Generating human Middle Temporal Gyrus (MTG) reference taxonomy

This section describes the steps required for generation of a MTG reference taxonomy, along with the associated use cases for this taxonomy. Additional information on how this cell type taxonomy is used by SEA-AD is available in the published study (Gabitto, Travaglini et al 2024)

MTG reference data set

This data set includes single-nucleus transcriptomes from 166,868 total nuclei derived from five post-mortem young adult human brain specimens. These data are used to characterize cell type diversity in the middle temporal gyrus (MTG) for multiple projects (see below), and can be considered follow-up to the "Human MTG Smart-Seq (2018)" study (website, database, publication). These nuclei were collected as part of two separate efforts: an Allen Institute-funded project specifically targeting cortical areas and a National Institute of Mental Health grant (NIMH U01 MH114812-01) targeting cells across the whole human brain. Samples were processed using the 10x Chromium Single Cell 3'™ Reagent Kit v3 (link) or v3.1 (link). 10x chip loading and sample processing was done according to the Manufacturer's protocol. Gene expression was quantified using the default 10x Cell Ranger v6 pipeline with the 10x 2020-A human genome annotation. For clarity, samples are referred to as "cells" in this document even though RNA was collected only from each cell's nucleus.

Associated cell typing projects

This reference data set is used as a baseline for cell typing in multiple additional data sets, which currently include:

- **Seattle Alzheimer's Disease Brain Cell Atlas (SEA-AD; this study):** The Seattle Alzheimer's Disease Brain Cell Atlas (SEA-AD) is a consortium focused on identifying and characterizing early changes in the brain in Alzheimer's disease and normal aging, that is funded by the National Institute on Aging (NIA U19 AG060909-01A1). This project includes 10x collected from human MTG from an aged cohort of 84 donors who span the full spectrum of disease severity.
- **Great ape (GA) study**: This study investigates cellular diversity in MTG across human and several non-human primate species: chimpanzee, gorilla, rhesus macaque, and common marmoset. It includes nuclei collected from individual cortical layers using SMART-Seq v4 (link) and from tissue sections using 10x (link) which provide a relatively unbiased survey of cell types.
- **Human cross areal (CA) study:** This study investigates cellular diversity across human cortical areas. It includes nuclei collected from individual layers using SMART-Seq v4 (link) and from tissue section using 10x (link) which provide a relatively unbiased survey of cell types.
- **Human variation study** (link): This study seeks to characterize variation of gene expression in cell types of adult human cortex, and how this variation relates to genetic signatures. Cells from this study will be assigned to the same taxonomy as SEA-AD to allow for direct comparison of cells in young adult and aged donors.

As part of the SEA-AD data release, we provide tools for label transfer from this reference dataset to novel datasets (see MapMyCells documentation) and ***we strongly encourage AD researchers to map their own data sets into this reference taxonomy, if possible***.

Previously defined cell type assignments

Both the CA and GA studies produced distinct cell type taxonomies using a combination of SMART-Seq v4 nuclei, which were collected as part of the “Human MTG Smart-Seq (2018)” study (link), and the 10x data in this dataset. The details of both analyses are presented elsewhere, but both rely on a combination of automated and manual QC, Leiden clustering using Seurat, and merging of clusters with insufficient evidence of differentially expressed genes. The CA study identified 124 within-region cell types for MTG, while the GA study identified 151 within-human cell types, using less stringent parameters. *These 151 cell types defined as part of the GA study are used as the starting point for defining SEA-AD “supertypes”.* In addition to defining these high-resolution cell types, lower-resolution subclass and class assignments are defined as described previously for mammalian primary motor cortex (publication), and match the published interlex terms. Example subclass terms include "SST", "L6 CT", and "Astrocyte", while example class terms are one of "Neuronal: GABAergic", "Neuronal: Glutamatergic", or "Non-neuronal and Non-neural". All cells passing QC are assigned to the same class, and nearly all are assigned to the same subclass across taxonomies, even though independent cell type assignments are generated for each study.

Creation of “supertypes”

We defined “supertypes” as a set of fine-grained cell type annotations for single nucleus expression data that could be reliably predicted on held-out reference data (where “ground truth” labels were assigned as described above) using state-of-the-art machine learning approaches (publication). From 5 neurotypical donors in the GA study with roughly 140K nuclei captured with 10x snRNAseq we systematically held out 2 donors and used scANVI to iteratively and probabilistically predict their class (3 labels), subclass (24 labels), and then cluster (151 labels). When predicting each nucleus’s class, we selected the top 2,000 highly variable genes along with the top 500 differentially expressed genes unique to each class (calculated from the reference cells which had their labels retained using a Wilcoxon rank sum test) to use as features in training the model and specified the donor name and number of genes detected as categorical and continuous covariates, respectively. Nuclei were then separated by their predicted class and features were re-selected with the same criteria to predict subclasses and again in predicting clusters. A differential expression test was run on clusters with an F1 score below 0.7, and those without 3 positive markers when compared against nuclei from their constituent subclass (corrected p-value <0.05, fraction in group expression >0.7, fraction out of group expression <0.3) were pruned from the taxonomy. Of the 26 clusters flagged, 24 fell below these cutoffs and were pruned from the final supertype taxonomy. The remaining 2 (L2/3 IT_2 and Oligo_3) were retained and recovered after supertype prediction (see below).

Data availability

All data, metadata, cell type assignments, and associated taxonomy files, along with a detailed readme are available on Allen Brain Map (link). The taxonomy in AIT format is also available (link). We note that this taxonomy does not yet include data from aged donors and therefore also does not included expanded non-neuronal types (see below).

Quality control and cell type assignment of nuclei from aged donors using snRNA-seq

This section applies to nuclei collected using singleome 10xV3 snRNA-seq (all donors) and 10x Multiome (28 donors), which are then mapped to the above reference transcriptome using only gene expression. The results of these mappings are extensively used in the SEA-AD web product as described below. A separate mapping of these data to cell types using a combination of gene expression and chromatin accessibility is described below, which is used for assigning cell types to ATAC-seq data and for assessment of cell type assignment confidence.

Initial removal of low-quality nuclei

SEA-AD nuclei with fewer than 500 genes detected were removed upstream of supertype mapping (see below).

Mapping SEA-AD nuclei to reference supertypes

After defining supertypes in neurotypical donors, we iteratively and probabilistically predicted each SEA-AD nucleus's class, subclass, and supertype using scANVI, as above. Each SEA-AD nucleus's class was predicted after projecting them into a shared latent space with reference nuclei using models trained with 2000 highly variable genes and 500 differentially expressed genes per class (from reference data, where donor name and number of genes were passed as categorical and continuous covariates, respectively). Nuclei were then split by predicted class, projected into a new class-specific latent space where subclass was predicted, and again for supertype. The subclass-specific latent spaces were then used to compute two-dimensional uniform manifold approximation and projections (UMAPs) and the scANVI predictions were evaluated by known marker gene expression (using signature scores defined by differentially expressed genes in reference nuclei). In regions reference nuclei occupied there was strong agreement in signature gene expression with SEA-AD nuclei, indicating accurate prediction by scANVI, but there was more variable expression in regions with poor reference support (which also had higher uncertainty in their predictions). These areas represented either droplets with ambient RNA, multiple nuclei, dying cells, or transcriptional states missing from the reference, unique to a donor or found only in aging or disease. To triage these possibilities, we fractured the graph into tens to hundreds of clusters (called "metacells") using high resolution Leiden clustering (resolution=5, k=15) and then merged them based on differential gene expression using the defaults in the transcriptomics_clustering package. Clusters and metacells were then flagged and removed if they had poor group doublet scores, fraction of mitochondrial reads, number of genes detected, or donor entropy, eliminating common technical sources of transcriptional heterogeneity.

Expanding the reference taxonomy for non-neuronal cells

With common technical axes of variation removed, we then sought to identify nuclei that were transcriptionally distinct from the reference and add them to our supertype taxonomy. We constructed a new latent space for each subclass using scVI, where the model was aware of the supertype prediction for each nucleus, gene dispersion was allowed to vary per supertype, donor name, sex, race and 10x technology (multiome versus singlome) were passed as categorical covariates, and the number of genes detected in each nucleus and the donor age at death were passed as continuous covariates. Using the neighborhood graph from this latent space, we clustered the nuclei into tens to hundreds of groups and merged them based on differential gene expression, as above. We defined merged clusters with fewer than 10% of all reference cells or of any single supertype as having poor reference support and added them

to the taxonomy (systematically named Subclass_Number-SEAAD). In cases where more than 90% of SEA-AD nuclei within these poor support groups were predicted to be one supertype, their new label reflected that assignment (e.g. Subclass_SupertypeNumber_Number-SEAAD). These cell type assignments are used as baseline for the analyses, plots, and tools developed for the web product and in-processed scientific manuscripts.

Data availability

All data, metadata, and cell type assignments for SEA-AD is available at the SEA-AD Documentation, Data, and Downloads page.

Extension of snRNA-seq data analysis to other brain structures

The reference taxonomy generation, quality control, and cell type assignment workflow described above for MTG was extended to two additional SEA-AD efforts: a study of the caudate nucleus (Ca; Kana et al 2026) and a multi-region study spanning ten neo- and allocortical brain regions (Travaglini, Gabitto, Ding, et al 2026). The sections below describe each effort with a focus on what differs from the MTG workflow; steps not explicitly noted as different were performed as described in the MTG sections above, and more details are available in the relevant manuscripts.

Caudate nucleus study

Reference taxonomy and cell type assignment

Unlike MTG, no new reference taxonomy was generated for the caudate nucleus. Instead, nuclei were assigned to cell types using a combined reference atlas that pairs the existing SEA-AD MTG taxonomy with the BRAIN Initiative Cell Atlas Network's (BICAN) Human and Mammalian Brain Cell Atlas (HMBA) cross-species consensus taxonomy of the basal ganglia (Johansen et al. 2025). In practice, this meant that the lateral ganglionic eminence (LGE)-derived medium spiny neurons and other basal-ganglia-specific neuronal populations not represented in cortex were labeled using the HMBA basal ganglia taxonomy, while non-neuronal populations – which are largely shared with cortex – were labeled using the existing SEA-AD MTG taxonomy. SST+ CHODL+ and a subset of VIP+ inhibitory interneuron populations were found to be shared across both taxonomies.

To build the combined reference, the SEA-AD MTG reference taxonomy (used as the reference) was integrated with the human caudate (head and body) portion of the HMBA basal ganglia cross-species taxonomy (used as the query) using the same iterative scANVI procedure described for MTG. Only HMBA cells sampled from human caudate head or body were used. The latent space corresponding to the neuron versus non-neuron classification was clustered de novo with Leiden clustering, and the overlap between the two atlases was quantified per cluster by computing the entropy of atlas mixing among each cell's nearest neighbors and averaging within cluster. At an overlap threshold of 0.075, HMBA cells in clusters above the threshold were relabeled with their corresponding SEA-AD labels, while clusters below the threshold retained their HMBA labels. This relabeled, combined atlas was then used as the reference to query labels for the SEA-AD caudate nuclei via iterative scANVI, applied hierarchically across all levels of the taxonomy. The mapping algorithm is therefore the same hierarchical, iterative scANVI approach used for MTG; only the reference – a combined SEA-AD MTG/HMBA basal ganglia atlas rather than the SEA-AD MTG taxonomy alone – differs. The harmonized snRNA-seq reference was subsequently used to annotate the caudate spatial transcriptomics dataset.

Quality control

Quality control followed the same overall strategy used for MTG – initial heuristic filtering followed by per-cell-type assessment on the cell-type-specific latent space – but with caudate-specific thresholds. Nuclei were first filtered with initial heuristics: nuclei were retained if they had between 2,000 and 100,000 unique molecular identifiers (UMIs), between 1,000 and 13,000 genes detected, a doublet score below 0.3, and a fraction of mitochondrial UMIs below 0.03. For neuronal nuclei specifically, an additional requirement of at least 2,000 genes detected was applied. (This differs from the MTG workflow, where low-quality nuclei were initially removed using a single threshold of fewer than 500 genes detected.) As in MTG, additional quality control was then performed on a per-cell-type basis using each cell type's latent space: the nearest-neighbor graph was fractured into tens to hundreds of clusters with high-resolution Leiden clustering, and clusters were flagged and removed based on poor group doublet scores, fraction of mitochondrial reads, and number of genes detected, with cutoffs adjusted for each cell type at each level of the hierarchy to account for the dramatically different RNA content found across cell types.

Data availability

All data, metadata, and cell type assignments are available at the SEA-AD Documentation, Data, and Downloads page.

Multi-region study

Extending the reference taxonomy

The multi-region study profiled ten neo- and allocortical brain regions. Rather than generating a new reference taxonomy, the same methods used to build the MTG taxonomy were applied to extend the existing SEA-AD MTG taxonomy to additional subclasses and supertypes that capture regionally specialized populations. snRNA-seq and snMultiome nuclei from each region were first mapped to SEA-AD MTG classes and subclasses using scVI and scANVI (version 1.2.2.post2), using the same iterative, hierarchical procedure and the same feature-selection criteria described for MTG (top 2,000 highly variable genes plus the top 500 differentially expressed genes unique to each class, with sequencing method as the batch key and library ID as an additional categorical covariate).

After subclass mapping, nuclei were split into two analysis branches depending on predicted subclass identity. Inhibitory and non-neuronal subclasses were merged across regions within each subclass, then projected into subclass-specific latent spaces where cell type was predicted by mapping to SEA-AD MTG cell types. As in MTG, the subclass-specific latent spaces were used to build a nearest-neighbor graph and UMAP, and areas of the graph with few MTG reference nuclei were triaged by fracturing the graph into metacells with high-resolution Leiden clustering (resolution = 5) and merging them based on differential gene expression. Thresholds were adjusted within each subclass to mirror the clustering resolution used to build the original MTG taxonomy, and metacells and merged clusters were flagged and removed based on group doublet scores, fraction of mitochondrial reads, number of genes detected, or donor entropy, with cutoffs adjusted per subclass. After removing common technical axes of variation, clusters that passed quality control but had few reference nuclei were tested for differential expression, and those with at least 3 positive markers relative to their constituent subclass were added to the taxonomy as new cell types.

Cells mapping to excitatory subclasses were instead merged across all regions and a joint latent representation was learned with a new scVI model. Low-resolution Leiden clusters

(resolution = 1) were flagged and removed for technical effects as above, and remaining low-resolution clusters with poor reference support were added as new subclasses – namely Sub-CA1, CA2-4, DG, and EC IT. For excitatory subclasses already represented in the MTG reference, nuclei were projected into the corresponding subclass-specific latent space, mapped to MTG cell types, and new cell types were added when clusters passed quality control and differential testing. For subclasses not represented in the MTG reference (e.g. those specific to hippocampus or entorhinal cortex), a new scVI model was trained to learn a latent space, and high-resolution clusters were computed and merged based on differential genes, with cell type addition performed identically to the inhibitory and non-neuronal branches. Finally, annotations from the inhibitory, non-neuronal, and excitatory branches were unified, subjected to a final round of quality control, and consolidated into a common taxonomy. Through this process the MTG taxonomy was expanded to include five new subclasses corresponding to regionally restricted populations (EC-IT, Sub-CA1, CA2–4, DG, and Ependymal) and 70 new molecularly distinct cell types, for a total of 207 cell types organized into 28 subclasses and 3 higher-level classes.

Quality control and cell type assignment

Quality control and cell type assignment for the multi-region study followed the same approach described for MTG, applied within the iterative mapping pipeline above. Nuclei were initially filtered for low quality (as in MTG), then assigned to class, subclass, and cell type using hierarchical, iterative scVI/scANVI mapping to the SEA-AD MTG taxonomy. Per-cell-type quality control was performed on the subclass-specific latent spaces by fracturing the nearest-neighbor graph into metacells and removing those with poor group doublet scores, fraction of mitochondrial reads, number of genes detected, or donor entropy, with cutoffs adjusted for each subclass to account for differing RNA content across cell types. The principal difference from MTG is that this single workflow simultaneously assigned nuclei to the existing taxonomy and identified the new, regionally specialized subclasses and cell types described above.

Data availability

All data, metadata, and cell type assignments are available at the SEA-AD Documentation, Data, and Downloads page.

Development of web tools for presentation of -omics data

This section includes methods for generating plots and statistics for several current and near-future web tools for analysis, presentation, and visualization of singleome and multiome snRNA-seq data as part of SEA-AD.

Comparative Viewer

The Comparative Viewer is a tool aimed at allowing users to explore gene expression in aged donors in the context of donor metadata and cell types. A primary visualization used in the Comparative Viewer and CZI cell-by-gene (below) is a UMAP, which is defined as described above for the Transcriptomics Explorer, except that the latent space is defined using data from both the reference MTG dataset as well as both the singleome and multiome snRNA-seq data from aged donors, for a total of roughly 1.3 million cells.

The Comparative Viewer presents an interactive view of six panels which allows a user to view expression of a single gene in the context of a specific donor metadata for part or all of the MTG taxonomy. This is an interactive tool such that all panels update to reflect the correct comparison as cell type resolution, gene, or donor metadata change. The left two panels show

cell type assignments and gene expression for cells from the reference SEA-AD dataset described for the tools above. The center two panels show the same information, but for data from the aged donors, shown in the same latent space for direct comparison with the reference. The right panels focus on comparison of donor metadata in the context of cell types and gene expression. The upper right panel shows the same cells from the middle panels color-coded by a specific donor meta data, to provide a qualitative view of how the metadata relates to gene expression and cell types. The lower right panel provides a summary of gene expression and associated statistics for a specific gene across cell types. Dot plots show the average expression and proportion of a gene's expression for a given cell type within a given meta-data categories. Finally, genes with significant differential expression across any comparison (see details in the next section) are indicated.

In the dot plot panel of the comparative viewer, differentially expressed genes were identified among subclasses within a class or supertypes within a subclass using a general linear mixed model, implemented in the NEBULA R package, with the following formula: gene expression ~ cell type + sex + age at death + race + 10x technology + fraction of mitochondrial reads. Cell type was encoded as a 0 or 1 if inside or outside the group of interest, sex as "M" or "F", age at death as a binned ordinal with the groups "65 to 77 years old", "78 to 89 years old", and "90+ years old", race as "white" or "non-white", 10x technology as "singleome" or "multiome" and fraction of mitochondrial reads as a continuous covariate. To determine an appropriate p-value cutoff, we re-ran the test after shuffling the cell index linking the expression values and metadata. Differential gene expression within a subclass or supertype across overall Alzheimer's disease neuropathic change (ADNC), Braak stage, or cognitive status was determined with the same package and p-value control using the model: gene expression ~ (ADNC or Braak or cognitive status) + sex + age at death + race + 10x technology + fraction of mitochondrial reads. When a secondary covariate included in the models is selected in the comparative viewer, the p-values used for determining significance come from the test where ADNC was the primary covariate. Metadata available in the comparative viewer, but which was not included in the differential gene expression model do not have notations for significance.

CZI cell-by-gene

Cellxgene (cell-by-gene) is an interactive browser that provides users with a UMAP representation of the data that can be colored by gene expression or metadata (reference). Cellxgene enables scientists to annotate, publish, find, download, explore and analyze single cell datasets. It allows them to select nuclei based on any variables or by circling cells live and then either subset the dataset or search for differentially expressed genes. This instance targets data in human MTG from young adult and aged donors and is the main interactive tool for gene expression exploration in the initial SEA-AD release.

Inputs for the CZI cell-by-gene, including data, metadata, and associated data models match precisely what is described for the Comparative Viewer, except that a few additional pieces of metadata are included. AnnData objects for cellxgene follow a standardized schema created by the Chan-Zuckerberg Initiative. Raw UMIs are stored in AnnData.raw.X, log-normalized UMIs per 10,000 plus 1 in AnnData.X, de-identified and binned-for-display metadata in AnnData.obs, the subclass-specific latent space in AnnData.obsm["X_scVI"], the UMAP coordinates in AnnData.obs["X_umap"], and display colors in AnnData.uns[metadata_variable + "_colors"].

Transcriptomics Explorer

The Transcriptomics Explorer is a web application for visualization of gene expression and associated metadata in the context of an annotated reference data set, in this case using the MTG data and SEA-AD supertypes described in the "Generating human Middle Temporal Gyrus

(MTG) reference taxonomy" section of the "snRNA-seq analysis" white paper. A detailed explanation of all files included in the Transcriptomics Explorer is described in the Readme that is downloadable on Allen Brain Map (link). Additional features of Transcriptomics Explorer can be explored using 'on hover' tool tips available throughout the site.

In this case the taxonomy of clusters (dendrogram) was generated by arranging supertypes using transcriptomic similarity based on hierarchical clustering. To do this the median expression level of the top 3,000 most variable genes was calculated for each cluster, using the FindVariableFeatures function in Seurat. A correlation-based distance matrix was calculated, and complete-linkage hierarchical clustering was performed using the build_dend R function in scrattch.hicat. The resulting dendrogram branches were reordered to match previous work in human primary motor cortex (reference, Transcriptomics Explorer) to the extent possible. Common Cell type Nomenclature (CCN) was applied to the dendrogram as described (reference; website), and unique identifiers for each supertype and dendrogram node are shown on hover. Adjustable heat maps corresponding to gene expression and cell type metadata are shown below the dendrogram in the "Heatmap" and "Sampling Strategy" sections of the Transcriptomics Explorer, respectively. The Heatmap section shows trimmed mean expression of a manually curated set of marker genes, along with any additional user-defined genes added using the "Add Genes" button at the right of the tool bar. The Sampling Strategy section shows the percentage of cells in each cell type that correspond to different tissue, donor, and cell metadata. For all categories, columns sum to 100%.

Uniform Manifold Approximation and Projection (UMAP) coordinates for each sample are shown on the Transcriptomics Explorer in the "Scatter Plot" section. UMAP is a method for dimensionality reduction of gene expression that is well suited for data visualization (reference). The UMAP coordinates were calculated from a nearest neighbor graph constructed with scanpy.tl.neighbors (reference). More specifically, a 20-dimension latent representation of the entire counts matrix was learned with scVI (reference), with gene dispersion allowed to vary across each cluster label and the donor and number of genes encoded as categorical covariates. In the "Scatter Plot" each point corresponds to a cell, and cells can be color-coded either by SEA-AD supertype or by expression of any single gene. When coloring by gene expression, a histogram of gene expression for that gene for a single cell type can be shown on hover. All cells mapped to a specific supertype can be seen by clicking on any cell of that supertype.

Gene Expression Trajectory Viewer

The Gene Expression Trajectory Viewer is a web application for identifying and visualizing cell type-specific gene expression changes across Alzheimer's disease pseudo-progression at multiple levels of the MTG taxonomy. It draws on two inputs: quantitative neuropathology, used to order donors by disease severity, and single-nucleus gene expression. snRNA-seq from roughly 1.7 million MTG cells (~1.2 million after QC), assigned to 139 supertypes, was used to generate more than 6 million sets of images spanning every combination of gene and cell type across taxonomy levels, which can be accessed by searching for genes and cell types, or by sorting and filtering by statistical metrics.

Donors are ordered along a continuous pseudo-progression score (CPS). MTG tissue sections were stained for pathological proteins and cell types relevant to AD (e.g. AT8 for pTau-bearing neurofibrillary tangles and 6E10 for Aβ plaques, plus markers for comorbid pathologies and cellular populations), and machine learning was used to quantify staining on the whole slide images. A Bayesian model inferred pathological burden from each variable's trajectory and assigned each donor a score from 0 to 1, from least to most pathology. The CPS defines two

disease epochs, an early epoch with low, slowly increasing pathology and a later epoch with an exponential increase in plaques and tangles; the boundary is marked on the trajectory plots.

Gene expression changes are then quantified against the CPS using a negative binomial generalized linear mixed model (the NEBULA R package), controlling for donor and technical covariates. Each gene and population yields an effect size for the full trajectory ("All") and for the early and late epochs, where positive values indicate increases in AD and negative values decreases (as of February 18, 2025, the tables report Effect.size rather than the underlying beta coefficient). These statistics and the associated trajectory plots come from the same differential expression analysis used to characterize vulnerable, disease-associated, and unaffected populations in the SEA-AD MTG study (Gabitto, Travaglini et al., 2024; see Figure 5).

MapMyCells

MapMyCells allows users to assign SEA-AD cell types to their own single-cell or single-nucleus transcriptomics data. The application uses vetted, state-of-the-art mapping algorithms to compare user data to SEA-AD cell type taxonomies. By default, mapping to the MTG reference is done using a deep generative model algorithm (derived from scVI), which puts query data into the same latent space as data from the reference taxonomy. Mapping to the multi-region reference taxonomy (and all other taxonomies hosted on Allen Brain Map) are done by default using hierarchical correlation mapping. Given the hierarchy tree of reference clusters, query cells traverse the tree down to the terminal cluster selecting the branch with the minimal distance to the query data using pre-calculated marker genes with correlation as distance metric (more details in Daniel et al 2026). Efforts are underway to apply a deep generative model to the multiregion taxonomy as well. For mapping data from caudate nucleus, we encourage the use of the Consensus Basal Ganglia taxonomy (for neurons) and the 10x Human MTG SEA-AD taxonomy (for non-neurons).

Allen Brain Cell (ABC) Atlas

The Allen Brain Cell (ABC) Atlas allows users to explore gene expression levels in human cortical cells in the context of donor demographics and disease metadata using a UMAP. The single-nucleus RNA-seq instance assays middle temporal gyrus and A9 (prefrontal cortex) from all donors, with all 2.78 million nuclei organized into a joint taxonomy that allows exploration of gene expression in the context of cell types and donor metrics. A companion spatial transcriptomics instance of the ABC Atlas allows users to explore cell type co-localization and gene expression in the largest cell-resolution MERFISH data set in AD human brain to date (302,000 cells), assaying middle temporal gyrus from 24 aged donors (a subset of the 84-donor SEA-AD cohort) and based on a 140-gene panel created to define neuronal types in MTG. The single-nucleus and spatial transcriptomics instances can be viewed in parallel. Additional single cell and spatial transcriptomics data sets from the caudate nucleus and multi-region studies will be made available in the ABC Atlas later in 2026.