

White Paper: Brain Cell Type Naming Conventions in the Era of Single Cell Sequencing

Introduction

Brain cell type names have historically reflected the features that were most observable or experimentally accessible at the time. Terms such as pyramidal neurons and chandelier cells are visually evocative of distinctive morphological features; neurogliaform cells similarly convey an unusual glia-like shape. Other names capture different defining features: fast-spiking interneurons are so named for their recorded firing behavior, and SST interneurons are named for the expression of a defining neurotransmitter. Altogether these names emerged from this tradition and remain useful because they are interpretable and connected to decades of literature. Other cell types were named for the scientist who discovered them, such as Martinotti cells, Betz cells, and von Economo neurons, or based on their developmental lineage, such as intermediate progenitor cells and oligodendrocyte precursor cells. Even with efforts to formalize terminology (e.g., (Ascoli et al. 2008)), these naming systems are incomplete, representing some of the most distinctive and historically recognizable neurons while failing to capture the full breadth of cellular diversity. Because they draw on different naming traditions, they also present challenges for systematic application and comparison across studies.

More recently, single-cell and single-nucleus RNA-seq (sc/snRNA-seq) have enabled cells to be classified at unprecedented resolution using the complete transcriptomes of millions of cells across multiple species. Such approaches improve classification because they measure many features at once, are extremely high-throughput, are cost-effective, reduce bias over traditional approaches, can be extended across all organ systems, and "facilitate genetic access for functional studies, as well as analyses of development, evolution and disease" (Zeng and Sanes 2017). Such classification results in the identification of thousands of neuronal and glial cell types in the mammalian brain, far outpacing the capability of the prior naming systems without full phenotypic characterization of each neuron. For example, while prior literature describes fast-spiking parvalbumin interneurons as a small number of morphologically defined categories, such as basket cells and chandelier cells, transcriptomic studies now distinguish dozens of related types based on gene expression. These transcriptomic cell types are organized hierarchically into taxonomies and, when possible, the cell types are aligned to spatial, morphological, and/or electrophysiological data. Although transcriptomically defined cell types are often highly reproducible and robust, they also create naming challenges because clusters may not be cleanly distinguished by one or a few marker genes and may not map directly onto historical categories.

Consistent and interpretable cell type naming is essential for integrating results across studies, modalities, and species in modern neuroscience. As large-scale transcriptomic and multimodal datasets have accumulated over the past decade, naming conventions have evolved to balance biological accuracy, interpretability, and scalability across increasingly complex datasets. Early approaches often emphasized descriptive, marker-driven labels tied to specific experiments, whereas more recent efforts aim to standardize nomenclature across datasets, brain regions, and species. A useful name should balance biological meaning and usability, drawing on features such as anatomical location, marker genes, neurotransmitter identity, physiology, morphology, or literature precedent when those features are informative. The goal is not for every name to include every possible attribute, but rather to select an informative and accurate set of components that can be shared alongside associated data and metadata, enabling cell types to be recognized across studies as nomenclature continues to evolve.

The Allen Institute has created more than a dozen cell type taxonomies over the past decade, and the number continues to grow (**Figure 1**). This document summarizes several major naming conventions used across these Allen Institute studies, highlighting their structure, strengths, and trade-offs.

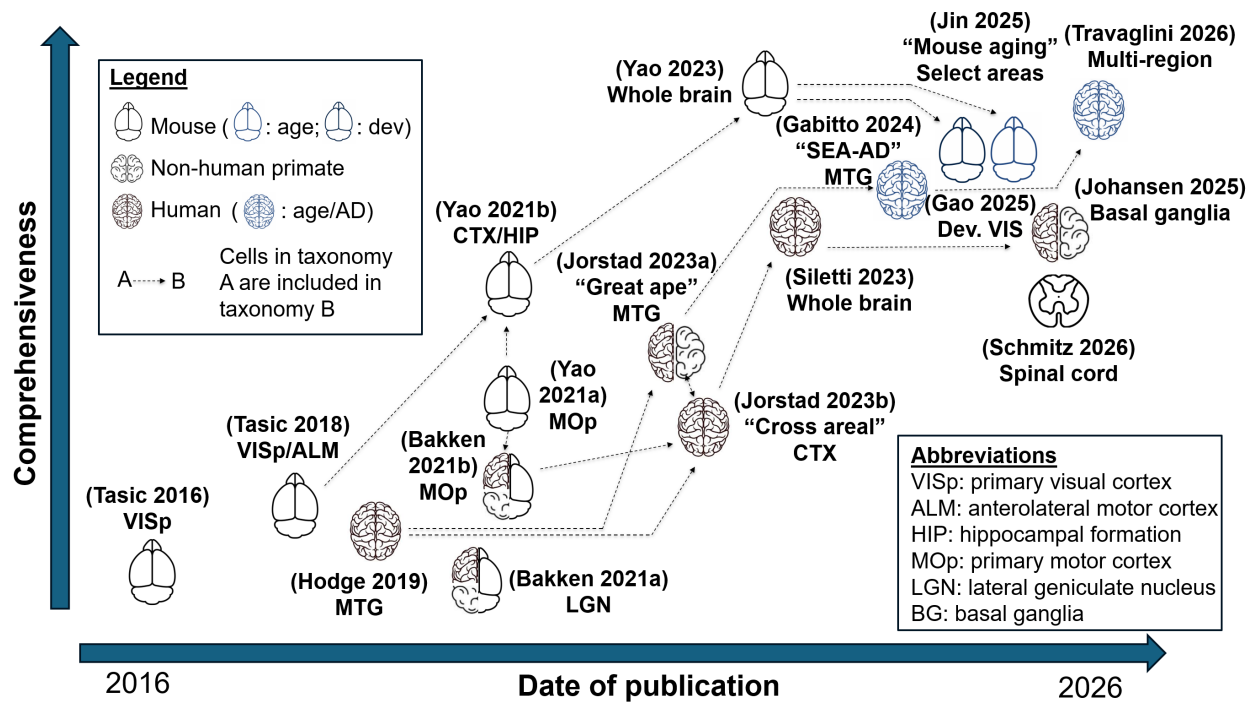


Figure 1: Published Allen Institute and BICAN taxonomies available through October 2026. Cell type taxonomies have gotten progressively more comprehensive over time by increasing the number of cells, brain regions, species, and modalities included, leading to evolving challenges in cell type naming. Figure reproduced from an upcoming version of (Ray et al. 2026).

Naming Conventions Across Studies

Tasic et al. 2018 (Mouse VISp and ALM scRNA-seq)

In this study (Tasic et al. 2018), cell types were defined from 23,822 single-cell transcriptomes collected from mouse primary visual cortex (VISp) and anterior lateral motor cortex (ALM), yielding 133 transcriptomic types (or “clusters”): 56 glutamatergic, 61 GABAergic, and 16 non-neuronal types. This was the first cell type taxonomy to establish a strict hierarchy, with major branches referred to as classes (e.g., glutamatergic neurons), and related groups of types referred to as subclasses (e.g., L2/3 IT). We note that many concepts discussed here were initially developed in an earlier study of mouse VISp (Tasic et al. 2016) that isn’t discussed here. A visual representation of the hierarchy of non-neuronal and endothelial cells from this study is shown below (**Figure 2**).

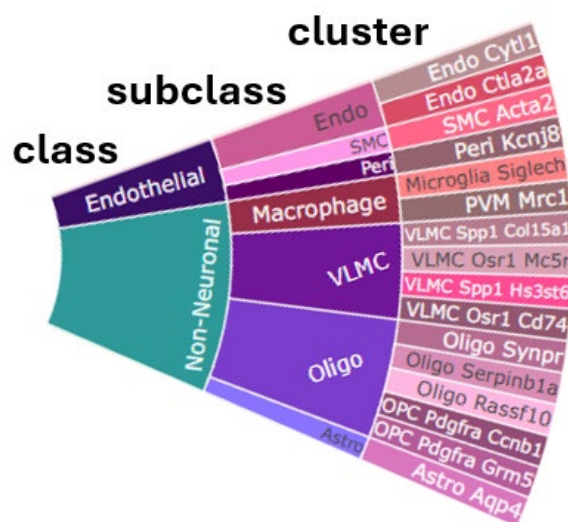


Figure 2: Visual representation of the concept of the cell type hierarchy, as applied to non-neuronal and endothelial cells from Tasic et al 2018. Multiple clusters are typically included in a single subclass, and multiple subclasses are typically included in a single class, as shown with common shared edges along the vertical arcs.

This convention organized cell type (cluster) names as a structured sequence of biologically meaningful components as follows: (1) the subclass name, which differs across broad cell classes and is described below; (2) brain region enrichment for the large subset of glutamatergic neuronal types showing preferences to VISp or ALM (otherwise, this component is skipped); (3) one to three genes which alone or in combination distinguish the labeled cell type from all other clusters in the cell type taxonomy. These marker genes were selected through a combination of computational selection and visual inspection.

The format in this paper was therefore: **[subclass] [VISp/ALM/(blank)] [one to three markers]**

For example, *L6 IT VISp Col18a1* indicates a glutamatergic neuron of the L6 IT subclass, found more frequently in VISp than ALM, and specifically marked by *Col18a1*.

Subclass naming conventions differed across broad cell classes in this study. Glutamatergic neuron names incorporated the cortical layer of the cell soma followed by the axonal projection targets from Retro-seq experiments (except for L6b), since both properties strongly structured the taxonomy. For example, the ‘L6 IT’

subclass corresponds to a group of glutamatergic neurons whose cell body is most likely to be in layer 6 of neocortex and whose projection is most likely to project intratelencephalically (e.g., within cortex). GABAergic subclasses were defined based on the most prominent canonical marker gene profile, reflecting developmental origin, with Pvalb and Sst labeling cell types derived from the medial ganglionic eminence and Vip, Sncg, and Lamp5 labeling cell types derived from the caudal ganglionic eminence. Since this study (and many early studies from the Allen Institute) focused primarily on neuronal types, non-neuronal subclass labels correspond to established non-neuronal types detected in cortex, such as astrocytes (Astro), endothelial cells (Endo), oligodendrocytes (Oligo), oligodendrocyte precursor cells (OPC), pericytes (Peri), perivascular macrophages (PVM), microglia, smooth muscle cells (SMC), and vascular leptomeningeal cells (VLMC).

Enduring nomenclature concept #1: Subclass, brain region, and marker genes are now included in cell type names in most cell type taxonomies. Similarly, the specific subclass definitions from this study are used for nearly all neocortical studies at the Allen Institute, BICCN, and BICAN (formalized in the BICCN flagship paper glossary, reproduced below; BRAIN Initiative Cell Census Network (BICCN) et al. 2021).

Name		Definition	InterLex Identifier
<i>Cell subclass:</i> Subset of class, major groupings with highly convergent evidence across data modalities			ILX:0770095
GABAergic subclasses	Lamp5	Genomic, cellular marker, phenotype and developmental origin-defined GABAergic cell sets	ILX:0770149
	Sncg		ILX:0770150
	Vip		ILX:0770151
	Sst		ILX:0770152
	Sst Chodl		ILX:0770153
	Pvalb		ILX:0770154
	Meis2		ILX:0770155
Glutamatergic subclasses	L2/3 IT	Layer 2/3 intratelencephalic-projecting	ILX:0770156
	L4/5 IT	Layer 4/5 intratelencephalic-projecting	ILX:0770174
	L5 IT	Layer 5 intratelencephalic-projecting	ILX:0770157
	L6 IT	Layer 6 intratelencephalic-projecting	ILX:0770158
	L6 IT Car3	Layer 6 intratelencephalic-projecting, Car3-expressing	ILX:0770159
	L5 ET	Layer 5 extratelencephalic-projecting	ILX:0770160
	L6 CT	Layer 6 corticothalamic	ILX:0770162
	L6b	Layer 6b neurons	ILX:0770163
Neural non-neuronal subclasses	L5/6 NP	Layer 5/6 near-projecting	ILX:0770161
	Astro	Astrocytes	ILX:0770141
	Oligo	Oligodendrocytes	ILX:0770140
	OPC	Oligodendrocyte progenitor cells	ILX:0770139
Non-neural subclasses	Endo	Endothelial cells	ILX:0770142
	VLMC	Vascular leptomeningeal cells	ILX:0770143
	SMC	Smooth muscle cells	ILX:0770144
	Peri	Pericytes	ILX:0770145
	Micro	Microglia	ILX:0770146
	PVM	Perivascular myeloid cells	ILX:0770147
<i>Cell type:</i> Subset of subclass, finest resolution clustering achieved for a modality or a consensus clustering across modalities and/or species			ILX:0770096

Table 1: Subset of ‘Extended Data Table 2: Glossary’ from (BRAIN Initiative Cell Census Network (BICCN) et al. 2021). We note that slight variations in subclass names such as changes in layer precision/identification and inconsistency in abbreviations occur between studies. For example, Tasic et al 2018 selectively separates endothelial cells from non-neuronal as shown in Figure 2.

Enduring nomenclature concept #2: Across these studies, cell types are identified through iterative clustering or related hierarchical classification approaches and then presented as nested taxonomy levels representing progressively finer subdivisions of cellular diversity. While the names used for these levels have evolved somewhat across studies, the underlying concept of taxonomy levels persists. The table below summarizes the standard levels of cell type granularity used across the studies discussed here and can serve as a reference throughout.

Standard Name	Description	Tasic 2018	Hodge 2019	Yao 2021a	Yao 2021b	Jorstad 2023a/b	Siletti 2023	Yao 2023	Gabitto 2024	Jin 2025	Gao 2025	Johansen 2026
cluster	The finest granularity of parcellation, defined entirely through a clustering analysis.	✓	✓	✓	✓	✓	sub-cluster	✓		✓	sub-cluster	✓
supertype	Highest granularity to which self-mapping is high confidence, defined computationally.				✓			✓	✓	✓	cluster	
group	Granularity at which cell types can be confidently assigned between (primate) species, defined largely computationally. May be called ‘orthotype’ in the future.											✓
subclass	Granularity of cell types typically defined in literature, with high confidence between modalities and species; defined computationally and biologically.	✓		✓	✓	✓	cluster	✓	✓	✓	✓	✓
class	Low granularity grouping of broad cell types (e.g., inhibitory neurons in neocortex), defined computationally and biologically.	✓	✓	✓	✓	✓	super-cluster	✓	✓	✓	✓	✓
neighborhood	Lowest granularity parcellation dividing the whole brain by broad region and cell type, defined largely biologically.						✓ [^]	✓ [*]				✓

Table 2: Table showing standard levels of granularity from most granular (top row) to broadest (bottom row). Checks (✓) indicate that a specific level of granularity was included in a given study (column). If a different name was used for a specific level of granularity, it is written in place of the check. [^] Siletti 2023 neighborhoods were “Neuron” and “Non-neuron”, and all analyses were performed separately for cells in the two neighborhoods. ^{*} For Yao et al 2023, neighborhood does not follow a strict hierarchy, with some cells from a single class show up in multiple neighborhoods. We note that capitalization of these hierarchy levels may differ between studies (e.g., “subclass” and “Subclass” are identical).

Hodge et al. 2019 (Human MTG snRNA-seq)

In this study (Hodge et al. 2019), cell types were defined from 15,928 single-nucleus transcriptomes collected from human middle temporal gyrus (MTG), yielding 75 transcriptomic types (or “clusters”): 24 glutamatergic, 45 GABAergic, and 6 non-neuronal types.

This convention organized cell type names as a structured sequence of biologically meaningful components, ordered from broad to specific, that was fully computationally derived. The naming scheme includes: (1) major cell class, defined by canonical marker gene expression (e.g., Inh, Exc, Microglia, Astro, OPC, Oligo, Endo); (2) cortical layer enrichment, indicating layers where at least 10% of cells in the cluster are found (e.g., L1, L2–3, L4, L5–6); (3) a curated broad marker gene selected based on maximal median expression of PAX6, LAMP5, VIP, SST, PVAlB, LINC00507, RORB, THEMIS, FEZF2, TYROBP, FGFR3, PDGFRA, OPALIN or NOSTRIN; and (4) a cluster-specific marker, representing the gene most uniquely enriched in that cluster relative to others. A few cluster names were manually adjusted to reflect known biology; for example, Microglia were initially labeled as ‘L1-3’ due to undersampling and were renamed as ‘L1-6’.

Thus, the final format for this study was: **[class] [layer(s)] [broad marker] [specific marker]**

For example, *Inh L1–2 PAX CDH12* indicates an inhibitory neuron enriched in layers 1–2, expressing PAX6 more than any other broad marker and specifically marked by CDH12.

This convention was widely used in human single-nucleus RNA-seq datasets circa 2019–2021 (Hodge et al. 2019; Bakken et al. 2021) and reflects a highly structured, semi-automated approach. We note that the concept of class in this taxonomy merges the concept of class (for neurons) and subclass (for non-neurons) from Tasic et al 2018, and that the concept of subclass relies entirely on canonical markers for all classes (not just GABAergic interneurons).

Enduring nomenclature concept: Automated curation of important gene expression and cell metadata (e.g., cortical layer) for use with cell naming and annotation. This was also the first sc/snRNA-seq study to work towards integration with Cell Ontology (Bard et al. 2005) through use of a Provisional Cell Ontology.

Recent primate snRNA-seq studies (including SEA-AD)

More recent human (and non-human primate) datasets adopt a simplified and more flexible naming scheme centered on subclass identity (Jorstad, Song, et al. 2023; Jorstad, Close, et al. 2023; Gabitto et al. 2024; Travaglini et al. 2026). In this approach (Miller et al. 2020), the primary label reflected a subclass defined by gene expression patterns (often derived from clustering in reduced-dimensional spaces like UMAP), such as L2/3 IT, SST, or Microglia. Note that these are the same subclasses defined in Tasic et al 2018 (shown above) and are inferred by integrating human and mouse datasets and transferring labels from mouse to human. In some cases (e.g., “chandelier”), historical names are used to replace quantitatively derived subclass labels.

To distinguish multiple clusters within a subclass, an arbitrary numeric suffix is appended. Importantly, these numbers are study-specific and do not necessarily correspond across datasets (e.g., “SST 3” in one study may not match “SST 3” in another). In cases where number assignments occur prior to finalizing quality control steps, some numbers may be skipped.

An optional disease indicator has also been included in some studies. For instance, in SEA-AD, cell types that are specific to aging or disease conditions are tagged with a suffix such as “-SEAAD”. We note that each cluster ending in -SEAAD is distinct from each cluster without the suffix, even if they have the same ID number. However, cell type taxonomies in SEA-AD are linked: cell types in Travaglini et al 2026 and Gabitto et al 2024 with common names represent the same cell types—the newer study extended the MTG-based study to additional brain regions.

The final format for this convention was: **[subclass]_[number]([blank])/-SEAAD]**

Example: *Astro_6-SEAAD* represents a specific astrocyte cluster identified predominantly in an aged human population but not in the corresponding healthy human brain reference.

This convention prioritizes simplicity and scalability, though at the expense of cross-study alignment and interpretability of numeric identifiers. It has been used in multiple studies of primary motor cortex (M1) and MTG, although more recent studies in primate have shifted back towards more informative names with considerations taken for cross-species alignment.

Enduring nomenclature concept: Use of numeric identifiers to distinguish multiple cell types of the same subclass. Current studies attempt to balance overloading data set-specific content (as in Hodge et al 2019) vs. the overly simplistic nomenclature used in these studies.

Mouse cortical studies circa 2021

Studies focused on mouse motor cortex (MOp) and mouse neocortex + hippocampus published in 2021 combined the concepts presented above in their naming conventions (Yao, Liu, et al. 2021; Yao, van Velthoven, et al. 2021). For example, the great majority of cell types in the mouse MOp study use the following format: **[subclass] [marker gene][([blank)]/_number]**. For the mouse neocortex + hippocampus study, clusters were labeled as **[number]_[subclass*]** while supertypes (slightly lower granularity) were labeled as **[subclass*]_[one to two marker genes]**. *Technically, this study used the format **[number]_[supertype]**, where in most (but not all) cases supertype corresponded to **[Subclass]_[marker genes(s)]**, and where ‘supertype’ is a level of granularity in the cell type taxonomy between ‘cluster’ and ‘subclass’. In this study, ‘subclass’ and ‘supertype’ include the original concept of subclass from Tasic et al 2018 as well as the brain region ontology term indicating where a cell type is primarily derived. In cases where supertypes are more spatially localized than subclasses, a more granular brain region term is used for the supertype.

Enduring nomenclature concept: The specific format of cluster names in the mouse neocortex + hippocampus study was also used for cluster names in the whole mouse brain study described next, with the caveat that additional information is embedded in the subclass.

Whole mouse brain

The whole mouse brain (WMB) atlas is a high-resolution transcriptomic and spatial cell-type atlas across the entire mouse brain, integrating several whole-brain single-cell RNA-sequencing (scRNA-seq) datasets (Yao et al. 2023). The datasets contain a total of ~4 million cells passing rigorous quality-control (QC) criteria. The integrated transcriptomic taxonomy contains 5,322 clusters that are organized in a hierarchical manner with nested groupings of 34 classes, 338 subclasses, 1,201 supertypes and 5,322 types/clusters. The scRNA-seq data reveal transcriptome-wide gene expression and co-expression patterns for each cell type. The anatomical location of each cell type has been annotated using a comprehensive brain-wide MERFISH dataset with a total of ~4 million segmented and QC-passed cells, probed with a 500-gene panel and registered to the Allen Mouse Brain Common Coordinate Framework (CCF v3).

The whole mouse brain taxonomy adopts a hierarchical and modular naming framework designed to be intuitive and extensible. A key principle is that higher-level categories are conserved across development and evolution, while lower levels capture increasing specificity.

- **Class:** Broad grouping defined by region and neurotransmitter (e.g., 19 MB Glut).
- **Subclass:** Combines anatomical location, projection patterns, marker genes, features, and neurotransmitter (e.g., 007 L2/3-IT-CTX_Glut, 051 Pvalb-chandelier_Gaba). Some components may be omitted if not applicable. More details below.
- **Supertype:** Following the format of the mouse neocortex + hippocampus study, appends a numeric prefix (and if needed, suffix) to subclasses (e.g., 0029 L2/3-IT-CTX_Glut 1, 1181 COP NN_1). Future plans include replacing numeric identifiers with more descriptive markers and refined anatomical information.
- **Cluster:** Same as supertype, but with distinct leading prefixes (e.g., clusters 0103 L2/3-IT-CTX_Glut 1 and 0104 L2/3-IT-CTX_Glut 1 are children of 0029 L2/3-IT-CTX_Glut 1).

After breaking up the subclass components, supertype and cluster names followed the following scheme: **[number] [layer(s)*] [projection target(s)*] [anatomic structure(s)* (dash separated)] [0-3 marker gene(s)*] [common cell type name*]_[number]**, where every *ed component is optional so long as at least one is present. *We note that abbreviations are used for a large majority of the naming components; definitions for each abbreviation are available both through the [ABC Atlas application](#) through ABC Atlas Assistant or the “Learn More” panel.*

This framework emphasizes scalability and cross-dataset consistency, but introduces challenges. Increasing specificity through marker genes and anatomical descriptors can make names long (for example, "2959 SPA-SPFm-SPFp-POL-PIL-PoT Sp9 Glut_4" includes abbreviations for six anatomical structures). Additionally, fine-grained anatomical labeling requires precise spatial mapping and high-resolution reference atlases (e.g., CCF), which is possible in mouse and smaller primate species, and is under exploration through ongoing methods development in human.

Enduring nomenclature concept: The concept of embedding the available and most biologically relevant context into the names will continue moving forward. In particular, inclusion of anatomic localization based on spatial transcriptomic context is first possible in this study.

Aging and developing mouse brain

A recent study of aging mouse (Jin et al. 2025) follows an adapted cell type taxonomy from the whole mouse brain taxonomy (Yao et al. 2023). All cells were annotated at class, subclass, supertype, and cluster levels; the class, subclass, and supertype labels were inherited from the WMB taxonomy, while the cluster labels were newly defined for this aging study, capturing 172 of 338 WMB subclasses and resolving them into 434 superotypes and 847 study-specific clusters. The cluster naming convention uses a study-specific numeric cluster ID together with the corresponding supertype-style label — for example, 331_TU-ARH Otp Six6 Gaba.

The developing mouse visual cortex study is derived from a single cell transcriptomic dataset containing 568,654 cells from donors ranging in age from embryonic to adult (Gao et al. 2025). The cell-type assignment is a four-level taxonomy defining 15 classes, 40 subclasses, 148 clusters and 714 subclusters. This cell-type assignment is broadly consistent with those from the previous studies from whole mouse brain (Yao et al. 2023) and the Broad Institute whole mouse brain taxonomy (Langlieb et al. 2023) at the subclass level while providing finer cell-type and temporal resolutions with additional subcluster annotations. More specifically the naming approach differs from the adult whole mouse brain taxonomy by emphasizing developmental continuity and diversification over time. Subclass-level labels are kept largely compatible with prior adult taxonomies (e.g., L2/3 IT, Lamp5, pericytes), but finer levels are used to capture transient developmental states, branch points, and postnatal refinement of cell

identities. This is important because the study finds that most broad classes and subclasses emerge early, whereas more refined cell types continue to diversify throughout postnatal development, including around eye opening and the onset of the critical period. More specifically, for types found in adult, cluster names directly align with cluster names from the consensus mouse whole brain taxonomy (e.g., 428_L6b CTX Glut_2), while types found only in development are named based on abbreviations for that type (e.g., NEC, IMN CR, RG, IMN, IT Deep Layer). Subclusters are defined as [cluster name]_[unique ID] in all cases (e.g., RG_1, RG_2).

Whole human brain

The Whole Human Brain Atlas study (Siletti et al. 2023) contains ~3 million nuclei (more than 2 million sequenced neurons and almost 1 million non-neurons) from 100 locations across the forebrain, midbrain, and hindbrain. The integrated transcriptomic taxonomy contains 31 superclusters, 461 clusters, and 3313 subclusters. This study used a minimalist approach for naming, instead relying on extensive, well-organized metadata for cluster annotation. Superclusters (called “class” in other taxonomies) are named based on the overall types of cells included; for neurons, this typically includes a brain structure and/or type of cells (e.g., “Thalamic excitatory”), while for non-neurons, only the cell type is included (e.g., “Microglia”). “Miscellaneous” and “Splatter” superclusters correspond to sets of cells not easily categorized due to their distinct nature and/or because they are undersampled in this study. Clusters (called “subclass” in other studies) are labeled with an abbreviated version of the super-cluster term followed by a unique identifier (e.g., “Thex_445” and “Mgl_4”), while subclusters (called “cluster” in other studies) appends another unique ID to the end of the cluster label (e.g., “Thex_445_122” and “Mgl_4_3267”). This convention ([**supercluster**]_[**unique ID(s)**]) will not be used for any future study, as more information on cell localization and cross-species alignment becomes available, and more precise nomenclature is possible.

Cross-species basal ganglia

Human and Mammalian Brain Atlas (HMBA) is a major atlas of the BRAIN Initiative Cell Atlas Network (BICAN) that proposes to establish a comprehensive, highly granular cell atlas in complete adult human, macaque, and marmoset brains that links brain structure, function and cellular architecture. Our consensus basal ganglia cell type taxonomy (Johansen et al. 2025) is the result of iterative clustering and cross-species integration of transcriptomic data across key structures within the basal ganglia. By combining data from multiple primate and rodent species, we have developed a consensus taxonomy that highlights both conserved and species-specific cell types. The taxonomy includes 4 neighborhoods, 12 classes, 36 subclasses, and 61 homologous ‘Groups’ (corresponding to a level of granularity between cluster/supertype and subclass in other taxonomies), along with several hundred species-specific clusters per species.

The basal ganglia ‘group’ naming scheme uses a structured but flexible set of components that integrate anatomical, physiological, historical, and transcriptomic information. The convention is designed to harmonize modern molecular classifications with established terminology. Specifically, these components include:

- **Location:** Standardized anatomical terms from the BICAN Atlas, with controlled abbreviations.
- **Physiology (optional):** Functional characteristics such as firing properties (e.g., FS for fast-spiking).
- **Literature (optional):** Historical or previously published names for the cell type.

- **Transcriptomics:** Gene expression-based classification, typically including one broad marker and one more specific marker (e.g., NMU-TAC3).
- **Neurotransmitter:** Primary neurotransmitter identity (e.g., GABA, glutamate, dopamine).

The final format is: **[Location] [Physiology] [Literature] [Transcriptomics] [Neurotransmitter]**

Example: *STR FS PTHLH ST18 GABA* represents a specific population of cells in striatum (STR) that is fast spiking (FS), expresses the genes *PTHLH* and *ST18* and which is GABAergic.

This convention allows integration of legacy knowledge with modern transcriptomic insights, while maintaining a standardized and interpretable structure.

Naming conventions for class and neighborhood reflect similar conventions from other taxonomies for cell types at these lower levels of granularity, largely corresponding to broad brain region, cell type(s), and neurotransmitter, with any combination of those things included in that order, when present. Subclass nomenclature follows this same pattern but also sometimes includes marker genes as with group. Single species clusters are not annotated and instead are given a [species]-[unique ID] nomenclature.

Reconciling Multiple Names for the Same Cell Type

Cell types may acquire multiple names as they are characterized using different anatomical, morphological, physiological, projection-based, and molecular approaches. As knowledge accumulates, nomenclature may be standardized around a term that more accurately captures the defining features of a cell population, sometimes replacing terminology used in earlier taxonomies. The following examples describe two such changes in Allen Institute-hosted cell type taxonomies and the biological rationale underlying each transition. Similar revisions will likely occur as cell types are characterized more comprehensively and naming conventions continue to evolve.

Pyramidal Tract (PT) vs. Extratelencephalic (ET) neocortical neurons

Early Allen Institute studies referred to a distinct population of long-range projection neurons as layer 5 pyramidal tract (PT) neurons. More recent studies have adopted the term layer 5 extratelencephalic (L5 ET) because evidence from transcriptomics, anatomy, and projection mapping showed that PT neurons constitute only a subset of a broader, evolutionarily conserved subclass of layer 5 excitatory neurons. The term PT was historically applied primarily to motor-cortex neurons that project through the pyramidal tract to the brainstem or spinal cord. However, transcriptomically related L5 neurons in sensory and other cortical areas do not necessarily project to the pyramidal tract. Even within motor cortex, some neurons belonging to this transcriptomic subclass project to the thalamus or other extratelencephalic targets rather than through the pyramidal tract. Single-cell transcriptomic studies further showed that these neurons form a coherent subclass across cortical areas and species. L5 ET therefore provides a more inclusive designation, shifting the emphasis from a single projection pathway to the broader subclass defined by shared molecular identity and extratelencephalic projection patterns. This transition is described in greater detail in the Methods of (BRAIN Initiative Cell Census Network (BICCN) et al. 2021).

Medium spiny neuron (MSN) vs. spiny projection neuron (SPN)

The terms medium spiny neuron (MSN) and spiny projection neuron (SPN) refer to the principal projection neurons of the striatum. MSN arose from early morphological observations in which soma size and dendritic appearance were used to distinguish these cells from larger aspiny neurons. As knowledge of striatal circuitry developed, the descriptor “medium” became less informative because intermediate soma size is neither unique to nor defining of this population. In contrast, these cells’ identity as the principal and exclusive projection neurons of the striatum represents a fundamental and consistent feature. SPN therefore captures both their characteristic spiny morphology and their defining role in basal ganglia circuits. It also supports commonly used pathway terminology, including direct-pathway SPNs (dSPNs) and indirect-pathway SPNs (iSPNs). Although MSN and SPN remain widely recognized synonyms, SPN is adopted here as the more biologically precise and functionally informative designation. This transition is described in greater detail in the Supplemental Materials of (Johansen et al. 2025).

Enduring nomenclature concept: Cell type names may be revised when accumulating evidence demonstrates that an established term describes only a subset of the relevant population or emphasizes a feature that is not biologically defining.

Future Work

Moving forward, naming conventions will likely integrate elements from both the whole mouse brain and basal ganglia frameworks—combining hierarchical structure, standardized anatomical terminology, and transcriptomic specificity. Efforts are ongoing to improve cross-study alignment, replace arbitrary numeric identifiers with biologically meaningful descriptors, and incorporate increasingly precise spatial and multimodal information.

This remains an active area of development, and future iterations of these naming systems will continue to evolve alongside advances in data generation and integration.

Acknowledgements

This white paper was written using the assistance of Microsoft Copilot. All content was confirmed and all references manually added by Allen Institute researchers.

References

Ascoli, Giorgio A., Lidia Alonso-Nanclares, Stewart A. Anderson, et al. 2008. “Petilla Terminology: Nomenclature of Features of GABAergic Interneurons of the Cerebral Cortex.” *Nature Reviews Neuroscience* 9 (7): 557–68.
<https://doi.org/10.1038/nrn2402>.

- Bakken, Trygve E., Nikolas L. Jorstad, Qiwen Hu, et al. 2021. “Comparative Cellular Analysis of Motor Cortex in Human, Marmoset and Mouse.” *Nature* 598 (7879): 111–19. <https://doi.org/10.1038/s41586-021-03465-8>.
- Bard, Jonathan, Seung Y. Rhee, and Michael Ashburner. 2005. “An Ontology for Cell Types.” *Genome Biology* 6 (2): R21. <https://doi.org/10.1186/gb-2005-6-2-r21>.
- BRAIN Initiative Cell Census Network (BICCN), Edward M. Callaway, Hong-Wei Dong, et al. 2021. “A Multimodal Cell Census and Atlas of the Mammalian Primary Motor Cortex.” *Nature* 598 (7879): 86–102. <https://doi.org/10.1038/s41586-021-03950-0>.
- Gabitto, Mariano I., Kyle J. Travaglini, Victoria M. Rachleff, et al. 2024. “Integrated Multimodal Cell Atlas of Alzheimer’s Disease.” *Nature Neuroscience* 27 (12): 2366–83. <https://doi.org/10.1038/s41593-024-01774-5>.
- Gao, Yuan, Cindy T. J. van Velthoven, Changkyu Lee, et al. 2025. “Continuous Cell-Type Diversification in Mouse Visual Cortex Development.” *Nature* 647 (8088): 127–42. <https://doi.org/10.1038/s41586-025-09644-1>.
- Hodge, Rebecca D., Trygve E. Bakken, Jeremy A. Miller, et al. 2019. “Conserved Cell Types with Divergent Features in Human versus Mouse Cortex.” *Nature* 573 (7772): 61–68. <https://doi.org/10.1038/s41586-019-1506-7>.
- Jin, Kelly, Zizhen Yao, Cindy T. J. van Velthoven, et al. 2025. “Brain-Wide Cell-Type-Specific Transcriptomic Signatures of Healthy Ageing in Mice.” *Nature* 638 (8049): 182–96. <https://doi.org/10.1038/s41586-024-08350-8>.
- Johansen, Nelson J., Yuanyuan Fu, Matthew Schmitz, et al. 2025. “Cross-Species Consensus Atlas of the Primate Basal Ganglia.” Preprint, bioRxiv, December 16. <https://doi.org/10.64898/2025.12.15.694496>.
- Jorstad, Nikolas L., Jennie Close, Nelson Johansen, et al. 2023. “Transcriptomic Cytoarchitecture Reveals Principles of Human Neocortex Organization.” *Science* 382 (6667): eadf6812. <https://doi.org/10.1126/science.adf6812>.
- Jorstad, Nikolas L., Janet H. T. Song, David Exposito-Alonso, et al. 2023. “Comparative Transcriptomics Reveals Human-Specific Cortical Features.” *Science* 382 (6667): eade9516. <https://doi.org/10.1126/science.ade9516>.
- Langlieb, Jonah, Nina S. Sachdev, Karol S. Balderrama, et al. 2023. “The Molecular Cytoarchitecture of the Adult Mouse Brain.” *Nature* 624 (7991): 333–42. <https://doi.org/10.1038/s41586-023-06818-7>.
- Miller, Jeremy A., Nathan W. Gouwens, Bosiljka Tasic, et al. 2020. “Common Cell Type Nomenclature for the Mammalian Brain.” *eLife* 9 (December): e59928. <https://doi.org/10.7554/eLife.59928>.

- Ray, Patrick L., Jeremy A. Miller, Dorota Jarecka, et al. 2026. “A Layered Standards Framework for Integrating Single-Cell and Spatial Omics Data into Brain Cell Atlases.” Preprint, bioRxiv, May 4. <https://doi.org/10.64898/2026.04.30.722039>.
- Siletti, Kimberly, Rebecca Hodge, Alejandro Mossi Albiach, et al. 2023. “Transcriptomic Diversity of Cell Types across the Adult Human Brain.” *Science* 382 (6667): eadd7046. <https://doi.org/10.1126/science.add7046>.
- Tasic, Bosiljka, Vilas Menon, Thuc Nghi Nguyen, et al. 2016. “Adult Mouse Cortical Cell Taxonomy Revealed by Single Cell Transcriptomics.” *Nature Neuroscience* 19 (2): 335–46. <https://doi.org/10.1038/nn.4216>.
- Tasic, Bosiljka, Zizhen Yao, Lucas T. Graybuck, et al. 2018. “Shared and Distinct Transcriptomic Cell Types across Neocortical Areas.” *Nature* 563 (7729): 72–78. <https://doi.org/10.1038/s41586-018-0654-5>.
- Travaglini, Kyle J., Mariano I. Gabitto, Yi Ding, et al. 2026. “Multiregional Single-Cell Profiling Reveals Shared and Specialized Cellular Vulnerability in Alzheimer’s Disease.” Preprint, bioRxiv, July 2. <https://doi.org/10.64898/2026.07.01.734821>.
- Yao, Zizhen, Hanqing Liu, Fangming Xie, et al. 2021. “A Transcriptomic and Epigenomic Cell Atlas of the Mouse Primary Motor Cortex.” *Nature* 598 (7879): 103–10. <https://doi.org/10.1038/s41586-021-03500-8>.
- Yao, Zizhen, Cindy T. J. van Velthoven, Michael Kunst, et al. 2023. “A High-Resolution Transcriptomic and Spatial Atlas of Cell Types in the Whole Mouse Brain.” *Nature* 624 (7991): 317–32. <https://doi.org/10.1038/s41586-023-06812-z>.
- Yao, Zizhen, Cindy T. J. van Velthoven, Thuc Nghi Nguyen, et al. 2021. “A Taxonomy of Transcriptomic Cell Types across the Isocortex and Hippocampal Formation.” *Cell* 184 (12): 3222–3241.e26. <https://doi.org/10.1016/j.cell.2021.04.021>.
- Zeng, Hongkui, and Joshua R. Sanes. 2017. “Neuronal Cell-Type Classification: Challenges, Opportunities and the Path Forward.” *Nature Reviews Neuroscience* 18 (9): 530–46. <https://doi.org/10.1038/nnrn.2017.85>.