Cross-regional uniformity of neurotransmitter receptor transcript detection across the adult mouse brain

Kentaro Abe
Independent Researcher

Abstract

Quantifying cross-regional similarity in molecular compositions is fundamental in systems neuroscience, yet atlas-scale reference baselines with explicit uncertainty remain scarce. While regional specialization is often emphasized, the complementary question—how conserved molecular compositions are across major brain macro-regions—has not been formalized with quantitative endpoints and confidence intervals.

Here, we establish an uncertainty-aware reference distribution for cross-regional similarity in neurotransmitter receptor–related transcript detection using the Allen Institute Adult Mouse Brain Cell Atlas (WMB-10Xv3; Cell Census v2023). We pre-specify a primary analysis comprising 11 non-cerebellar macro-regions and a curated panel of 26 receptor-related genes. For each region–gene pair, detected expression is defined on raw UMI counts (UMI > 0), region-wise detection fractions are computed, and within-region percentage compositions are obtained by renormalization across the fixed gene panel.

Cross-regional similarity is quantified using two complementary composition-based endpoints: (i) upper quantiles of absolute percentage-point differences across region pairs, and (ii) Aitchison distances computed in centered log-ratio (CLR) space with explicit zero replacement. Uncertainty is quantified via hierarchical nonparametric bootstrap confidence intervals (95% CI), and robustness is evaluated through pre-specified sensitivity analyses.

Across all 55 region pairs, Aitchison distances exhibited a unimodal distribution with a median of 3.66 (95% CI: 3.41–3.92) and a 95th percentile of 5.95 (95% CI: 5.62–6.18). Gene-wise upper 95th-percentile absolute differences showed a median of 1.85 percentage points (pp; 95% CI: 1.71–1.99), with maxima remaining below 3.25 pp. Sensitivity analyses confirmed that these upper-tail summaries varied by less than ± 0.1 under alternative zero-handling and log-ratio choices.

Rather than testing for the absence of regional differences, this study establishes a quantitative, uncertainty-aware reference distribution for cross-regional similarity within a standardized atlas, providing a principled baseline for contextualizing regional deviations in future comparative and perturbation-based studies.

1 Introduction

Large-scale brain organization reflects both regional specialization and conservation. Interpreting regional deviations therefore requires an explicit reference scale against which differences can be evaluated. In molecular neuroscience, regional variation in gene expression is frequently highlighted, yet the complementary question—how similar molecular compositions are across major brain macro-regions—has rarely been formalized with quantitative endpoints and uncertainty.

Single-cell transcriptomic atlases enable standardized, brain-wide comparisons across anatomically defined regions. However, atlas-scale analyses introduce intrinsic challenges: transcript detection is sparse, region-level summaries are compositional by construction, and unequal sampling

across regions and donors complicates uncertainty estimation. Without explicit baselines, visually salient gradients risk being over-interpreted.

Receptor subtype composition provides a natural entry point for addressing these issues. By treating region-wise detection profiles as compositions, similarity can be evaluated using geometry that respects relative structure rather than absolute magnitude. Aitchison geometry offers a principled framework for measuring distances between compositions while avoiding artifacts induced by closure.

Accordingly, this study is organized around a descriptive-to-quantitative progression. We first visualize receptor subtype compositions across macro-regions to establish the compositional structure of the data. We then formalize cross-regional similarity using explicitly defined, uncertainty-aware metrics. **Importantly, the goal is not to demonstrate uniformity in a hypothesis-testing sense, but to establish a reference distribution that characterizes the typical scale and dispersion of cross-regional similarity within a fixed atlas.**

2 Methods

Cross-regional similarity is evaluated through a two-stage framework, comprising an absolute-scale assessment based on upper-quantile summaries of percentage-point differences, followed by a relative-scale evaluation of compositional similarity in log-ratio space using Aitchison distances.

2.1 Data source and region definition

All analyses used publicly available single-cell RNA-sequencing data from the Adult Mouse Brain Cell Atlas released by the Allen Institute for Brain Science (WMB-10Xv3; Cell Census v2023). Cells were assigned to macro-regions using the provided atlas annotations.

The primary analysis comprised 11 non-cerebellar macro-regions: Isocortex, Hippocampal formation (HPF), Hypothalamus (HY), Striatum (STR), Thalamus (TH), Midbrain (MB), Pons (P), Medulla (MY), Olfactory areas (OLF), Cortical subplate (CTXsp), and Pallidum (PAL). The cerebellum was excluded *a priori* and examined only in sensitivity analyses.

2.2 Detected expression

For each cell i and gene g, raw UMI counts are denoted by $x_{i,g}$. Detected expression was defined as

$$I(x_{i,g} > 0),$$

where $I(\cdot)$ denotes the indicator function. Detection was evaluated on raw counts to ensure invariance to downstream normalization.

2.3 Region-wise detection fractions

Let N_r denote the number of cells assigned to region r. The detection fraction for gene g in region r was defined as

$$f_{r,g} = \frac{1}{N_r} \sum_{i \in r} I(x_{i,g} > 0).$$

2.4 Compositional normalization

Within each region, detection fractions were renormalized across the fixed receptor-gene panel of size $G = 26$ to yield percentage compositions:

$$c_{r,g} = 100 \times \frac{f_{r,g}}{\sum_{g'=1}^{G} f_{r,g'}}.$$

By construction, $\sum_{g=1}^{G} c_{r,g} = 100$ for all regions.

2.5 Uniformity metrics

For each gene g, all unordered pairs of regions were considered ($\binom{11}{2} = 55$ pairs). Absolute percentage-point differences were defined as

$$\Delta_{ij,g} = |c_{r_i,g} - c_{r_j,g}|.$$

Uniformity was summarized using pre-specified upper quantiles:

$$U_q(g) = \text{Quantile}_q\big(\{\Delta_{ij,g}\}\big), \quad q \in \{0.75, 0.95\}.$$

2.6 Aitchison distance with explicit zero handling

Because log-ratio transforms require strictly positive inputs, compositions were subjected to zero replacement using a fixed pseudo-count $\delta = 10^{-6}$, followed by renormalization. Centered log-ratio (CLR) transformation was applied:

$$\text{CLR}(c_{r,g}) = \log\left(\frac{c_{r,g}}{\left(\prod_{g'=1}^{G} c_{r,g'}\right)^{1/G}}\right).$$

where the denominator corresponds to the geometric mean of all components within region r. The resulting CLR-transformed vectors have zero sum across components. Pairwise Aitchison distances were computed as Euclidean distances in CLR space, which correspond to scale-invariant distances in the original simplex.

2.7 Uncertainty estimation

Uncertainty was quantified using a hierarchical nonparametric bootstrap (donors → cells), preserving region labels. Percentile-based 95% confidence intervals were computed from 1,000 bootstrap replicates. Bootstrap replicate-level summaries and confidence intervals were stored in machine-readable tables (*bootstrap_summary.csv*; replicate-wise values in *bootstrap_distributions.csv*), as listed in Code and Data Availability.

2.8 Sensitivity analyses

Robustness was evaluated under alternative analytic choices, including pseudo-count values $\delta \in \{10^{-5}, 10^{-7}\}$ and alternative log-ratio transforms (ILR, robust CLR).

2.9 Visualization

Region-wise receptor subtype compositions were visualized using heatmaps constructed from the percentage compositions $c_{r,g}$ (Eq. 4; *region_compositions.csv*; Figure 1). Pairwise Aitchison distances computed between CLR-transformed regional compositions (Eq. 6; *aitchison_distance_matrix.csv*) were visualized both as a distance-matrix heatmap and as a distribution across all region pairs (Figure 2A,B).

Software environment

All analyses were conducted using the following software environment:

- Python 3.12
- NumPy 2.3.5
- Pandas 2.3.3
- Scanpy 1.11.5
- Statsmodels 0.14.5
- Matplotlib 3.10.7
- Seaborn 0.13.2

Exact package versions, random seed settings, and analysis parameters are documented in the accompanying run metadata and analysis logs provided in the public repository.

3 Results

3.1 Receptor subtype compositions across macro-regions

Receptor subtype compositions across the 11 non-cerebellar macro-regions exhibited broadly conserved patterns (Figure 1). Major receptor families consistently accounted for the largest proportions of within-region compositions, while lower-prevalence classes contributed smaller but reproducible fractions.

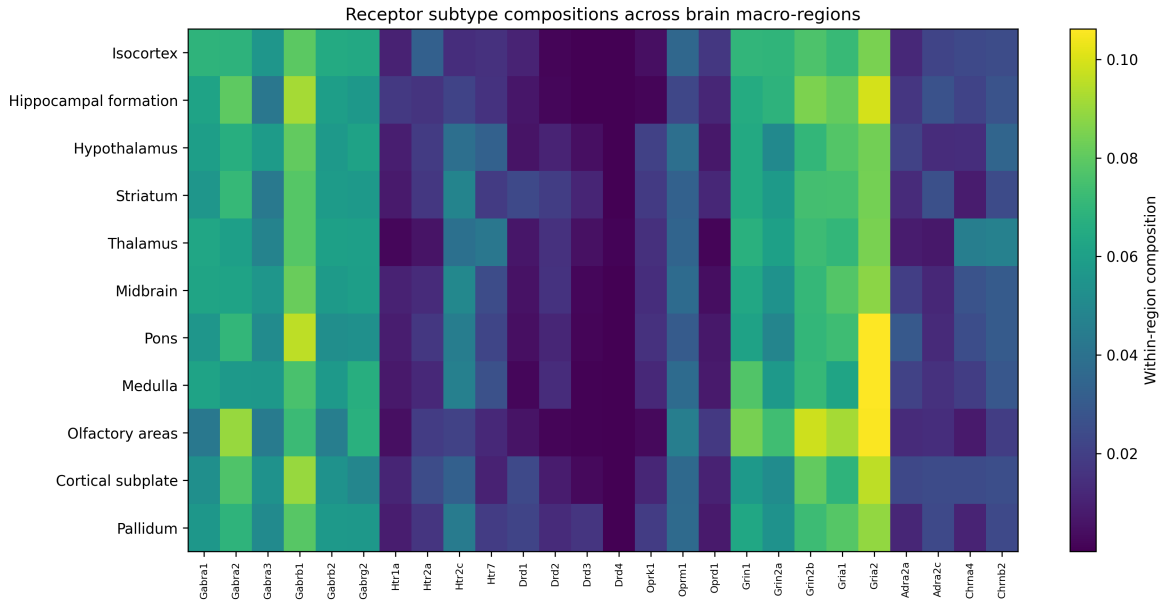
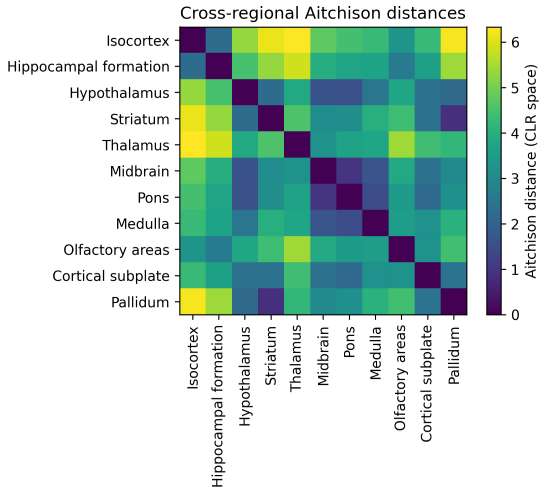


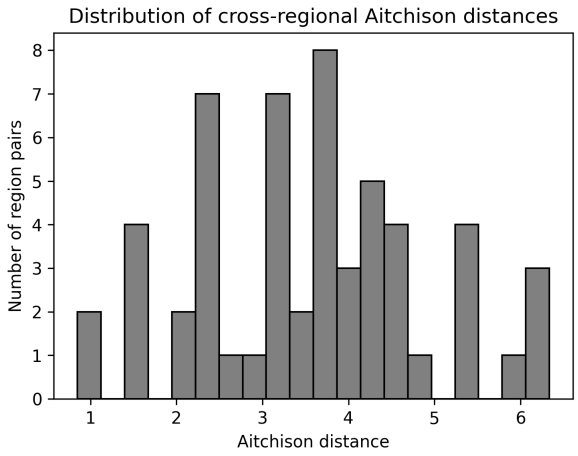
Figure 1: Receptor subtype compositions across the 11 non-cerebellar macro-regions. Rows correspond to macro-regions and columns to receptor-related genes. Values represent within-region compositional percentages.

3.2 Cross-regional similarity quantified by Aitchison distance

Pairwise Aitchison distances displayed a unimodal distribution within a restricted empirical range (Figure 2A). Across all 55 region pairs, the median distance was 3.66 (95% CI: 3.41–3.92), with a 95th percentile of 5.95 (95% CI: 5.62–6.18) and a maximum observed value of 6.34. The distance matrix (Figure 2B) revealed structured but moderate variability across region pairs.



(A) Pairwise distance matrix across macro-regions.



(B) Distribution of pairwise Aitchison distances across all region pairs.

Figure 2: Cross-regional similarity quantified by Aitchison distance. (A) Pairwise distance matrix in CLR space. (B) Distribution of distances across all 55 region pairs.

3.3 Absolute percentage-point differences

Uniformity metrics based on absolute differences were similarly constrained. Across genes, the median U_{75} was 1.16 pp (95% CI: 1.05–1.27), and the median U_{95} was 1.85 pp (95% CI: 1.71–1.99). Maximal gene-wise values remained below 3.25 pp.

3.4 Sensitivity analyses

Across alternative pseudo-count values and log-ratio transforms, the median Aitchison distance and its 95th percentile varied by less than ± 0.1, indicating high robustness of the upper-tail summaries to analytic choices.

4 Discussion

This study formalizes cross-regional similarity as a bounded empirical distribution of compositional differences rather than as the absence of regional variation. Heatmap visualization establishes the compositional structure from which all quantitative analyses follow, while Aitchison distances and uniformity metrics provide scale-aware summaries with explicit uncertainty.

Aitchison geometry is essential in this context because receptor subtype profiles are closed compositions. Distances in CLR space capture relative differences without artifacts induced by closure, enabling interpretable comparisons across regions.

Several limitations warrant consideration. Transcript detection is an indirect proxy for receptor abundance, aggregation across heterogeneous cell types may obscure cell-type–specific effects, and results are specific to the selected atlas and gene panel. Accordingly, the reported distributions should be interpreted as reference characteristics rather than universal constants.

5 Conclusion

This study demonstrated, within a quantitative framework that integrates compositional data analysis and uncertainty estimation, the extent to which neurotransmitter receptor–related transcript detection compositions are conserved across major macro-regions of the adult mouse brain.

Although receptor subtype compositions exhibited clear regional differences, evaluation using Aitchison distances and upper-quantile–based metrics showed that cross-regional variation was, overall, constrained within a limited range. These findings support the view that functional diversity of the brain is not necessarily accompanied by large-scale reorganization of molecular composition, but may instead arise from more subtle adjustments and differences at the circuit level.

The analytical framework and metrics presented here are general and are not specific to particular genes or brain regions, and are therefore extensible to other molecular systems, brain structures, and even cross-species comparisons. We anticipate that this approach will serve as a quantitative reference framework for understanding the relationship between conservation and variation in the molecular architecture of the brain.

Code and Data Availability

Repository. All analysis scripts, processed data tables, and figure-generation code are available at:
https://github.com/sotomititouru-collab/mouse-brain-receptor-uniformity

Release. This manuscript corresponds to release v1.0.0, which includes:

- Full Python analysis pipeline with fixed random seed specification (*Mouse_brain_receptor_uniformity_pipeline.py*)
- Python package requirements (*requirements.txt*)
- Region-wise detection fractions and cell counts (*region_detection_fractions.csv*, *region_cell_counts_primary.csv*)
- Region × gene compositional tables used for visualization (*region_compositions.csv*)
- Gene-wise uniformity summaries based on absolute percentage-point differences (*uniformity_U75_U95.csv*)
- Cross-regional Aitchison distance matrices (*aitchison_distance_matrix.csv*)
- Hierarchical bootstrap outputs for uncertainty estimation, including:
  - Summary statistics and 95% confidence intervals (*bootstrap_summary.csv*)
  - Replicate-level bootstrap distributions (*bootstrap_distributions.csv*)
- Reproducible figures corresponding to Figure 1 and Figure 2 (*Figure1_heatmap.png*, *Figure2A_aitchison_heatmap.png*, *Figure2B_aitchison_hist.png*)
- Run metadata and analysis logs documenting software versions and parameters (*run_metadata.json*, *analysis_log.txt*)

Ethics, Funding, and Competing Interests

All analyses used publicly available, de-identified data. No new experiments were conducted. This work received no external funding. The author declares no competing interests.

References

- Aitchison J. *The Statistical Analysis of Compositional Data*. Chapman & Hall, 1986.
- Egozcue JJ, Pawlowsky-Glahn V, Mateu-Figueras G, Barceló-Vidal C. Isometric logratio transformations for compositional data analysis. *Mathematical Geology*, 35(3):279–300, 2003.
- Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. Chapman & Hall, 1993.
- Hansen JY, Markello RD, Vogel JW, et al. Mapping neurotransmitter systems to the structural and functional organization of the human neocortex. *Science*, 378(6616):eabk3532, 2022.
- Yao Z, van Velthoven CTJ, Nguyen TN, et al. A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Nature*, 621:1–12, 2023.