



HAL
open science

Purification of astrocyte subtype based on a double reporter mice approach for downstream transcription profiling

Marion Aguirrebengoa, David Ohayon

► **To cite this version:**

Marion Aguirrebengoa, David Ohayon. Purification of astrocyte subtype based on a double reporter mice approach for downstream transcription profiling. STAR Protocols, 2021, 2 (4), pp.101009. 10.1016/j.xpro.2021.101009 . hal-03805200

HAL Id: hal-03805200

<https://hal.science/hal-03805200>

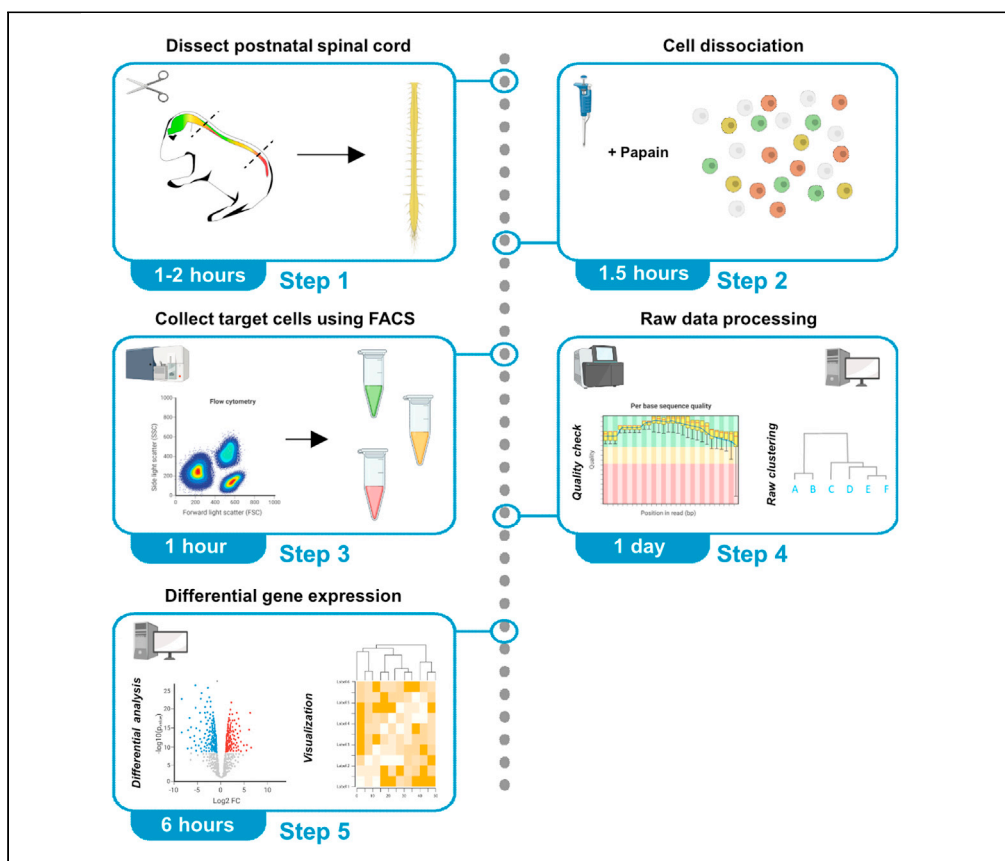
Submitted on 17 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Protocol

Purification of astrocyte subtype based on a double reporter mice approach for downstream transcription profiling



Characterizing the molecular signature of a cell subtype leads to a better understanding of cell diversity, as this molecular data can identify new cellular markers and offer insights about cell function. Here, we describe an efficient protocol to separate a subtype of astrocytes, the Olig2-AS, from other glial cells by using a double reporter mouse approach and to determine the transcriptome profile of the Olig2-AS from the postnatal spinal cord using RNA-sequencing analysis.

Marion Aguirrebengoa, David Ohayon

david-robert.ohayon@univ-tlse3.fr

Highlights
Combined enzymatic and physical dissociation of mouse postnatal spinal cord

Efficient protocol to isolate a subtype of astrocytes (Olig2-AS) from other glial cells

Two-step comparative analysis of RNA-seq databases to characterize glial subtypes

Aguirrebengoa & Ohayon, STAR Protocols 2, 101009 December 17, 2021 © 2021 The Author(s). <https://doi.org/10.1016/j.xpro.2021.101009>



Protocol

Purification of astrocyte subtype based on a double reporter mice approach for downstream transcription profiling

Marion Aguirrebengoa² and David Ohayon^{1,3,4,*}¹Molecular, Cellular and Developmental Biology Department (MCD) UMR 5077 CNRS, Centre de Biologie Intégrative (CBI), Université Paul Sabatier, 118 Route de Narbonne, 31062 Toulouse, France²BigA Core Facility, Centre de Biologie Intégrative, Université de Toulouse, 118 Route de Narbonne, 31062 Toulouse, France³Technical contact⁴Lead contact*Correspondence: david-robert.ohayon@univ-tlse3.fr
<https://doi.org/10.1016/j.xpro.2021.101009>

SUMMARY

Characterizing the molecular signature of a cell subtype leads to a better understanding of cell diversity, as this molecular data can identify new cellular markers and offer insights about cell function. Here, we describe an efficient protocol to separate a subtype of astrocytes, the Olig2-AS, from other glial cells by using a double reporter mouse approach and to determine the transcriptome profile of the Olig2-AS from the postnatal spinal cord using RNA-sequencing analysis. For complete details on the use and execution of this protocol, please refer to Ohayon et al. (2021).

BEFORE YOU BEGIN

Overview of the project

This protocol was developed to purify a subpopulation of astrocytes (Olig2-AS) in the mouse postnatal spinal cord to distinguish those cells from non-Olig2-AS, oligodendrocytes (OL) and oligodendrocyte progenitor cells (OPC). Then, we use RNA-Seq to generate transcriptome databases in order to identify the molecular signature of the Olig2-AS. This study includes an experimental procedure detailing dissection, dissociation of double reporter spinal cord and purification by FACS sorting; and a bioinformatic analysis pipeline of RNA-Seq data composed of data processing and differential gene expression analysis (Figures 1A and 1B).

Prepare double transgenic mouse P7 pups

⌚ Timing: 4 weeks (3 weeks of mouse gestation and 1 week after birth)

To fluorescently label the Olig2-AS subpopulation of astrocytes, we crossed two transgenic mouse lines: *Aldh1L1*-eGFP and *Olig2*-tdTomato. As a result, Olig2-AS express both the eGFP and tdTomato, nonOlig2-AS express only the eGFP and OPC/OL express only the tdTomato.

For this protocol, all procedures were performed in agreement with the European Community guiding principles on the care and use of animals (Scientific Procedures) Act, 1987 and approved by the national Animal Care and Ethics Committee (APAFIS#20396) following Directive 2010/63/EU.

Note: Ensure that all procedures involving mice are approved by your institutional Animal Care and Ethics Committee.



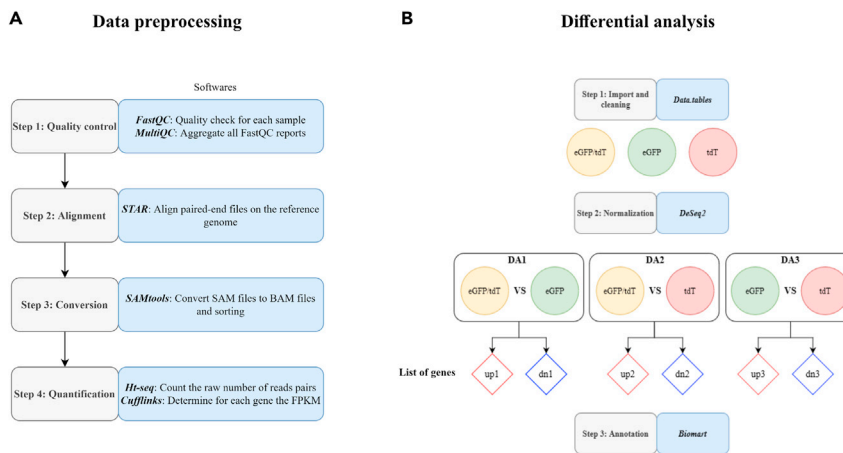


Figure 1. Bioinformatical overview of the project

(A) Flowchart of bioinformatic data processing, all steps have been run on the *Genotoul Galaxy cluster*.

(B) Flowchart describing the differential analysis on this RNA-Seq study, all steps have been run on *R (v3.6)*.

Note: Always prepare negative samples (non-transgenic animals) and single-reporter control samples expressing each fluorescent reporter separately. These are necessary to establish gating parameters on the fluorescence-activated cell sorter (FACS machine) and adjust compensation if necessary.

Prepare and pre-warm papain solution

⌚ Timing: 10 min

To purify glial cells (astrocytes and oligodendrocytes) in the spinal cord, we combine an enzymatic digestion using the Papain dissociation system of Worthington Biochemicals and mechanical dissociation.

1. Add 5 mL of EBSS to a papain vial provided in the Worthington papain dissociation system.
2. Place the vial in 37°C water-bath for 10 min.
3. Transfer the solution in a 15 mL falcon tube.

Note: The solution should be used promptly but can be kept at room temperature (17°C–25°C) during the dissection.

Prepare DNase solution

⌚ Timing: 2 min

4. Add 500 µL of EBSS to a DNase vial provided in the Worthington papain dissociation system.

Note: The solution should be used promptly but can be kept at room temperature (17°C–25°C) during the dissection.

Prepare albumin ovomucoid inhibitor solution

⌚ Timing: 2 min

5. Add 32 mL of EBSS to the albumin ovomucoid inhibitor vial provided in the Worthington papain dissociation system. This solution can be stored at 4°C up to 4 months for later use.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
Phosphate-buffer saline (PBS)	Sigma-Aldrich	Cat#D1408
FBS	ThermoFisher Scientific	Gibco Cat#A31604-01
Glycogen	ThermoFisher Scientific	Cat#R0551
Trizol	ThermoFisher Scientific	Cat#15596026
Critical commercial assays		
Papain dissociation system	Worthington Biochemical	Cat#LK003150
Deposited data		
Raw and analyzed data	This paper and Ohayon et al., 2021	GEO: GSE158517 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE158517
Example code for this protocol	This paper	https://github.com/MarionAguirrebengoa/CombinedRNAseqAnalysis_AstrocyteSubtypeProject
Experimental models: Organisms/strains		
Mouse: Tg(Aldh111-eGFP/Rpl10a) JD130Hz	GENSAT, kindly provided by Dr. N. Rouach, Paris, FRANCE	RRID: IMSR_JAX:030247
Mouse: Olig2-tdtomato	GENSAT, Tg(Olig2-tdTomato) TH39Gsat	MGI: 5311714 http://www.informatics.jax.org/allele/MGI:5311714#imsr
Oligonucleotides		
Primer for genotyping GFP-Fw	This paper	5'-CGCACCATCTTCTTCAAGGACGAC-3'
Primer for genotyping GFP-Rev	This paper	5'-AACTCCAGCAGGACCATGTGATCG-3'
Primer for genotyping Tomato-Fw	This paper	5'-CTGTTCTGTACGGCATGG-3'
Primer for genotyping Tomato-Rev	This paper	5'-GGCATTAAAGCAGCGTATCC-3'
Software and algorithms		
STAR aligner	Dobin et al., 2012	https://github.com/alexdobin/STAR
FastQC	Andrews, 2010	RRID:SCR_014583 https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
MultiQC	Ewels et al., 2016	RRID:SCR_014982 https://multiqc.info/
HTSeq count	Anders et al., 2015	https://htseq.readthedocs.io/en/master/
DeSeq2	R-Bioconductor (Love et al., 2014)	RRID:SCR_015687 https://bioconductor.org/packages/release/bioc/html/DESeq2.html
Bioconductor (3.10)	R-Bioconductor	www.bioconductor.org/
ggplot2	CRAN (Wickham, 2016)	RRID:SCR_014601 https://ggplot2.tidyverse.org/
GPlot	CRAN(Walter et al., 2015)	https://wencke.github.io/
Cufflinks (v2.2.1)	Trapnell et al., 2010	RRID:SCR_014597 http://cole-trapnell-lab.github.io/cufflinks
biomaRt	R-Bioconductor (Durinck et al., 2005)	RRID:SCR_019214 https://bioconductor.org/packages/release/bioc/html/biomaRt.html
Ensembl Mus Musculus GRCm38.91 (mm10)	Ensembl	http://www.ensembl.org/Mus_musculus/Info/Index
SAMtools	Li et al., 2009	RRID:SCR_002105 https://www.htslib.org/
Heatmap.2 from gplots2 package	R-Bioconductor	https://rdocumentation.org/packages/gplots/versions/3.1.1
Affinity Designer	Serif	RRID:SCR_016952 https://affinity.serif.com/

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Prism8	Graphpad Software	RRID:SCR_002798 https://www.graphpad.com/
R (3.6.0)	The R foundation	https://www.r-project.org/
Other		
Tissue-chopper	Mc Ilwain	RRID:SCR_015798
Nikon AZ100 microscope	Nikon	RRID:SCR_018603
Nikon SMZ800 microscope	Nikon	RRID:SCR_020333
Water bath WB7	Memmert	n/a
40µm cell strainer	Falcon	Cat#352340
5mL round bottom polypropylene tubes	Falcon	Cat#352063
Tissue culture dish (35 mm)	Falcon	Cat#353001
FACS Aria Fusion BSL1	BD Biosciences	n/a
BGISEQ-500 RNA-seq platform	BGI	n/a

STEP-BY-STEP METHOD DETAILS

Dissection of the P7 spinal cord

⌚ Timing: 1 h depending on expertise and the number of pups to be dissected (~15 min/pup)

1. Isolate the animal's head by decapitation using scissors at the base of the skull.
2. After drenching the fur with 70 % ethanol, a small incision is made in the dorsal skin to remove the pelt from the head to the hind limbs.
3. Insert small scissors through the abdominal wall muscles and cut laterally to the spinal column in both directions.
4. Then with the scissors cut the ribs on both sides of the spinal column before removing the viscera connected on its anterior side.
5. The spinal column is then removed by making a transverse cut at the level of the femurs.
6. Once clear of soft tissues, cut the column at the level of the last rib which is a landmark of the last thoracic segment. Keep the cervical and thoracic spinal column for further isolation and remove the lumbar part.

Note: Until step 6, the dissection is performed at room temperature. Once the spinal cord is removed from the body, all the following dissection steps are performed on ice.

7. Place the cervico-thoracic column in a petri dish containing cold PBS, and under a dissecting microscope (Nikon SMZ800) start cutting the column from one end and peel off the vertebrae with a forceps little by little, revealing the spinal cord.
8. Remove dorsal root ganglion and meninges from the spinal cord explants.

Note: It is very important to carefully remove the meninges which can be identified as a translucent sheet of tissue covering the spinal cord. If the meninges are not fully removed it can interfere with the following dissociation and result in a poor number of cells purified.

9. Check the fluorescence on the spinal cord explants to confirm genotyping using a Nikon AZ100 multizoom microscope

Dissociate dissected P7 spinal cord tissue

⌚ Timing: 1.5 h

These steps describe the enzymatic treatment and physical homogenization procedures to dissociate dissected tissue pieces into single-cell solution ready for FACS isolation of targeted cells.

10. Slice the spinal cord into 100 μm sections using a tissue-chopper (Mc Ilwain).
11. Place each diced spinal cord in a 35 mm tissue culture dish.
12. Add 250 μL of the DNase solution in the falcon tube containing the papain solution.
13. Add 1 mL of the papain/DNase solution per petri dish containing one chopped spinal cord.
14. Incubate for 60 min at 37°C in a 95%O₂:5%CO₂ incubator.

△ CRITICAL: Monitor dissociation and adjust the timing accordingly. Prolonged enzymatic digestion can affect cell viability.

Note: In our context, the perinatal spinal cord, we adapted the method used to purify astrocytes from adult spinal cord in order to establish their transcriptome (Cahoy et al., 2008; Zamanian et al., 2012; Molofsky et al., 2014; Chaboub et al., 2016). This method has been used, in particular, by B. Barres and collaborators to compare transcriptomes of wild-type astrocytes versus reactive astrocytes induced by LPS or MCAO (Zamanian et al., 2012). To minimize the effect of stress, we focused on reducing the incubation time with the dissociation enzyme (60 min instead of 90 min) and the time period between dissection and cell sorting (less than 3 h).

15. Gently triturate the mixture by pipetting up and down 10–15 times using a 5 mL pipette

△ CRITICAL: The dissociation needs to be quick yet gentle to separate the cells rather than rip them apart. Also avoid creating bubbles during the pipetting as it could result in poor cell health.

16. Centrifuge the dissociated cells at 300 g for 5 min at room temperature.
17. During this spin, mix 2.7 mL of EBSS with 300 μL of ovomucoid solution in a 15 mL falcon tube and 150 μL of DNase solution (low-ovo inhibitor solution).
18. Discard supernatant and resuspend each cell pellet in 500 μL diluted DNase/ovomucoid solution (low-ovo inhibitor solution) prepared in step 17.

△ CRITICAL: Be at the centrifuge, as soon as it stops and handle 15 mL conical falcon tubes very carefully to avoid disturbing the fragile pellet. Remove the supernatant slowly using a P-1000 pipette (do not pour off) ensuring no pellet is aspirated.

19. Add 1 mL of ovomucoid in a 15 mL falcon tube and layer the cell suspension on top.
20. Centrifuge the dissociated cells at 70 g for 6 min at room temperature.
21. Discard supernatant and resuspend the cell pellet in 300 μL of PBS + 10% FBS and filtered through a 40 μm cell strainer to remove any remaining clumps of tissue.

Reagent	Final concentration	Amount
PBS	1x	270 μL
FBS	10%	30 μL

22. Transfer in a 5 mL polypropylene round-bottom tube or other adapted to your cell sorter and keep on ice. Proceed to sorting as fast as possible.

Collect target cells by FACS

⌚ Timing: Approximately 1 h

Dissociated tissue is now ready for FACS, target cells are collected and processed for RNA-Seq.

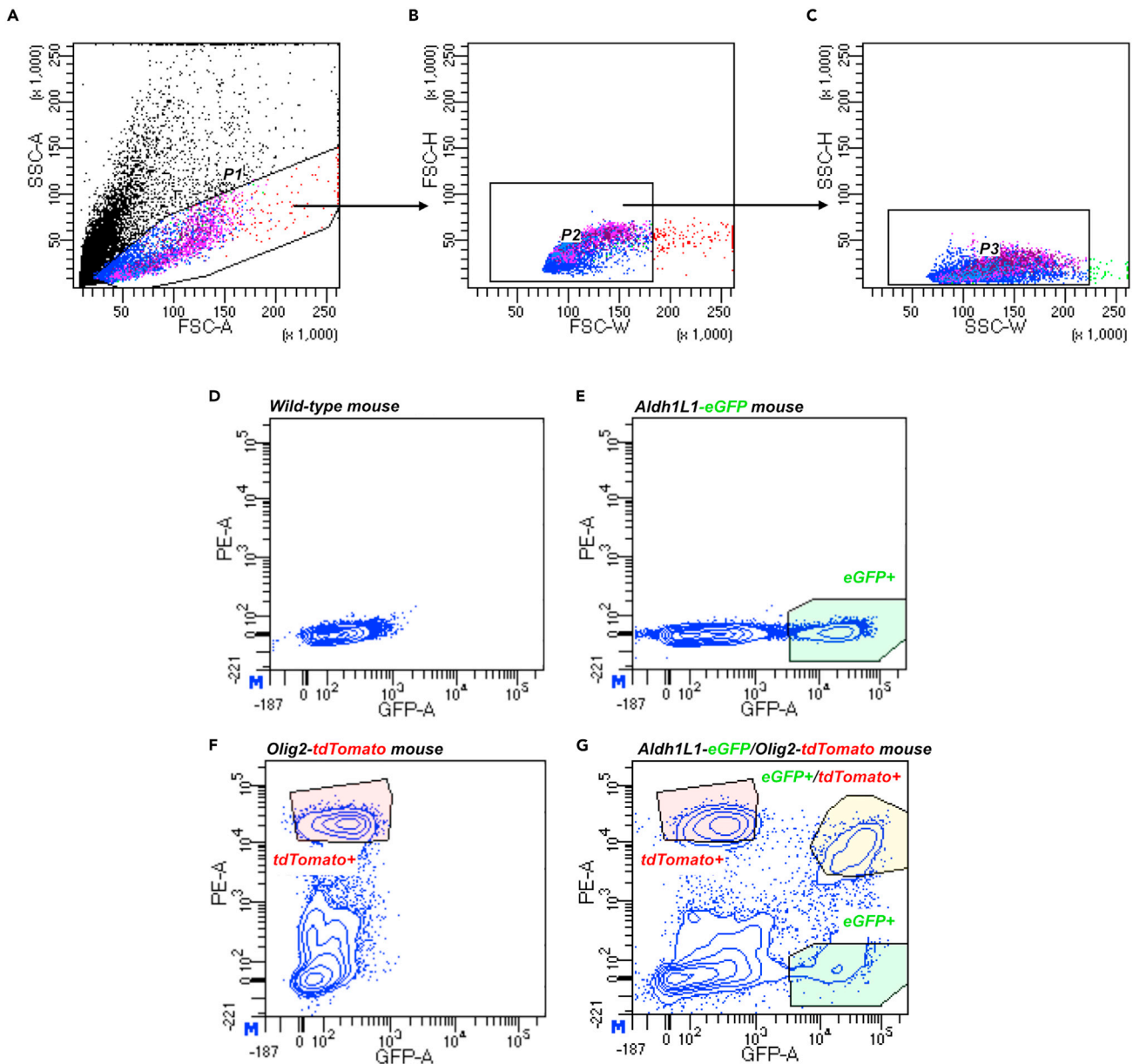


Figure 2. Gating strategies for isolation of eGFP+/tdT-, tdT+/eGFP-, and eGFP+/tdT+ cell populations

(A–C) Debris and doublets were removed as described in steps 25a and 25b.

(D–F) Representative FACS plots showing the last step of the gating strategy. Gates for eGFP+ (E) and tdT+ (F) cells were defined from cell populations purified from *wild-type* mouse (D), *aldh111-eGFP* (E) and *olig2-tdTomato* single transgenic mouse (F).

(G) Representative FACS plot of the purification of tdT+/eGFP-, eGFP+/tdT- and tdT+/eGFP+ cell populations from P7 *aldh111-eGFP/olig2-tdTomato* double transgenic mice.

23. Proceed to the FACS facility.
24. Run control samples to establish gating parameters: GFP-reporter mouse, tdTomato-reporter mouse and wild-type mice. (Figure 2)
25. Run double transgenic samples and set up gates. (Figure 2)
 - a. Remove dead cells and debris according to the plot of forward scatter area (FSC-A) and side-scatter area (SSC-A) (Figure 2A).
 - b. Exclude doublets by two successive gating approaches, first forward scatter height (FSC-H) vs forward scatter width (FSC-W) and then side scatter height (SSC-H) vs side scatter width (SSC-W) (Figures 2B and 2C).

- c. Then gate for fluorescence, to identify the three awaited cells population using a PE-A vs GFP-A: high eGFP fluorescence for the nonOlig2-AS; high tdTomato fluorescence for the OPC/OL and high eGFP and tdTomato fluorescence for the Olig2-AS (Figures 2D–2G).
26. Prepare a set of three 1.5 mL collection tubes per spinal cord sample to collect the eGFP+/tdT-; eGFP-/tdT+ and the eGFP+/tdT+ cell populations.
27. Sort the populations of interest into the prepared 1.5 mL tubes
28. Centrifuge at 2000g for 5 min
29. Discard the supernatant
30. Process the samples for RNA extraction using standard Trizol reagent with glycogen added as carrier according to manufacturer instructions (https://www.thermofisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FLSG%2Fmanuals%2Ftrizol_reagent.pdf&title=VXNlciBHdWlkZTogVFJjem9sIFJlYWdlbnQ=).
31. Elution of RNA into 35 μ L RNase-free water.

▮▮ **Pause point:** The eluted RNA samples can be stored at -80°C for up to 3 months.

32. Library preparation and RNA-Sequencing were performed by Beijing Genomics Institute (BGI, Hong Kong, China) using a BGISEQ-500 RNA-seq platform (paired-end sequencing, 100 bp reads, 30M reads/sample).

Raw data processing

⌚ **Timing:** ~ 12 h on Genotoul Galaxy cluster 48 broadwell nodes (each 64 threads, 256 GB of memory)

33. Quality control of the raw data from sequencing
 - a. Use FastQC (v0.11.5) to check the quality of each raw fastq file from each sample: two files per sample, one for read1 and one for read2 of the RNA fragment sequenced paired-end (Andrews, 2010).

Note: This tool provides some quality control checks on raw sequences coming from the sequencing platform in FASTQ format and gives an overview of the data quality with graphs and indicators in a HTML report. The report contains basic statistics, sequence quality summary, sequence content summary, information about read lengths and estimation of duplication levels. It also lists potential overrepresented sequences and verifies the presence of reads' adapters based on a list of known adapters.

- b. Use MultiQC (v1.7) to aggregate all FastQC reports and generate one single HTML report summarizing the output from all the FastQC reports (Figures 3A and 3B) (Ewels et al., 2016).

34. Alignment of the raw data from sequencing on the reference genome

Use STAR aligner (v2.1) to align paired-end files from raw sequencing on the reference mouse genome from Ensembl database: FASTA and GTF file.

Note: Based on sequential maximum mappable seed search in uncompressed suffix arrays, STAR is reliable, accurate, aligner designed for RNA-Seq dataset. The output is a single aligned file per sample in sequence alignment map (SAM) format (Dobin et al., 2013).

Note: For example, we perform alignment on the following version of the mouse reference genome: Ensembl Mus Musculus GRCm38.91 (mm10).

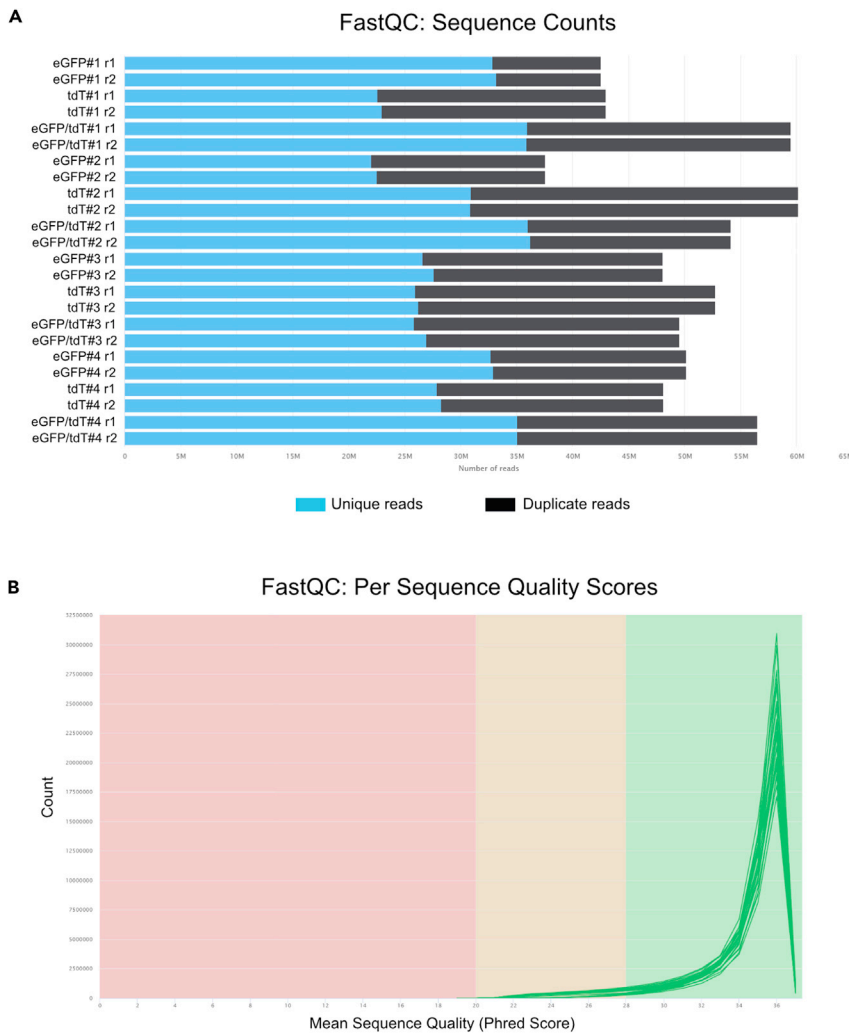


Figure 3. MultiQC report

(A) Sequence counts for each sample, showing number of unique and duplicate reads.
(B) Distribution of reads' sequence quality Phred score for each sample

35. Conversion, sorting, and indexing alignment output

Use Samtools (v1.9) to process alignment output files in SAM format (Li et al., 2009).

- a. Convert each SAM file in a compressed Binary Alignment Map (BAM) file.
- b. Sort each BAM file following genomic coordinates order.
- c. Generate an index, in BAI format, for each BAM file.

36. Quantification of reads' pairs on the chosen features

- a. Use HT-seq count (v1.0) to count the raw number of reads' pairs aligned on each annotated feature from Ensembl gtf (Anders et al., 2015).

Note: The output is a table containing, for each gene, the raw number of reads' pairs overlapping with its exon. The overlap resolution parameter is set to «union» to assign any ambiguous pair.

- b. At the same time, use Cufflinks (v2.2.1) compute for each gene the number of fragments per kilobase million reads (FPKM) (Trapnell et al., 2010).

Table 1. Read alignment and quantification for transcriptomic set #3

	Transcriptomic set #3		
	eGFP	tdTomato	eGFP/tdTomato
	GSM4802116	GSM4802117	GSM4802118
Raw nb of pairs	48,040,482	52,740,687	49,509,590
Aligned nb of pairs	46,081,202	50,818,803	47,343,603
Counted nb of pairs	27,898,597	34,662,279	30,299,452
% Aligned	95.92%	96.36%	95.63%
% Counted	60.54%	68.21%	64.00%

Note: This metric normalizes raw pair counts, taking account of sequencing depth and gene length. Each number of reads' pairs is multiplied by 10^9 and divided by the total number of pairs in the sample (library size) and the length of the gene, in kilobases.

37. Data processing control (Tables 1 and 2)

Compute percentage of read pairs aligned and quantified for each sample. Usually, the number of reads aligned should be > 60% and the number of reads falling on feature should be > 50%

Control the specificity of each sample

⌚ Timing: ~6 h on Genotoul Galaxy cluster 48 broadwell nodes (each 64 threads, 256 GB of memory)

38. Alignment of raw reads on specific sequences

To validate the selection of cells, use STAR aligned to map raw sample files on the sequences of eGFP and td-Tomato.

39. Quantification of these reads

Use Samtools idxstat to quantify the number of reads of each file on the 2 specific sequences.

40. Visualization of the output

Visualize these quantifications normalized by total sequencing depth into a bar plot (Figure 4A).

Prerequisite before differential analysis

⌚ Timing: ~1.5 h on personal computer with R (v3.6)

41. Import raw datasets in the workspace

Import the raw pairs count table for each sample. Check that each table is completely loaded then merge all tables in one common data frame with all the samples.

Note: The number of genes and the genes' order should be the same for all the samples. Remove rows corresponding to ambiguous or not assigned pairs.

42. Explore the complete raw dataset

Table 2. Read alignment and quantification for the four experimental sets

	Mean of 4 experimental sets		
	eGFP	tdTomato	eGFP/tdTomato
Raw nb of pairs	44,561,085	50,988,365	54,894,246
Aligned nb of pairs	42,506,561	48,819,524	52,252,220
Counted nb of pairs	21,068,477	30,569,845	27,762,839
% Aligned	95.42%	95.78%	95.20%
% Counted	49.42%	62.38%	53.40%

- a. Check the quality of the count table: proportion of null counts per sample, most expressed genes in each sample, statistical indicator for the distribution of reads' pairs counts for each sample and correlation between each replicate (Figure 4D).
 - b. Apply Principal Component Analysis and Hierarchical Clustering (Euclidean distance – Ward criterion) on the raw data to explore the entire data frame. Use ggplot2 R-package (v3.3.5) to draw graphs of PCA results (Figures 4B and 4C) (Wickham, 2016).
43. Clean data from non-expressed genes. Before running differential analysis, remove from the raw count table any gene having an averaged raw read count per sample lower than 1 (Figure 4E).

Differential analysis

⌚ Timing: ~1.5 h on personal computer with R (v3.6)

44. Perform normalization and parameters estimation with DESeq2

Use DESeq2 (v1.22.2) to perform Relative Log Expression (RLE) normalization on the raw counts and estimate dispersion parameter modeling the Negative-Binomial distribution of the data (Figures 5A and 5B) (Love et al., 2014).

Note: RLE is based on the computation of a scaling factor applied on each raw count. This factor is the median over all the genes of the ratio of raw gene count per gene divided by the geometric mean of the gene across all samples. This normalization method corrects both sequencing depth and mRNA composition bias that are usual in RNA-Seq. RNA-Seq count data are then modeled by a negative binomial distribution with fitted mean and a gene-specific dispersion parameter.

⚠ **CRITICAL:** This normalization method is based on the hypothesis that the majority of genes are not regulated between the 2 populations. In case of drastic variability between this normalization method might not be applicable. Moreover, this normalization method corrects both sequencing depth and mRNA composition bias but not gene length associated bias. Indeed, genes will not be compared within the same sample. To compare genes within sample FPKM normalization (described in 34.b) would be more appropriate.

45. Apply Wald tests

- a. Use DESeq2 to apply inferential testing between pairs of cell populations in order to get the results of three different differential analysis (DA): between Olig2-AS and nonOlig2-AS cells (DA1), between Olig2-AS and OPC/OL (DA2) and between nonOlig2-AS and OPC/OL (DA3).
- b. For each DA a Wald test is applied to each gene, the p value is then corrected for multiple testing with Benjamini–Hochberg method and we obtain an output with log₂ fold change (log₂FC) values as well as p value and adjusted p value for each gene (Benjamini and Hochberg, 1995)

46. Find associated gene symbol with each ID

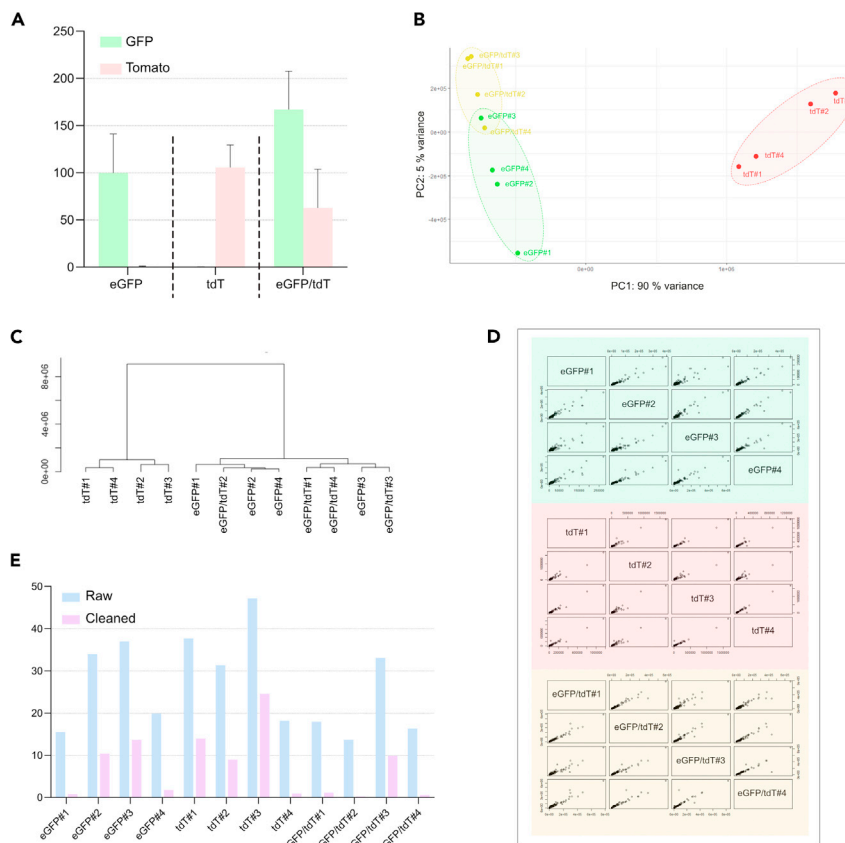


Figure 4. Control and cleaning of the raw data

- (A) Comparison of eGFP and tdTomato gene expression levels (by RPKM values) for the three cell populations.
 (B) Principal component analysis (PCA) on raw data.
 (C) Hierarchical Clustering dendrogram Euclidean distance and ward criteria on raw data.
 (D) Pairwise scatter plot between replicates.
 (E) Bar plot illustrating the raw and cleaned data from non-expressed genes for all samples.

Use the R-package Biomart (v2.48.2) to convert Mouse ENSEMBL IDs to mouse official gene names (Durinck et al., 2005).

Visualization of the results

⌚ Timing: ~3 h on personal computer with R (v3.6)

47. Global visualization of the normalized samples
 - a. To get an overview of the samples after normalization, apply Principal Component Analysis and Hierarchical Clustering (Euclidean distance – Ward criterion) on the normalized data set (Figures 5C and 5D).
 - b. Control correlation between sample computing Pearson correlation and use the ggplot2::heatmaps.2 R-package (v3.1.1) to plot the correlation results in a heatmap plot. (Warnes et al., 2015)
48. Global visualization of the DESeq2 outputs

In order to visualize the result of each DA, generate plots from DESeq2 output:

 - a. MA plot (log (basemean) against log2FC for each gene)
 - b. Volcano plot (log2FC against log10 (adjusted p value)) (Figures 5E–5G).

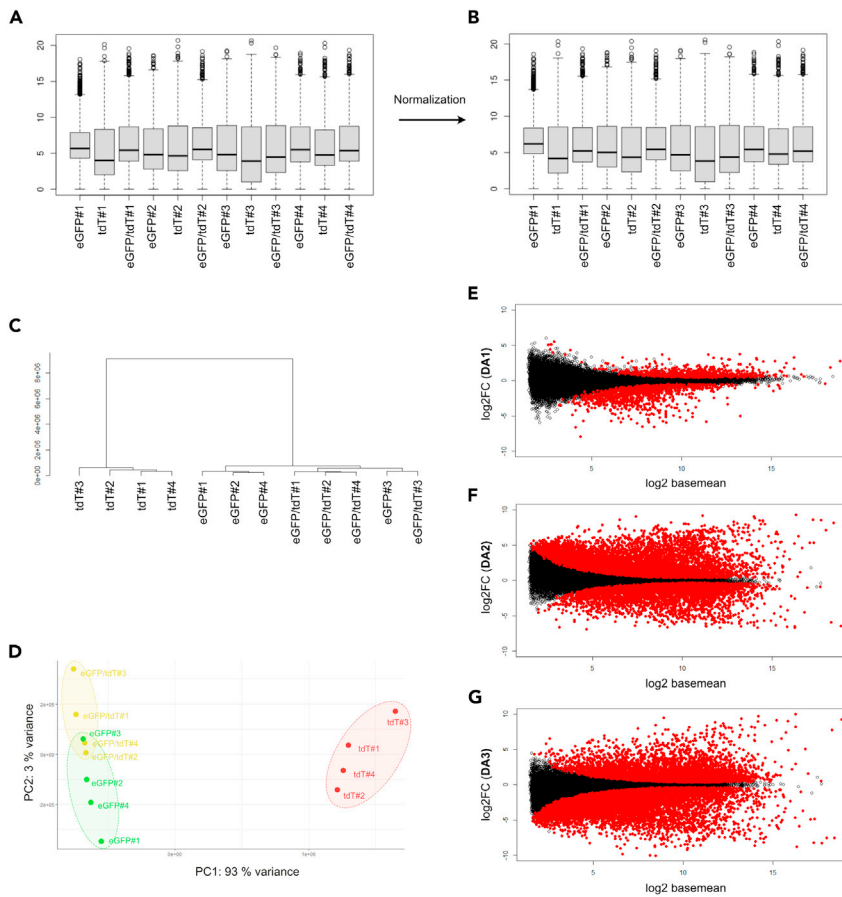


Figure 5. Differential analysis

(A and B) Boxplots on cleaned raw counts before (A) and after (B) normalization.

(C) Hierarchical Clustering dendrogram Euclidean distance and ward criteria on the normalized data.

(D) Principal component analysis (PCA) on normalized samples.

(E–G) MA plots of differentially expressed genes representing the \log_2FC and \log (basemean) from the differential analysis DA1 (E), DA2 (F) and DA3 (G). Red dots represent transcripts with a significant $\text{padj} < 0.05$ and a $\log_2FC > 0$ or a $\log_2FC < 0$.

- c. For the top 20% most variable mRNA, based on the $|\log_2FC|$, compute euclidean distances between the cell types and plot results using the `ggplot::heatmaps.2` R-package function (Figure 5C).

49. Visualization of list of markers

Use \log_2FC and adjusted p values of the different DA to select subset of genes. Then for each gene of interest compute the mean per condition, and center and scale each gene using the z-score. With the computation of z-score we are able to visualize genes of different levels of expression in the same color scale in a heatmap (Figures 6A and 6B). The two heatmaps have been made using the Prism 8 software (GraphPad).

$$z - \text{score}_{(\text{gene condition})} = \frac{\text{Normalized read counts}_{(\text{gene condition})} - \text{Row mean normalized read counts}_{(\text{gene})}}{\text{Row standard deviation}_{(\text{gene})}}$$

50. Validation of the selected candidate genes by either qPCR and/or in situ hybridization (ISH).

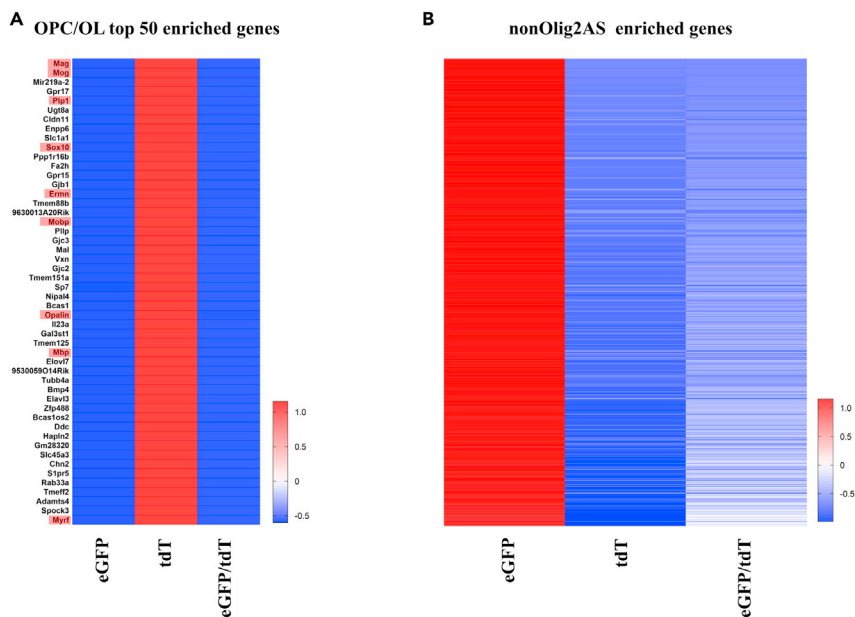


Figure 6. Visualization of list of genes

(A and B) Heatmaps depicting the abundance of mRNA corresponding to the top 50 genes enriched in OPC/OL(A) and the 649 genes enriched in the nonOlig2-AS. The gene names highlighted in red are genes known to be involved in OL/OPC function. All heatmaps are presented as z-score and genes are ordered by the log2FC (DA3) for (A) and the log2FC (DA1) for (B).

EXPECTED OUTCOMES

The differential gene expression analysis will lead to a list of differentially expressed mRNAs at post-natal day 7 in each purified cell type (Figures 6A and 6B). The differentially expressed mRNAs in the specific glial cell types are critical for downstream applications. Either for functional annotation using a GO analysis or for the validation of the candidate genes, use qPCR or ISH (Ohayon et al., 2021).

LIMITATIONS

Drawbacks can be observed during the tissue dissociation protocol, especially in terms of cell viability throughout the procedure. In this protocol, several options were taken to improve viability of the dissociated cells like the duration of the enzymatic treatment or the timeline to perform the whole experiment. However, it is advisable to perform a test run to determine the number of cells obtained from your tissue of interest before planning the downstream experiment. In terms of bio-informatic analysis, the quality of raw data has a strong influence on the analysis. Overviews using principal component analysis (PCA) can be performed at different steps of the analysis. If the data are of good quality, transcriptomes are likely to be grouped according to the cell type.

TROUBLESHOOTING

Problem 1

Low cell viability (step 13).

Potential solution

This indicates cells have been stressed throughout the protocol. Make sure to work swiftly, pipette gently, and more importantly to not over-treat the sample with the papain solution.

Problem 2

Cells clogging of the FACS machine (step 24).

Potential solution

This can happen if the extent of tissue dissociation is not carefully assessed, always be careful about any remaining meningeal tissue during the dissection. Also, do not forget to correctly filter the cell suspension through a 40 μm cell strainer to remove any remaining clumps of tissue before sorting.

Problem 3

No detection during FACS sorting (step 24).

Potential solution

Ensure samples are fluorescently labeled prior to running the FACS, the step 9 after dissection to check fluorescence right after dissection is really important for that matter. Also, cell pellets could have been lost during the dissociation procedure. The supernatants should be removed slowly and always inspect the pellet during the process.

Problem 4

Statistical analysis and figures generation ([Raw data processing](#)).

Potential solution

While bioinformatics steps can be done with Galaxy user-friendly interface, the next steps: quality controls of the read count, differential analysis, intersection between differential analysis and figures must be done in R which is both a programming language and free GNU package software designed for statistical computing. R provides precompiled executables that need to be run using the command line. These steps could be a major challenge for a non-bioinformatic user. To overcome this issue free graphical user interfaces such as RStudio, are available to provide an integrated development environment helping the user to address this challenge.

Problem 5

The large number of hypothesis tests applied during differential analysis and the impact of the cleaning of the raw count matrix on that issue ([Differential analysis](#)).

Potential solution

The DESeq2 differential analysis protocol applies simultaneous testing of many hypotheses. Indeed, in its process, DESeq2 will apply as many inference tests as there are features (genes in our case) in the input count matrix. Automatically, the more inferences are made, the more likely erroneous inferences become. To overcome these issues, statisticians developed methods to correct raw p values in order to obtain a more reasonable risk on the entire set of tests. DESeq2 applies the Benjamini–Hochberg method but the correction of p values is strongly related with the number of tests applied, making the cleaning of the raw count step very important. Indeed, the more features are removed, the more the number of tests applied decrease, the less stringent the correction method needs to be. Genes not expressed do not need to be tested as they will not be differentially expressed, on the other hand we do not want to remove any potential differential even lowly expressed. We chose to remove genes with raw count mean lower than 1 read pair over all samples to get rid of non-expressed genes. There exist both more and less stringent methods to remove lowly expressed genes; however, these methods might have consequences on the computation of corrected p values by DESeq2.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed and will be fulfilled by the lead contact, David Ohayon (david-robert.ohayon@univ-tlse3.fr).

Materials availability

This study did not generate new unique materials.

Data and code availability

RNA-seq data and processed files used during this study are available (Ohayon et al., 2021) and deposited on Gene Expression Omnibus (GEO: GSE158517).

An example code for this protocol is on: https://github.com/MarionAguirrebengoa/CombinedRNAseqAnalysis_AstrocyteSubtypeProject.

ACKNOWLEDGMENTS

We especially thank N. Rouach and P. Durbec for sharing of Aldh1L1-GFP and Olig2-tdTomato mice, respectively, and E. Nasser and P. Viana for the technical support at the IPBS-FACS platform. We thank the ABC facility and ANEXPLO for housing mice. This protocol involves the support of a large number of individuals who contributed the original publication (Ohayon et al., 2021) whom we acknowledge the help of and we thank. The graphical abstract was generated using <https://biorender.com>. This work was supported by grants from ARC (PJA20161204698), ARSEP, CNRS, and University of Toulouse.

AUTHOR CONTRIBUTIONS

Conceptualization, D.O.; methodology, D.O.; software, M.A. and D.O.; investigation, M.A. and D.O.; writing – original draft, M.A. and D.O.; writing – review and editing, M.A. and D.O.; funding acquisition, D.O.; supervision, D.O.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

- Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. <https://doi.org/10.1093/bioinformatics/btu638>.
- Andrews, S. (2010). FastQC. Babraham Bioinform. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Benjamini, Y., and Hochberg, H.Y. (1995). Benjamini and Y FDR.pdf. *J. R. Stat. Soc. Ser. B (Methodol.)* 57, 289–300.
- Cahoy, J.D., Emery, B., Kaushal, A., Foo, L.C., Zamanian, J.L., Christopherson, K.S., Xing, Y., Lubischer, J.L., Krieg, P.A., Krupenko, S.A., et al. (2008). A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J. Neurosci.* 28, 264–278. <https://doi.org/10.1523/JNEUROSCI.4178-07.2008>.
- Chaboub, L.S., Manalo, J.M., Lee, H.K., Glasgow, S.M., Chen, F., Kawasaki, Y., Akiyama, T., Kuo, C.T., Creighton, C.J., Mohila, C.A., and Deneen, B. (2016). Temporal profiling of astrocyte precursors reveals parallel roles for Asef during development and after injury. *J. Neurosci.* 36, 11904–11917. <https://doi.org/10.1523/JNEUROSCI.1658-16.2016>.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. <https://doi.org/10.1093/bioinformatics/bts635>.
- Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., and Huber, W. (2005). BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* 21, 3439–3440. <https://doi.org/10.1093/bioinformatics/bti525>.
- Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 3047–3048. <https://doi.org/10.1093/bioinformatics/btw354>.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
- Love, M.I., Anders, S., and Huber, W. (2014). Differential analysis of count data - the DESeq2 package. *Genome Biol.* 15, 10-1186. <https://doi.org/10.1186/s13059-014-0550-8>.
- Molofsky, A.V., Kelley, K.W., Tsai, H.H., Redmond, S.A., Chang, S.M., Madireddy, L., Chan, J.R., Baranzini, S.E., Ullian, E.M., and Rowitch, D.H. (2014). Astrocyte-encoded positional cues maintain sensorimotor circuit integrity. *Nature* 509, 189–194. <https://doi.org/10.1038/nature13161>.
- Ohayon, D., Aguirrebengoa, M., Escalas, N., Jungas, T., and Soula, C. (2021). Transcriptome profiling of the Olig2-expressing astrocyte subtype reveals their unique molecular signature. *IScience* 24, 102806. <https://doi.org/10.1016/j.isci.2021.102806>.
- Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. <https://doi.org/10.1038/nbt.1621>.
- Walter, W., Sánchez-Cabo, F., and Ricote, M. (2015). GOpilot: An R package for visually combining expression data with functional analysis. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btv300>.
- Warnes, G.R., Bolker, B., Gorjanc, G., Grothendieck, G., Korosec, A., Lumley, T., MacQueen, D., Magnusson, A., Rogers, J., and Others, A. (2015). gdata: various R programming tools for data manipulation. R Package Version 2, 35.
- Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis, Second Edition (Springer). <https://doi.org/10.1007/978-0-387-98141-3>.
- Zamanian, J.L., Xu, L., Foo, L.C., Nouri, N., Zhou, L., Giffard, R.G., and Barres, B.A. (2012). Genomic analysis of reactive astrogliosis. *J. Neurosci.* 32, 6391–6410. <https://doi.org/10.1523/JNEUROSCI.6221-11.2012>.