Comparison of chromatin accessibility and motif enrichment in human and mouse excitatory and inhibitory neurons

Andrea Brun
Nils Schlatter
Pamina Lenggenhager

Bioinformatic Approaches to Regulatory Genomics and Epigenomics - Project

ETH Zürich

Introduction:

The processes in the human brain such as the regulation of the vegetative nervous system, coordination of movements , formation of memories or processing of sensory information and emotions are still not entirely understood. It has, however, been established that the mammalian nervous system is extremely heterogenous in order to facilitate these processes[1]. Recent advances in RNA sequencing technology have enabled the creation of a transcriptomic cell type taxonomy in adult mice[2]. Single nucleus ATAC sequencing has also been used to profile the chromatin accessibility brain cells[3–5].

Similarly, there have been multiple publications exploring similarities between excitatory and inhibitory neurons using scRNA-seq, such as identification of differing sodium channel expression[6]. Epigenetic signatures between the two types have also been investigated using snATAC-seq[7].

The aim of this project is to analyze snATAc-seq data of human and mouse neuron in order to investigate the differences in chromatin accessibility between inhibitory and excitatory neurons by utilizing a variety of analysis methods.

Methods:

All code and plots are accessible through: https://github.com/darthnius/proj_brun-schlatter-lenggenhager.

Analysis of Human Brain Neurons

Data Acquisition

For the human brain analyses data was taken from https://decoder-genetics.wustl.edu/catlasv1/catlas_downloads/humanbrain/bigwig/ [8]. The bigwig files were split into four groups: Excitatory, Inhibitory, Glia and Other (GABAergic medium spiny neurons and dopaminergic MSN) based on their cell type.

t-SNE

t-Distributed Stochastic Neighbor Embedding was performed to visualize the accessibility profiles between the groups.

Initial GO enrichment

GO enrichment was performed with a background of the entire hg38 genome for three ontologies, molecular function, biological processes and cellular components. For the first plots only a background of all genes in hg38 was used.

BG comparisons

To enable comparison between different backgrounds a clustered heatmap was created. The backgrounds used are accessible (all accessible genes across all cell types), genome (entire hg38) and self (only genes in the group).

Chromatin Accessibility and Functional Enrichment Analysis in Mouse Neurons

Data Acquisition

We obtained single-nucleus ATAC-seq data for mouse neuronal cell types from the ENCODE project. The data included three samples: two excitatory neuron types (ENCFF150BMM and ENCFF194UMD) and one inhibitory neuron type (ENCFF448HYG)[9]. For each sample, we downloaded:

- Aligned BAM files for raw read alignments,
- Peak files (BED format) representing accessible chromatin regions
- Signal tracks (bigWig format) for quantitative accessibility profiling

Sample Metadata

We created a metadata table (`sample_info`). This metadata was later incorporated into the `SummarizedExperiment` object for downstream analysis.

Peak Processing and Consensus Region Generation

The raw peak files for each sample were imported and converted to `GRanges` objects using `data.table::fread()` and `GenomicRanges::GRanges()`. All peaks were resized to 300 bp centered on the original peak midpoint for consistency. To control for differences in peak count across samples, we downsampled all peak sets to the size of the smallest set (33,747 peaks).

We then merged all downsampled peaks into a single set of consensus peaks using `GenomicRanges::reduce()`. These consensus regions represent a unified set of accessible chromatin loci across all samples and were saved in BED format.

Read Counting and SummarizedExperiment Construction

To mitigate differences in sequencing depth, BAM files were downsampled using `samtools view -s` on a Unix-based system. The sampling fractions were:

- ENCFF150BMM: 1.1%
- ENCFF194UMD: 1.15%
- ENCFF448HYG: not downsampled

We then quantified the number of ATAC-seq fragments overlapping each consensus peak using `featureCounts()` from the Rsubread package, treating the consensus peaks as annotation features.

The resulting count matrix was combined with the genomic coordinates and sample metadata to construct a `SummarizedExperiment` object. Chromatin bias due to GC content was corrected using the `chromVAR::addGCBias()` function, with the mm10 mouse reference genome from `BSgenome.Mmusculus.UCSC.mm10`.

Normalization and Model Design

Raw peak counts were normalized using the edgeR package. A `DGEList` was created, and normalization factors were estimated with the TMM method. Sample metadata (`neuron_type`, `age_group`) was encoded as factors. A design matrix contrasting neuron types was

constructed, and the limma-voom gtransformation was applied to model mean-variance trends and compute log2-transformed values with precision weights.

Differential Accessibility Analysis

Linear models were fitted using `lmFit()` and moderated with `eBayes()`. Differential accessibility between inhibitory and excitatory neurons was assessed via the `neuron_typeinhibitory` coefficient. Top differentially accessible peaks were extracted by fold change and P-value, mapped to genomic coordinates, and visualized in a volcano plot with ggplot2, highlighting the most significant peaks for each neuron type.

Motif Accessibility Analysis

To assess transcription factor (TF) activity, mouse motifs from the HOCOMOCOv10 database were retrieved using the *MotifDb* package and converted into a `PWMatrixList` compatible with motifmatchr. Peaks with non-zero counts were used to match motifs to accessible regions via `matchMotifs()`. Motif accessibility deviations were computed using chromVAR, generating Z-scores that reflect motif activity.

Motif deviation scores were then compared pairwise, for each comparison, the top 20 motifs with the greatest absolute differences were identified and visualized in heatmaps. A final heatmap summarized the most variable motifs across all samples.

Gene Ontology Enrichment Analysis

To investigate biological functions linked to regulatory regions, we summarized chromatin accessibility signals from BigWig tracks over consensus peaks using `epiwraps::signal2Matrix`. The 15,000 most variable regions were selected, and k-means clustering was applied across a range of cluster numbers (k = 2 to 10). We computed an elbow plot showing the variance explained by each k, and selected k = 4 as the optimal number of clusters.

Each of the four resulting clusters was analyzed separately using the rGREAT package to perform GO Biological Process enrichment, with the consensus peak set as the background. Enrichment was assessed using binomial testing based on extended TSS domains from the mm10 genome. For each cluster, the top GO terms were visualized individually, and a final heatmap summarized enrichment across all clusters using –log10(p-adjusted) values.

Reproducability

To ensure reproducibility of random operations such as peak downsampling, clustering, and motif selection, a fixed random seed (set.seed(123)) was consistently applied throughout the analysis.

Results:

Human Brain

The t-SNE shows that the 4 chosen groups each form a cluster.

Figure 1: a) GO enrichment of biological processes for the 4 groups of human neurons with a background of the entire hg38 genome; b) t-SNE visualization of accessibility profiles of the 4 groups

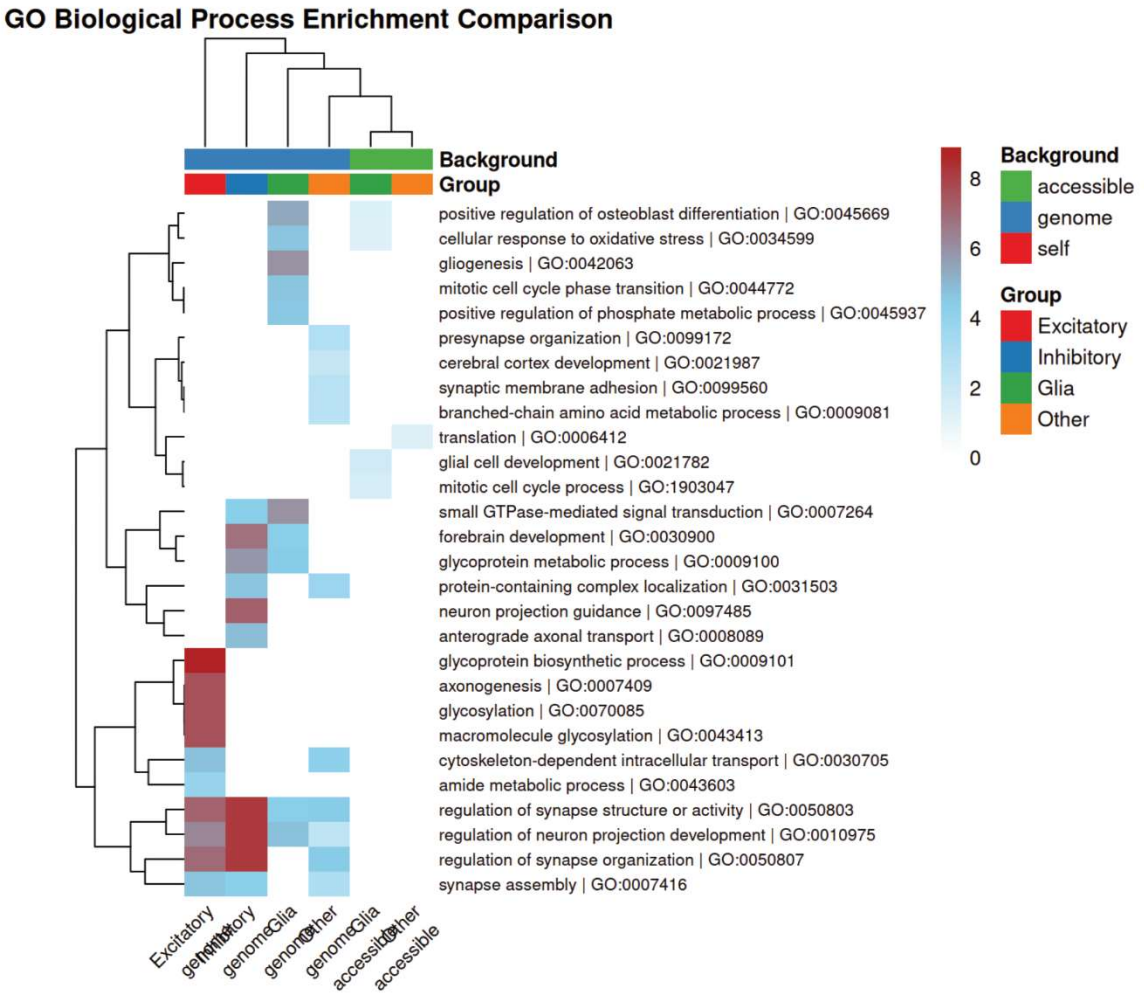

Figure 2: Background comparison for biological processes, accessible: accessible genes of any cell type as background, genome: entire hg38 as background, self: only genes accessible to the current group as background

Mouse Model

The peak width distribution looks similar for all three mouse datasets. More than 50% of the peaks have a width of less than 1000 bp. In the set for the inhibitory neurons the peaks are shorter on average. GC content in the peaks of all three neurons together is slightly lower than 50% with a longer tail in the direction of higher GC content. In excitatory neuron data peaks are enriched mainly in promoters, introns and distal intergenic regions. While inhibitory neuron peaks still show up mainly in these regions, they are found nearly three times as often in promoters in comparison to the excitatory neuron set.

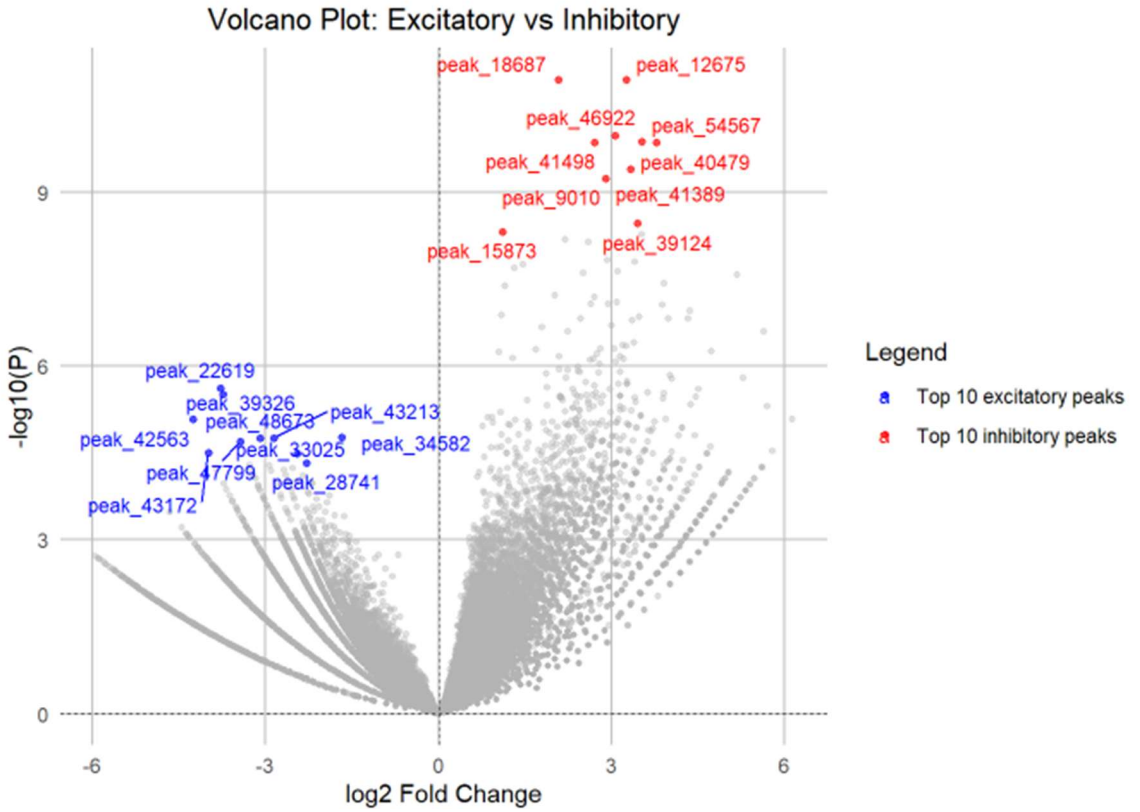

Figure 3: Differential peak accessibility between inhibitory and excitatory neurons.
Volcano plot showing log_2 fold change (X-axis) versus $–log_{10}$(p-value) (Y-axis) for all peaks. Peaks with significantly higher accessibility in inhibitory neurons (red) and excitatory neurons (blue) are highlighted. The top 10 peaks per group are labeled.

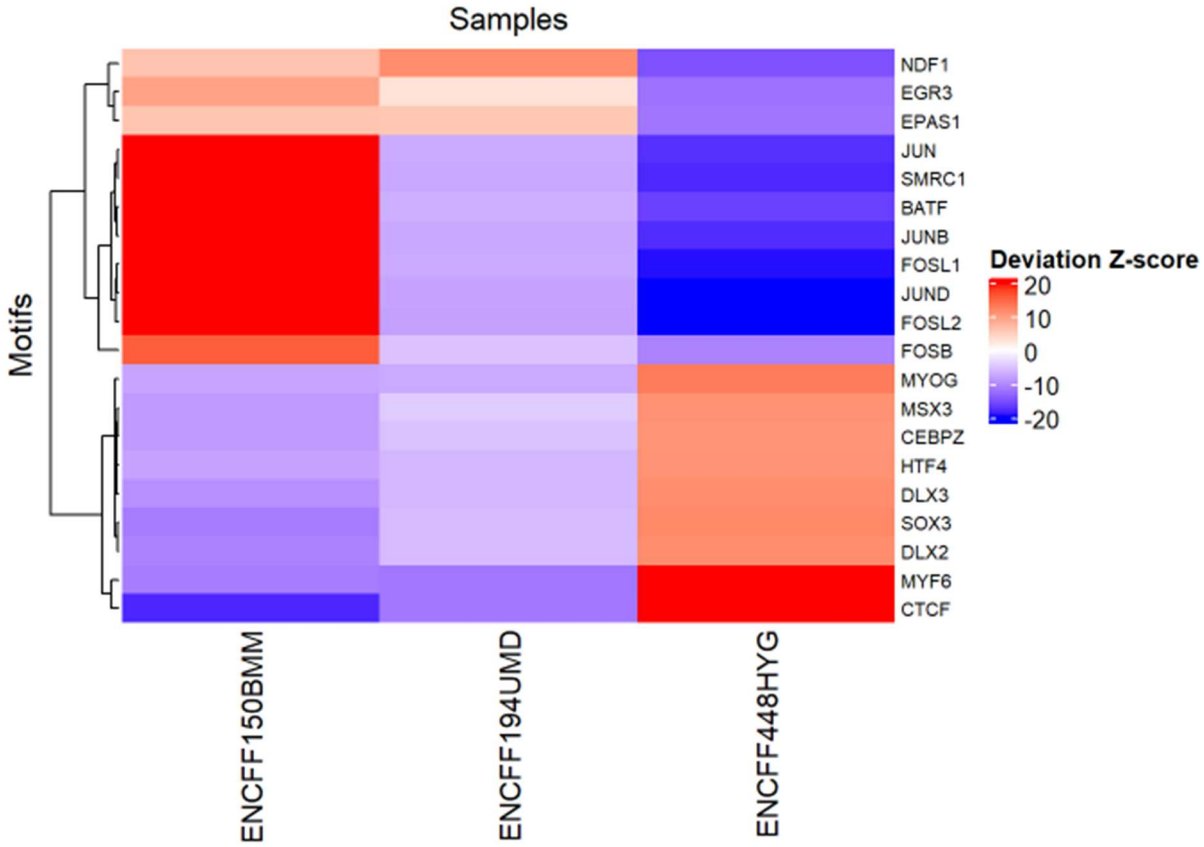

Figure 4: Motif accessibility across samples.
Heatmap showing chromVAR deviation Z-scores for the top 20 most variable motifs across the three samples: ENCFF150BMM (Excitatory), ENCFF194UMD (Excitatory), and ENCFF448HYG (Inhibitory). Rows represent transcription factor motifs; columns represent samples. Values indicate motif accessibility standardized as Z-scores.

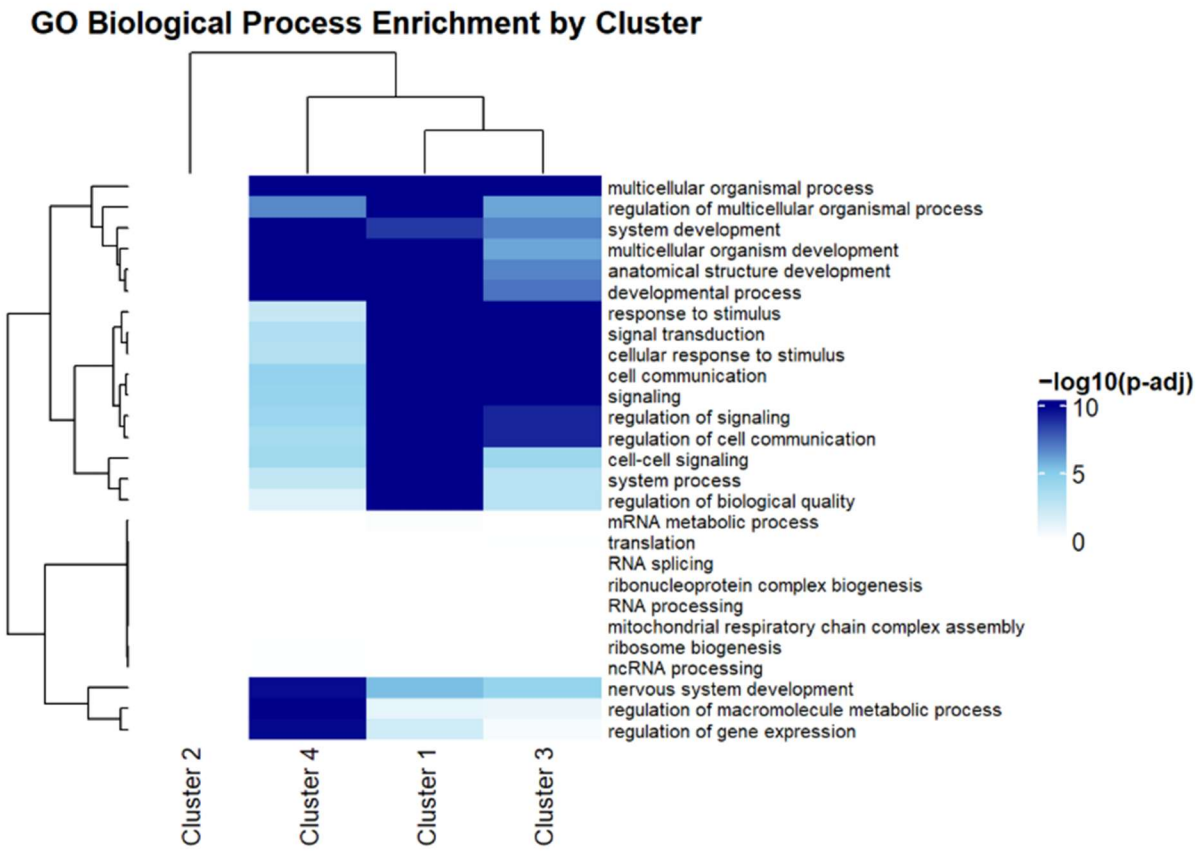

Figure 5: GO Biological Process enrichment across clusters.
Heatmap showing the –log10 adjusted p-values for the top Gene Ontology (Biological Process) terms enriched in each of the four clusters derived from signal intensity profiles. Each row represents a GO term, and each column corresponds to one of the clusters. Color intensity indicates the strength of enrichment, with darker blue representing more significant terms.

Discussion:

In this project we analyzed data from mouse and human brain to gain a better understanding of the differences between excitatory and inhibitory neurons on a molecular level. For the mouse data several scores and plots were used to assess quality and maybe identify limitations on the data level. GC bias was found to be low and should, with the according correction in the test, not influence the results negatively[10]. Complete quality testing was not possible for bam files due to limitations in computational power. As representation, smaller subsets were tested for each file. Overall, the tested data was in good shape with typical fragment sizes which also show periodicity as expected for a good ATAC-seq experiment[11]. Plots and score output were removed from the code due to high computational demands.

Differential Accessibility Between Inhibitory and Excitatory Neurons:

Figure 3 shows the results of the differential chromatin accessibility analysis comparing inhibitory and excitatory neurons. The volcano plot highlights thousands of significantly differentially accessible regions, with a substantial number more accessible in either inhibitory or excitatory cell types. This widespread difference underscores major regulatory distinctions between the two neuronal subtypes, suggesting that chromatin structure plays a critical role in defining cell-type-specific gene expression programs. The identification of significantly different peaks served as the basis for subsequent motif and GO enrichment analyses.

Motif Accessibility Patterns Across Neuronal Subtypes:

To investigate transcription factor activity associated with differential chromatin accessibility, we performed motif deviation analysis across all three neuronal subtypes. The heatmap in figure 4 displays the deviation Z-scores for the top variable motifs among inhibitory (ENCFF448HYG) and excitatory neurons from adult/child (ENCFF150BMM) and embryo/postnatal (ENCFF194UMD) mice. Notably, motifs associated with the FOS–JUN/AP-1 family (including FOSL1, FOSL2, JUN, JUND, JUNB) and BATF exhibit markedly higher accessibility in adult/child excitatory neurons, consistent with their established roles in neural plasticity and activity-dependent transcription[14,15]. Conversely, inhibitory neurons show greater accessibility for motifs such as MYF6 and CTCF, pointing toward distinct transcriptional regulation within GABAergic cell types. Embryo/postnatal excitatory neurons display lower accessibility for JUN motifs compared to their adult counterparts, suggesting a potential shift in transcriptional programs during maturation. These patterns highlight subtype-specific transcription factor activity and support the existence of regulatory divergence during neuronal development and differentiation.

GO Term Enrichment Across Accessibility Clusters:

GO enrichment analysis was performed on clusters of differentially accessible chromatin regions derived from all three neuronal subtypes. Clustering of the signal intensity profiles (figure 5) revealed that embryo/postnatal and adult/child excitatory neurons exhibited similar accessibility patterns and predominantly grouped into Cluster 2. In contrast, Cluster 1, which contained regions with higher accessibility in inhibitory neurons, showed significant enrichment for biological processes related to RNA processing, ribosome biogenesis, and metabolic activity. Clusters 3 and 4 displayed enrichment for developmental pathways such as angiogenesis, morphogenesis, and neurogenesis, although these did not reach statistical significance. The correspondence between cluster identity and cell-type specificity was also visible in a supplementary heatmap (included in the code output), where red-highlighted peaks indicate cluster membership. Together, these findings suggest a shared regulatory

landscape among excitatory neurons and a functionally distinct chromatin accessibility signature in inhibitory neurons.

Analysis of data from an atlas of single-cell data on different types of brain cells was done without previous quality assessment as extensive information on the data and derivation of the clusters shown in the original publication[8]. In our analysis we focused on excitatory, inhibitory and glial cells, all other types were grouped and used for further comparison. GO enrichment for biological processes rendered significant enrichment for a variety of terms but with overall low gene ratio as shown in figure 1a. Furthermore, terms with a lot of associated genes were predominantly found by our analysis. This in combination with the low gene ratio might suggest that the signals in the data were too weak or there was too little data to infer stronger relationships also with terms containing less associated genes.

Still, many of the reported associations can be considered meaningful when comparing molecular functions found with the expectations based on cell types. Inhibitory as well as excitatory sets showed enrichment in forebrain processes in figure 1a. While both cell types are abundant in the forebrain they are not limited to the site but found throughout the brain with special abundance in the cortex. For the excitatory set enrichment in genes associated with information processing and synapse organization as well as the membrane potential and protein localization were found as shown in figure 1a. This relates nicely to the function of excitatory neurons. With their axons they span large distances, and a great amount of information is transmitted by them via axons and synapses. Due to their long axons far from the nucleus protein and RNA transport are important to keep up functionality in the whole cell[16,17]. Our analysis detected enrichment of signals in genes related to signal transduction, various processes in synapse organization and hormonal input in inhibitory neurons. While signal transduction and synapse organization relate to the similar functions as in excitatory neurons, the hormonal input is less easy to explain[16]. For glia cells processes in the production of glycoprotein were reported as shown in figure 1a. While glycoproteins fulfill a wide variety of important functions like aiding differentiation and neurite outgrowth in several brain cell types they have a special role in Glial cells[18]. Myelin-associated glycoprotein is an important component of the axon-insulating myelin formed by glial cells in the central nervous system (CNS)[19]. Another special enrichment was found for processes associated with oxidative stress as seen in figure 1a. Glial cells perform various care-taking functions to ameliorate neuron function. Mediating oxidative stress for example by production of glutathione (GSH) for ROS scavenging in the vicinity of neurons is an important task of glial cells[20]. In the direct comparison glial cells show enrichment in some typical processes for the cells like gliogenesis. Mitosis is only visible in the glial cell group in figure 2 which might be due to the contrast between non-dividing neurons and the glial cells still capable of mitosis[21]. For the grouped brain cells the enriched terms are less informative, as the individual

cells belong to a wide variety of different cell types with diverse functions. Still some widespread functionalities like synapse organization or neuron development were found to be enriched.

The visualization by t-SNE nicely shows clustering of the same or closely related cell types as seen in figure 1b. For most of the cells the subtypes show even closer relations to each other than to others of the same excitatory, inhibitory or glial cell group. This finding suggests that the analysis of enrichment was able to capture the identity of the individual samples. In the direct comparison of GO terms associated to the different cell groups it's again possible to see a clear distinction between cell types in the structure of the tree. The enrichment patterns of closely related or functionally intertwined biological processes are often similar and improve confidence in the reported results.

Overall, mouse data was not able to show as clear tendencies in involvement in molecular processes as human data. Due to large size and complexity of the human brain cell atlas greater power was to be expected in comparison to the available mouse data. Furthermore, significant enrichment of terms suggesting activation of pathways consistent with known neuronal functions and differentiation was shown in human data, while enrichment in mouse data rendered some significant results mainly related to functionalities relevant for most cells in the body.

Main limitations were the amount of data and power to process it. Single cell ATAC-seq data is an excellent source for analysis of accessibility and motif as well as GO enrichment. The relatively small sample size of mouse data as well as the considerable difference in data on excitatory and inhibitory neurons may limit statistical power and negatively impact detection of biologically relevant enrichments.

In our case limited access to large computational power made exploration and adaptation of the analysis laborious and even required simplification of some processes. With additional resources part of the analysis could be improved.

Although the available single cell ATAC-seq data provides insights into chromatin accessibility, the assay does not directly measure transcription factor binding or gene expression which keeps conclusions about underlying mechanisms inferential.

Citations:

1. Bota, M. & Swanson, L. W. The neuron classification problem. *Brain Res. Rev.* **56**, 79–88 (2007).

2. Yao, Z. *et al.* A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell* **184**, 3222-3241.e26 (2021).

3. Thornton, C. A. *et al.* Spatially mapped single-cell chromatin accessibility. *Nat. Commun.* **12**, 1274 (2021).

4. Mulqueen, R. M. *et al.* High-content single-cell combinatorial indexing. *Nat. Biotechnol.* **39**, 1574–1580 (2021).

5. Preissl, S. *et al.* Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. *Nat. Neurosci.* **21**, 432–439 (2018).

6. Du, J. *et al.* Differential excitatory vs inhibitory SCN expression at single cell level regulates brain sodium channel function in neurodevelopmental disorders. *Eur. J. Paediatr. Neurol. EJPN Off. J. Eur. Paediatr. Neurol. Soc.* **24**, 129–133 (2020).

7. Mo, A. *et al.* Epigenomic Signatures of Neuronal Diversity in the Mammalian Brain. *Neuron* **86**, 1369–1384 (2015).

8. Li, Y. E. *et al.* A comparative atlas of single-cell chromatin accessibility in the human brain. *Science* **382**, eadf7044 (2023).

9. Moore, J. E. *et al.* Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710 (2020).

10. Van den Berge, K. *et al.* Normalization benchmark of ATAC-seq datasets shows the importance of accounting for GC-content effects. *Cell Rep. Methods* **2**, 100321 (2022).

11. Ou, J. *et al.* ATACseqQC: a Bioconductor package for post-alignment quality assessment of ATAC-seq data. *BMC Genomics* **19**, 169 (2018).

12. Beanan, M. J. & Sargent, T. D. Regulation and function of Dlx3 in vertebrate development. *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* **218**, 545–553 (2000).

13. Sobolev, V. V. *et al.* Role of the Transcription Factor FOSL1 in Organ Development and Tumorigenesis. *Int. J. Mol. Sci.* **23**, (2022).

14. Patterson, J. R., Kim, E. J., Goudreau, J. L. & Lookingland, K. J. FosB and ΔFosB expression in brain regions containing differentially susceptible dopamine neurons following acute neurotoxicant exposure. *Brain Res.* **1649**, 53–66 (2016).

15. Mulvaney, J. & Dabdoub, A. Atoh1, an essential transcription factor in neurogenesis and intestinal and inner ear development: function, regulation, and context dependency. *J. Assoc. Res. Otolaryngol. JARO* **13**, 281–293 (2012).

16. Kennedy, M. B. Synaptic Signaling in Learning and Memory. *Cold Spring Harb. Perspect. Biol.* **8**, a016824 (2013).

17. Mandal, A. & Drerup, C. M. Axonal Transport and Mitochondrial Function in Neurons. *Front. Cell. Neurosci.* **13**, 373 (2019).

18. Richter-Landsberg, C. & Duksin, D. Role of glycoproteins in neuronal differentiation. Inhibition of neurite outgrowth and the major cell surface glycoprotein of murine neuroblastoma cells by a purified tunicamycin homologue. *Exp. Cell Res.* **149**, 335–345 (1983).

19. Quarles, R. H., Ilyas, A. A. & Willison, H. J. Antibodies to gangliosides and myelin proteins in Guillain-Barré syndrome. *Ann. Neurol.* **27 Suppl**, S48-52 (1990).

20. Lee, K. H., Cha, M. & Lee, B. H. Crosstalk between Neuron and Glial Cells in Oxidative Injury and Neuroprotection. *Int. J. Mol. Sci.* **22**, (2021).

21. Ostrem, B. E. L., Lui, J. H., Gertz, C. C. & Kriegstein, A. R. Control of outer radial glial stem cell mitosis in the human brain. *Cell Rep.* **8**, 656–664 (2014).