouse 69.17
Neural Stem Cell Hypothalamus Mouse 69.17
R
Target genes in glutamatergic neuron-specific genes
Gata2
Gata3
Tal1
Gata3 (2)
Vsx2
Tle1
Gata3, Tal1 (2)
Gm17750
Kif26a
Tal1 (12)
Nkx6-1
Skor1
Igfbpl1
Nhlh1
Tmem163
Zfhx4
Tppp3
Ebf2
Dlk1
2300009A05Rik
Lhx4
Nhlh2
Vsx2 targets (3)
Vsx2
Dlk1
Gm17750
H
Gata2 target enrichments in GA/GL DEG
I
J
K
L
Gata2 targets (exp>0.5)
Inhibitory Neuron Brain Human 612.44
Interneuron Brain Human 434.35
Endocardial Cell Embryo Mouse 434.35
Inhibitory Neuron Hippocampus Human 264.02
Red Blood Cell (Erythrocyte) Bone Marrow Mouse 217.53
GABAergic Neuron Brain Mouse 125.37
Endothelial Cell Muscle Mouse 108.14
Hematopoietic Stem Cell Spleen Human 100.06
Inhibitory Neuron Brain Mouse 100.06
Interneuron Cortex Human 100.06
S
Target genes in GABAergic neuron-specific genes
Gata2
Gata3
Tal1
Gata2, Gata3 (3)
Gad2
Nrxn3
Tal2
Common targets (7)
Btbd17
Zfpm1
Pdzk1ip1
Tal1
Lmo1
Cbfa2t3
Gata2
Vsx2 targets (2)
Btbd17
Lmo1
Gata3, Tal1 (2)
9230102O04Rik
Sox9
Tal1 (3)
Hes5
Qk
Fam46a
Gata3 (9)
Kirrel3
Slc32a1
Otx1
Pdzrn4
Zfpm2
Plekha1
Gad1
Gata3
Nrip1
M
Gata3 target enrichments in GA/GL DEG
N
O
P
Q
Gata3 targets (exp>0.5)
Inhibitory Neuron Brain Human 1963.74
GABAergic Neuron Brain Mouse 1386.52
Interneuron Brain Human 1233.11
Inhibitory Neuron Hippocampus Human 672.3
Inhibitory Neuron Brain Mouse 522.52
Interneuron Cortex Human 522.52
Pan-trophectoderm Cell Embryo Human 522.52
Ectoderm Cell Embryo Mouse 522.52
Somatic Cell Testis Human 522.52
Stem Cell Bladder Human 522.52
p-val<0.0001
p-val<0.05
T
Gene regulatory interactions in the rV2 GABAergic neuron fate determination
Nhlh1
Nhlh2
Nkx6-1
Lhx4
Hes5
Dlk1
St18
Lmo1
Tal1
Vsx2
Gata2
Gata3
Gad1
Gad2
Slc32a1