Cell type nomenclature for the mammalian brain: Development and application of a systematic, extensible convention

Authors:
Jeremy A Miller¹,*
Nathan W Gouwens¹
Bosiljka Tasic¹
Forrest Collman¹
Cindy T.J. van Velthoven¹
Trygve E Bakken¹
Hongkui Zeng¹
Ed S Lein¹
Amy Bernard¹,*

¹ Allen Institute for Brain Science | Seattle, WA, USA
* Correspondence: jeremym@alleninstitute.org; amyb@alleninstitute.org

Abstract

The recent advent of single cell RNA-sequencing and other high-throughput technologies has led to an explosion of cell type definitions across multiple organ systems. Consortia like the BRAIN Initiative Cell Census Network (BICCN) and the Human Cell Atlas (HCA) have begun to standardize and centralize the intake of data and associated metadata from these projects; however, the naming and organization of cell types has largely been left to individual investigators, resulting in widely varying nomenclature and limited alignment between taxonomies derived from overlapping datasets. To facilitate cross-dataset alignment and comparison, we created a working nomenclature convention for matching and tracking cell types across studies. The convention is qualitatively similar to how gene transcripts are tracked across different versions of GENCODE genome builds, allowing comparison of matched types with a common reference or any other taxonomy. It augments but does not change existing cell type names provided in original publications and can be directly applied to data from new or published studies. We applied this nomenclature to a series of published cell type data from the Allen Institute, highlighting its extensibility to classifications defined using multiple quantifiable modalities. Finally, we propose conventions for assigning accurate yet flexible cell types names in the mammalian cortex as a step towards a much larger community-wide effort to organize multi-source, data-driven information related to cell types. The convention presented here can be easily applied to cell type taxonomies from any organ system or organism.
Main text

Introduction

Cell type classification has been central to understanding biological systems for many tissues (e.g., immune system) (Lees et al., 2015) and organisms (e.g., C. elegans) (Packer et al., 2019). Identifying and naming anatomic cellular components of the brain has been an integral part of neuroscience since the seminal work of Ramon y Cajal (Cajal, 1899). Many neuronal cell types such as neurogliaform, chandelier, Martinotti, and pyramidal cells, have been identified based on highly distinct shape, location, or electrical properties, providing a common vocabulary for neuroscientists. These distinct morphological and electrophysiological features have provided the basis for robust and consistent classifications of neuronal cell types (Greig et al., 2013; Markram et al., 2004). However, the recent application of high-throughput, quantitative methods such as single cell or nucleus transcriptomics (scRNA-seq) (Hodge et al., 2019; Tasic et al., 2018; Zeisel et al., 2015), electron microscopy (Zheng et al., 2018), and whole brain morphology (Winnubst et al., 2019) to cell type classification is enabling more quantitative measurements of similarities among cells and construction of taxonomies (Zeng and Sanes, 2017). These methodological advances are ushering a new era of data-driven classification, by simultaneously expanding the number of measurable features per cell, the number of cells per study, and the number of classification studies, and the computational resources for storing and analyzing this information. The use of scRNA-seq, in particular, for cell type classification has increased exponentially since its introduction a decade ago (Tang et al., 2009), with approximately 2,000 published studies using such technology, and several hundred tools for data analysis (Zappia et al., 2018).

This explosion of data has vastly expanded our collective understanding of biological cell types, while simultaneously introducing challenges in cell type classification. For example, in retina, neurons with common morphology and function also have common connectivity (Jonas and Kording, 2015), spacing, arbor density, arbor stratification (Seung and Sümbül, 2014), and gene expression signatures (Macosko et al., 2015), often with one-to-one correspondences between modalities (Zeng and Sanes, 2017). However, studies combining scRNA-seq with traditional morphological and electrophysiological characterizations have found a more complicated relationship in the brain than in retina, with cell types defined by morphology and electrophysiology sometimes containing cells from several cell types defined using gene expression (Gouwens et al., 2020), and some transcriptomically-defined types containing cells with multiple morphologies (Hodge et al., 2020, 2019). Further complicating classification is the overlay of discrete cell type distinctions with more continuous variations associated with cortical depth (Berg et al., 2020), brain region (Hawrylycz et al., 2012), activity-dependent cell state (Wu et al., 2017), or all of these things at once (Yao et al., 2020b). Finally, many features in matched cell types diverge between species (Bakken et al., 2020; Berg et al., 2020; Boldog et al., 2018; Hodge et al., 2019), and as cells advance along trajectories of development (Nowakowski et al., 2017), aging (The Tabula Muris consortium et al., 2019), and disease (Mathys et al., 2019).

To extend historical cell type classifications to the era of data-driven science will require community-accepted strategies for standardized quantitative cell classification, cell type naming, and data and metadata organization. Here we lay out the challenges in doing so and present a generalizable nomenclature convention to address some of these challenges in the mammalian cortex that is immediately extensible to other organ systems and data organization efforts.
Challenges for cell type classification

To advance from low-throughput, phenotype-based to data-driven, principled approaches of cell type classification will require a change of mindset from experiment-based to community-based science, and an accompanying set of tools to facilitate this change. Within a single study, cell types need to be quantitatively defined, organized, named, and annotated by comparison to published work and cell metadata. Data and resultant classifications from multiple studies then need to be unified into community-accepted reference taxonomies with standard nomenclature, linked to common ontologies, and stored in standard databases. Achieving this transition in cell type classification will require overcoming several distinct but overlapping challenges.

First, how do we quantitatively define cell types in an individual study? For essentially all published studies, clusters (or provisional cell types; see below) are defined using hierarchical clustering, k-means clustering, machine learning, or related mathematical techniques (Table 1 includes definitions for key terms, bold on first use). This process is often preceded by selection of variable features followed by some form of dimensionality reduction, such as principal component analysis (PCA), Uniform Manifold Approximation and Projection (UMAP) (McInnes et al., 2018), or a combination of these. Resulting clusters are often presented as “cell types,” though likely represent a combination of cell types, cell states (such as cell cycle or cell activation) (Kotliar et al., 2019), and the intersection of discrete cell types with more continuous variation (Cembrowski and Menon, 2018; Yao et al., 2020b). Furthermore, the number of cell types presented is highly dependent on the number of cells assayed, the experimental platform, and the specific clustering method and associated parameter settings chosen (Yao et al., 2020a). The resulting cell types are then typically organized into a taxonomy, defined as the output of a computational algorithm applied to a specific dataset, where the distance of each pair of cell types can be measured quantitatively, and cell types with smaller distances are placed nearby.

Second, how do we annotate cell types using experimental and historical context? In most experiments, some information about the original cell location and donor of origin is available, providing some context for cell types, such as cortical layer of origin (Hodge et al., 2019). In mouse and other model systems, transgenic lines allow further refinement based on expression of marker genes and direct measurements of cellular projection targets (Madisen et al., 2015, 2010; Tasic et al., 2018). In some cases, clusters can then be linked with well-established historical types captured in a cell type ontology, or a structured controlled vocabulary for grouping information, which has a specific semantic structure and is built based on extensive curation of biological evidence. The Cell Ontology (CL) (Diehl et al., 2016) already organizes cells from the immune system and other well-studied organ systems into an ontological structure, and ongoing efforts seek to extend this ontology to encompass well-established brain cell types (Aevermann et al., 2018). Ontologies can provide a starting point for building an information framework to (1) promote data exchange and synthesis across groups, (2) provide a controlled vocabulary and standard metadata, (3) allow for interaction between data and metadata, and (4) provide a baseline for expansion for ongoing data-driven cell typing efforts (Gillespie et al., 2019; Larson and Martone, 2009).

A third challenge in cell type classification is, how do we compare and match cell types identified in different studies? Great progress has been made in addressing this question of “alignment”, whereby datasets collected using gene-based modalities such as scRNA-seq and epigenetics can be placed in a common space (Barkas et al., 2018; Butler et al., 2018; Johansen and Quon, 2019). These strategies provide a mechanism for defining joint taxonomies and allow for annotation transfer between experiments. Alignment has proven effective for matching cell types collected on different sequencing platforms, across multiple data modalities, and even between species where few marker genes show conserved patterns (Bakken et al., 2020; Hodge et al., 2020, 2019; Yao et al., 2020a). Autoencoder-based computational strategies (Gala et al.,
can extend these alignments to bridge distinct modalities, as long as some linking cells with simultaneous measurement of relevant modalities are included. This is possible through experimental methods such as Patch-seq (Cadwell et al., 2016; Fuzik et al., 2016; Gouwens et al., 2020; Scala et al., 2020).

Fourth, once quantitative clusters are defined in a new study, how should those cell types be named in a way that provides historical or comparative context? Currently no standard convention of naming brain cell types is widely followed. Cell types have historically been named by their shape, location, electrical properties, selective neurochemical markers, or even the scientist that discovered them (Betz, 1874; Szentágothai and Arbib, 1974). Now, quantitative clusters that cannot obviously be matched with these types are named on an ad hoc basis. For example, some investigators use generic names like “interneuron 1” or “Ex1” and then use figures, tables, or text to link these clusters to historical cell types (Gouwens et al., 2019; Lake et al., 2016; Zeisel et al., 2015). Others advocate for chaining critical cell type features in the name itself, resulting in names like “Neocortex M1 L6 CT pyramidal, Zfp3 non-adapt GLU” (Shepherd et al., 2019). This Allen Institute assigns semi-automated names that reference marker genes and metadata from a given dataset, such as “Exc L5–L6 THEMIS C1QL3” in human and “L5 IT ALM Pdf5” in mouse cortex, respectively (Hodge et al., 2019; Tasic et al., 2018). All these proposals are reasonable for stand-alone projects but make direct comparisons between studies daunting.

Finally, what data/metadata infrastructure is required to achieve these goals of cell type classification in a scalable way? While several public databases for data storage exist (e.g., dbGaP, NeMO, NeuroElectro, Neuromorpho, etc.), a community-recognized repository for storing and tracking cell type assignments and associated taxonomies is not available. This challenge has been recognized by many (Armañanzas and Ascoli, 2015; DeFelipe et al., 2013; Shepherd et al., 2019) and was the focus of a recent conference in Copenhagen that brought together many experts in the field and led to a call to the community for participation in a solution (Yuste et al., 2019). Any solution devised to tackle this question should ideally be user-friendly, so people will actually use it, and should directly address some of the ongoing challenges of ontology, data matching, and cell type naming described above in its implementation, providing some amount of immediate standardization of any cell type classifications included therein.

To address the latter two challenges and as a practical need to organize vast amounts of multimodal data from the Allen Institute and collaborators, we have developed a nomenclature convention aimed at tracking cell type information across multiple data sets. This convention was designed to parallel how gene transcripts are tracked across different versions of GENCODE genome builds, allowing comparison of matched types with a common reference or any other taxonomy. While this work is our own, it was heavily influenced by A Cell Type Ontology Workshop (Seattle, June 17-18, 2019; hosted by the Allen Institute in collaboration with the Chan Zuckerberg Initiative (CZI) and the National Institutes of Health (NIH)), which convened representatives from the fields of ontology, taxonomy, and neuroscience to make recommendations, highlight best practices and propose conventions for naming cell types. The initial version of this convention was introduced publicly in October 2019 at https://portal.brain-map.org/explore/classes, with the goal that this information will spur discussion and improvement for subsequent versions, to facilitate collaboration, and improve shared understanding of the many cell types in the brain.

Overview of proposed nomenclature convention

The problem of defining and naming cell types has many similarities to those of genes in genomics, where there is a practical need to track individual sequencing and assembly results as distinct and self-contained entities, while simultaneously recognizing that the ultimate goal is to
have a singular reference that the community can use to put their sequencing results into a common context (Frankish et al., 2019; Harrow et al., 2012; Kitts et al., 2016). Here, a similar strategy is proposed for cell type nomenclature that uses a standardized series of identifiers for tracking cell types in individual studies, in addition to providing a mechanism for saving common identifiers. At the core of the proposal are two key concepts: a taxonomy (defined above and in Table 1), which must be generated prior to implementation of this schema, and a cell set, which can represent any collection of cells within a specific taxonomy. Provisional cell types (sometimes called cell types for convenience) are particularly important cell sets that represent quantitatively-derived data clusters defined by the classification algorithm that generated the taxonomy. Provisional cell types can be organized as the terminal leaf nodes of a hierarchical taxonomy using a dendrogram, or as a non-hierarchical community structure. Taxonomies and cell sets are assigned unique identifier tags, as described below, and additional taxonomy and cell set metadata can be stored alongside these tags for use with future databasing and ontology tools. In principle, these properties can be tracked in a qualitatively similar manner to how transcripts are tracked in different versions of GENCODE genome builds (Frankish et al., 2019).

A major goal of this nomenclature convention is to track taxonomies. Analysis of each dataset will create a taxonomy of cell clusters that needs to be tracked. Since a taxonomy is defined both by the data set and the analysis method applied to that data set, it is likely that multiple taxonomies will be generated for the same cells, as data sets grow and analysis methods improve. Thus, this convention needs to account both for taxonomies from unique data sets and for versioned taxonomies using overlapping cells. To achieve this goal, each taxonomy is assigned a unique taxonomy id of the format CS[YYMMDD][#], where “CS” stands for cell set; Y, M, and D represent year, month, and day; and # is an index for compiling multiple taxonomies on a single day. Although many published quantitative datasets for cell type classification use scRNA-seq and present hierarchal taxonomies, this schema is compatible with taxonomies defined on any data type using any classification algorithm. Furthermore, taxonomies can be defined using a single dataset or through alignment of multiple datasets, and high confidence, high resolution taxonomies can be compiled into a reference taxonomy, along the lines of how select genome builds are wrapped up as public releases (Frankish et al., 2019). The figures below present examples of different types of taxonomies to highlight these possibilities.

Within each taxonomy, cell sets (and therefore also provisional cell types) are assigned multiple identifier tags, which are used for different purposes. Cell set accession IDs track unique cell sets across the entire universe of taxonomies and are defined as [taxonomy id][unique # within taxonomy]. Cell set labels provide general tags that allow easy tracking of groups of more broadly-defined cell types (e.g., “neuron”) within large taxonomies, and are defined as [Canonical cell type][unique # within canonical cell type]. Ideally these tags would match defined terms (e.g. common usage types) in CL or other existing cell type ontologies. Cell set aliases include any number of cell set descriptors intended for public consumption, including any cell type names used in manuscripts. These can include data-driven terms, historical names, or more generic cell type nomenclature. This system for aliasing provisional cell types in the mammalian cortex uses properties that are predicted to be largely preserved across development, anatomical area, and species. Finally, each cell set can have at most one preferred alias, which is a biologically-driven term that is selected from a controlled vocabulary, and that only can be assigned for cell sets with confident matches across taxonomies. Preferred aliases will be discussed in more detail below but represent what might be conventionally considered when discussing the "cell set name." Application of this schema aims at providing a set of standardized terms and files that are immediately useful, but also already formatted to seed any future centralized, searchable repository for this information, such as NIH Hubmap (https://commonfund.nih.gov/hubmap) or other in-process community endeavors described below. Finally, this framework seeks to either implement or set up future implementation of FAIR principles by making taxonomies and cell sets
Findable, Accessible, Interoperable and Reusable, both for machines and for people (Wilkinson et al., 2016).

Table 1: Glossary of terms. A glossary of broad terms, along with their definitions for the purposes of use here, and examples of how the terms are used (when relevant). These terms are presented in bold upon first use in the text. This table is provided since these terms may be open to multiple interpretations and classification requires disambiguation.

| Term                        | Definition                                                                 | Example                                                                 |
|-----------------------------|---------------------------------------------------------------------------|-------------------------------------------------------------------------|
| Taxonomy                    | Set of quantitatively-derived data clusters defined by running a specific computational algorithm on a specific dataset. Taxonomies are given a unique label and can be annotated with metadata about the taxonomy, including details of the algorithms, and relevant cell and cell set IDs. | -Any clustering result in a cell type classification manuscript          |
| Ontology                    | A structured controlled vocabulary for cell types.                         | -Cell Ontology                                                          |
| Marker gene(s)              | A gene (or set of genes) which, when expressed in a cell, can be used to accurately assign that cell to a specific cell set. | -GAD1 -PVALB -CHODL                                                     |
| Taxonomy ID                 | An identifier uniquely tagging a taxonomy of the format CS[YYMMDD][#].     | -CS191012                                                               |
| Cell                        | A single entry in a taxonomy representing data from a single cell (or cell compartment, such as the nucleus). Cells have meta-data including a unique ID. | N/A                                                                     |
| Cell set                    | Any tagged group of cells in a taxonomy. This includes cell types, groups of cell types (e.g., subclass or class), and potentially other informative groupings (e.g., all cells from a particular donor, organ, cortical layer, or transgenic line). Cell sets have a number of IDs and descriptors (as discussed below) and can also have other meta-data. | -A cell type -A group of cell types -All cells from layer 2 in MTG -All cells from donor X |
| Provisional cell type       | Quantitatively-derived data cluster defined within a taxonomy. Sometimes called "cell type" here for simplicity. | N/A                                                                     |
| Dendrogram                  | A hierarchical organization of cell types defined for a specific taxonomy. Dendrograms have a specific semantic and visualizable structure. | N/A                                                                     |
| Community structure         | Non-hierarchical relationships between cell types defined as groups of cell types in a graph. | N/A                                                                     |
| Cell set accession ID       | A unique ID across all tracked datasets and taxonomies. This tag labels the taxonomy and numbers each cell type. [taxonomy id][unique # within taxonomy] | -CS1910121201                                                          |
| Cell set alias              | Any cell set descriptor (e.g., what cell types are called in a paper). It can be defined computationally based on the data, or manually based on prior knowledge or new experiments, or a combination of both. | -Inh L1-2 PAX6 CDH12 -ADARB2 (CGE) -Chandelier -[blank] |
| Cell set label              | An ID unique within a single taxonomy that contains a biological term that ideally is included in a relevant cell ontology. In some cases, it may be redundant with accession ID. [Canonical cell type][unique # within canonical cell type] | -Neuron 12                                                             |
Cell set preferred alias | Analogous to “gene symbol”. At most one biologically-driven name for linking matching cell sets across taxonomies and with a reference taxonomy. | -L2/3 IT 4
-Pvalb 3
-Microglia 2 |
---|---|---|
Reference taxonomy | A taxonomy based on one or more high-confidence datasets, to be used as a baseline of comparison for datasets collected from the same organ system. | -Cross-species cortical cell type classification |
Morpho-electric (ME) type | A provisional cell type defined using a combination of morphological and electrophysiological features | -ME_Exc_7 |
Governing body | A group of people who formulate the policy and direct the affairs of this nomenclature schema and associated ontologies and databasing efforts on a voluntary or part-time basis. | N/A |

### Application of nomenclature schema

Any nomenclature schema is only useful if it is adopted by relevant researchers - in this case, those studying cell types across multiple data sources and organ systems. Therefore, it is necessary for this schema to be easy to understand and applicable to novel and published datasets, and both utilitarian tools and examples for use of this nomenclature should be available. Executable code for applying this nomenclature is provided in the open “nomenclature” repository of the “AllenInstitute” GitHub repository (https://github.com/AllenInstitute/nomenclature). This repository also includes example input files from a published study on cell types in human MTG (Hodge et al., 2019), which is used to generate the first example shown below. Three inputs are required to run this schema: (1) a cell type taxonomy (not necessarily hierarchical), (2) a cell metadata file with cluster assignments (and optionally additional information), and (3) optional manual annotations of cell sets (e.g., aliases), which typically would be completed during taxonomy generation.

Specific details for how to apply nomenclature are included in the GitHub repository, but the general steps are as follows. First, a unique taxonomy_id is chosen, which will be used as a prefix for all the cell set accession IDs. To ensure uniqueness across all taxonomies, taxonomy_ids are tracked in a public-facing database, with future plans to transfer these to a more permanent solution that will also provide storage for accompanying taxonomy files. Second, a dendrogram is read in and used as the starting point for defining cell sets by including both provisional cell types (terminal leaf nodes) and groups of cell types with similar expression patterns (internal nodes). Third, the main script assigns accession ids and labels for each cell set and outputs an intermediate table. Fourth, the user manually annotates these cell sets to include common usage terms (preferred aliases) and can also manually add additional cell sets which correspond to any combination of cell types in the taxonomy. If the relationships between cell types are non-hierarchical, relevant sets of cells from this community structure can be input at this step. Fifth, dendrograms are optionally updated to include the new nomenclature information from this nomenclature table. Sixth, cells are assigned nomenclature tags corresponding to their cell set assignments (if any). Finally, the code produces a set of standardized files for visualization of updated taxonomic structure and for input into a future database for cross-taxonomy comparison (to be described in detail below) or inclusion in manuscripts.

### Example applications to published datasets

Three examples are presented to illustrate how this nomenclature schema can be applied to published datasets using single nucleus RNA-seq, single cell RNA-seq, and electrophysiology
Second, in a study of mouse primary visual cortex (VISp) using ~1700 cells (Tasic et al., 2016), a total of 42 neuronal and 7 non-neuronal cell types were proposed, and cell type-specific mRNA processing and genetic access to these transcriptomic types was accomplished through the use of many transgenic mouse lines (Madisen et al., 2015, 2010). Data and metadata associated with this taxonomy can be browsed with an interactive navigation application (http://casestudies.brain-
In **Figure 2**, hierarchical organization of these provisional cell types and their cell set aliases (e.g., Oligo Opalin) is used, and the nomenclature schema is applied to cell set labels (e.g., Non-neuron 3) and cell set accession IDs (e.g., CS1601040043) for each cell type (right three columns). Seven of the internal nodes are annotated with a cell set alias, label, and accession ID, as described (left column). Once again, several cell set aliases are left blank (e.g., nodes C and F), but nodes representing a useful collection of cell types can be manually tagged with an alias (e.g., node "B" contains all of the Pvalb cell types and no others).

**Figure 2. Example: Applying nomenclature schema to mouse VISp.** Center columns: dendrogram and cell type names (now called cell set aliases) for 49 neuronal and non-neuronal cell types in mouse VISp, as reproduced from (Tasic et al., 2016). Right two columns: annotation of cell types with cell set labels and accession IDs. Left column: annotation of cell sets representing a subset of internal dendrogram nodes (labeled A-F). As in **Figure 1**, only some nodes have aliases.

This nomenclature schema is equally applicable to non-transcriptomic and non-hierarchical taxonomies of cell types. Another study of mouse visual cortex examined ~1800 cells characterized electrophysiologically by whole-cell patch clamp recordings, and for a subset of...
these (450 cells) also performed morphological reconstructions (Gouwens et al., 2019). Using a multimodal unsupervised clustering method, the authors identified 20 excitatory and 26 inhibitory morpho-electric types (me-types), which are cell types defined using a combination of morphological and electrophysiological features. As a third example, Figure 3 shows the application of the nomenclature convention to the me-types of that study. The “cell set alias” and “inferred subclass” columns show the organization scheme from the original paper; in the study, me-types were organized by subclass (inferred from transgenic labels), but not placed into a binary hierarchical taxonomic tree. Cell set labels and cell set accession IDs have been assigned to each me-type (note that only examples of excitatory me-types are shown in the figure).

Figure 3. Applying nomenclature schema to mouse visual cortex me-types. Excitatory morpho-electric types (me-types) from Gouwens et al. (2019) that have been incorporated into the nomenclature schema. Eleven of the original twenty excitatory me-types are shown as examples. Representative morphologies and electrophysiological responses are shown to illustrate the differences between types. Abbreviations: IT: intratelencephalic; CF: corticofugal; NP: near-projecting; CT: corticothalamic; RS: regular-spiking. L2/3–L6b refer to cortical layers.
Definition of reference taxonomies and associated nomenclature

There is compelling evidence for the existence of distinct cell types based on robust groupings of cells by observable and measurable cell attributes; in particular, gene expression. Due to advances in low-cost, high-throughput, high-resolution single cell transcriptomics, many researchers favor building a gene expression-based reference taxonomy (or reference cell type classification system), and then layering on additional phenotypic data as they become available, especially from multimodal (e.g., Patch-Seq type) assays (Yuste et al., 2019). Correspondences between modalities in the brain are likely to be more complicated than in the retina (Zeng and Sanes, 2017). For example, long-range neuronal projections are only partially predicted by gene expression profiles in adult mouse cortex (Chen et al., 2019) and ventromedial hypothalamus (Kim et al., 2019). However, molecular, physiological, and morphological characteristics of cortical neurons are highly correlated based on simultaneous measurement in individual cells using Patch-seq (Berg et al., 2020; Gouwens et al., 2020; Scala et al., 2020). Furthermore, many groups currently perform scRNA-Seq analysis in different areas of the brain, from all organs in the human body (Rozenblatt-Rosen et al., 2017), from multiple mammalian species (Geirsdottir et al., 2019), and across trajectories of development (Nowakowski et al., 2017), aging (The Tabula Muris consortium et al., 2019), and disease (Mathys et al., 2019). The definition of a reference (or consensus) cell type taxonomy is likely to evolve as new data become available. Therefore, tracking cell type definitions using appropriate ontology, data structure, and nomenclature is critical. With this in mind, it will be important to associate specific experiment and analysis parameters (experimental design, sample source, scRNA-Seq platform used, data processing methods and clustering or other algorithms used, etc.) and clustering results with the latest reference cell types to build a comprehensive knowledge base. More generally, a multi-staged analysis workflow for defining a reference taxonomy and mapping cell types from other taxonomies is envisioned (Figure 4). This workflow accommodates methodological differences in cell type definitions that will likely vary across different studies and change as new methods are developed. This workflow makes some major assumptions about cell type ontology, data visualization, and governance that will be discussed below.

Figure 4. Workflow for assigning types to a given dataset with taxonomy. (1) Cell type classification will initially be performed separately on all taxonomies. (2) One, some, or all these datasets will be combined into a high-confidence reference taxonomy which can be used as a comparator for any related datasets, by (3) mapping existing and new datasets to the reference taxonomy. (4) The reference will periodically be updated as new datasets and taxonomies are generated.
This workflow can be broken down into four broad stages. First, many research teams will independently define cell types, identify their discriminating features, and name them using one of many available experimental and computational strategies. This represents the current state of the field. Hence, this nomenclature schema should be applied to each dataset independently. Second, an initial reference cell type classification will be defined by taking the results from one or more (ideally validated) datasets and integrating these data together in a single analysis, if needed. Being high-dimensional, high-throughput, and relatively low cost, transcriptomics strategies are immediately applicable to many organs and species, and the goal is for reference cell types to be defined using this modality (Yuste et al., 2019). Once reference cell types are defined, features discriminating cell types should be calculated separately in each dataset so that canonical discriminating features can be separated from ones represented in only a subset of datasets. These features, along with additional meta-data available for these cell types from any of the integrated datasets will be used to match cell types to existing ontologies where possible, and to update ontology terms as needed. Reference cell types will then be named as described in this nomenclature convention, with some additional constraints discussed below. For the third step in this workflow, this reference cell type classification can now be used as a comparator for any related datasets, providing a mechanism for transferring prior knowledge about cell types across datasets. Existing data can be renamed by mapping cell sets onto the reference classification and then updating the cell set preferred alias to match terms defined in the reference. For new datasets, taxonomies can be generated using any clustering or alignment strategy followed by the same mapping and annotation transfer steps. Finally, new versions of the reference cell type classification will be periodically generated using additional data and/or computational methods, and this new classification will now be used as comparator for related datasets. Steps 3 and 4 can iterate at some to-be-defined cadence. As a whole, this workflow provides a general outline for versioned cell type classification that could be specialized as needed for communities studying different organ systems and that provides a starting point for design of future cell taxonomy and nomenclature databases.

In a reference taxonomy, the preferred alias is designed to allow tracking of the same cell types across multiple taxonomies defined using data from multiple modalities. For this reason, it should match (directly map to) cell types defined in the relevant ontology (i.e., Cell Ontology (Diehl et al., 2016) or Neuron Phenotype Ontology (Gillespie et al., 2019)), at the highest level of resolution possible. For the brain, both of these ontologies define cell types as glutamatergic or GABAergic neurons in a particular brain region and therefore many new preferred aliases will need to be defined. Ideally this tag, or other cell set aliases, should include prior knowledge (provenance), canonical discriminating genes, and/or information from other modalities (such as electrophysiological properties, if available) to provide the best data to match cell types in the reference with cell types defined previously, or in future taxonomies from any modality. An attempt has been made to extend the cell ontology to a higher resolution in human MTG with a provisional cell ontology (http://bioportal.bioontology.org/ontologies/PCL) (Aevermann et al., 2018; Hodge et al., 2019). Additional steps would be required to combine this effort with the effort of identifying biologically meaningful cell type names to link with these ontologies, which is the focus of the next section.

As an example, in Figure 5 presents a reference taxonomy that combines the snRNA-seq data from human MTG (Hodge et al., 2019) (Fig. 1) with a scRNA-seq data from mouse VISp (Tasic et al., 2018) (a more recent taxonomy along the lines of Fig. 2), and show how this
reference taxonomy can be matched to data from other modalities through informed selection of preferred alias. This analysis includes many of the same cells analyzed in Figures 1 and 2 in a new joint analysis. The two studies underlying the reference taxonomy both identified ~75-100 types per cortical area per species from ~15,000 cells, which each could be defined using one (or a combination of) robust marker genes. In addition, cell types are characterized by associated relevant meta-data from the assigned cells, including cortical layer of dissection, brain region, alignment statistics, and (in mouse) projection targets of a subset of cells. Data integration was applied to these two datasets using scAlign (Johansen and Quon, 2019), yielding approximately forty reference cell types (Fig 5a-b). The resulting dataset included some one-to-one matches (e.g., a single mouse cell type matches a single human dataset), with many of the remaining reference types matching to internal nodes of the tree (e.g., cell sets that are not defined as putative cell types). From these data, preferred aliases were assigned using a combination of (i) robust gene markers from the literature, (ii) highly discriminating gene markers in these data, (iii) projection targets in mouse, (iv) historical names based on cell shape, and (v) broad cell class names (that directly map to ontologies), providing a starting point for how brain cell types could be named.

In this study Pvalb-expressing interneurons could be divided into three cell types (Fig 5c), one of which, through use of mouse transgenic lines, could be associated with Chandelier cells. In more recent work using Patch-seq (Gouwens et al., 2020), cell type assignations are confirmed by collecting multimodal data from the same cells and linking cell sets defined from morphology and electrophysiology with transcriptomic types presented in VISp (Tasic et al., 2018) (Fig 5d). The proposed schema can readily represent the matching of Chandelier cells (and other cell types) between the single-species transcriptomics datasets with the integrated transcriptomics reference and can likewise link with taxonomies from other modalities using preferred alias tags.

Figure 5: Linkage of morphology types to a human and mouse cortex cell type reference taxonomy. A-C) Reference classification of cells from mouse VISp and human MTG. A) t-SNE visualization of human (n = 3,594 nuclei) and mouse (n = 6,595 cells) inhibitory neuron clusters after alignment with scAlign, color-coded by species (top) or reference classification (bottom). B) Visualization of reference classification, with the number of clusters in each taxonomy (left), the preferred aliases for each reference cell type (middle), and the dendrogram (right). C) Human and mouse cell type homologies for Pvalb neurons from mouse V1 predicted based on shared cluster membership. Grey shade corresponds to the minimum proportion of human nuclei or mouse cells that co-cluster. Rows are human clusters and columns are mouse clusters. D) Characterization of electrophysiological, transcriptomic, and morphological features of cell types defined from all three modalities. Lines between C and D indicate the linkages between cell types in these two taxonomies. Chandelier cells are found in the Pvalb Vipr2 cell type and have one-to-one correspondence with a single human transcriptomic type (*ed). Panels A, B, and C are adapted from (Hodge et al., 2019) and panel D is adapted from (Gouwens et al., 2020).
A challenging and potentially contentious question in cell type classification is how these newly identified cell types should be named, or in this convention: what should be put in the “preferred alias” identifier of provisional cell types? As described above, mammalian brain cell types inhabit a complex landscape with fuzzy boundaries and complicated correspondences between species and modalities, leading to a variety of disparate solutions for naming cell types. As a starting point for discussion, we propose a strategy for naming cell types in the mammalian cortex that includes properties which are cell-intrinsic and potentially well-conserved between species (Table 2). These preferred aliases have been used in multiple studies of the primary motor cortex (as discussed below) and could be extended with some modifications to other brain structures.

For glutamatergic neurons, cell types are named based on predominant layer(s) of localization of cell body (soma) and their predicted projection patterns. The relatively robust laminarity of glutamatergic cell types has been described based on cytoarchitecture in multiple mammalian species for many years (e.g., Rakic, 1984), and has been confirmed using RNA in situ hybridization (Hodge et al., 2019; Tasic et al., 2018; Zeng et al., 2012), and a combination of layer dissections and scRNA-seq (Hodge et al., 2019; Tasic et al., 2018). While in humans many cell types don’t follow the layer boundaries defined by cytoarchitecture entirely, laminar patterning...
is still generally well conserved between human donors and mice (Hodge et al., 2019). In adult mouse visual cortex, projection targets for cell types have been explicitly measured using a combination of retrograde labeling and scRNA-seq (Tasic et al., 2018). By aligning cell types across species, the projection targets in mice can be hypothetically extrapolated to putative projection targets in human, or other mammalian species. For example, von Economo neurons are likely to project subcortically (Hodge et al., 2020). For GABAergic interneurons, developmental origin may define cell types by their canonical marker gene profile established early in development, with Pvalb and Sst labeling the cell type derived from medial ganglionic eminence and Vip and Lamp5 labeling the cell type derived from caudal ganglionic eminence (DeFelipe et al., 2013). Non-neuronal cell types have not been a focus of the studies cited; hence they are labeled at a broad class level only. However, knowledge from other studies using single-cell transcriptomics studies on microglia (Hammond et al., 2019; Li et al., 2019), astrocytes (Batiuk et al., 2020), and oligodendrocytes (Marques et al., 2016) could be included in subsequent versions of this convention. In all cases, multiple cell types are present within a given class. While it may not be possible to directly translate every feature to other organs, most of the concepts proposed here could still be followed.

Table 2: Proposed strategy for naming cortical cell types

| Class          | Format                  | Example   |
|----------------|-------------------------|-----------|
| Glutamatergic  | [Layer][Projection] #   | L2/3 IT 4 |
| GABAergic      | [Canonical gene(s)] #   | Pvalb 3   |
| Non-neuronal   | [Cell class] #          | Microglia 2 |

Applications of the nomenclature convention

This nomenclature schema has been adopted (including a modified version of the proposed naming conventions) for a joint taxonomy of human and mouse described above (Hodge et al., 2019) and for the taxonomies in an open website (http://celltypes.brain-map.org/rnaseq/). It is now being applied to all taxonomies generated as part of the BRAIN Initiative Cell Census Network (BICCN; https://biccn.org/), which represents an integrated consortia of centers and laboratories who are working collaboratively to generate, map, and share data and information about cell types in human, mouse, and non-human primate brain. This nomenclature schema is included in a series of bioRxiv submissions aimed at defining cell types in the primary motor cortex (M1) of mouse (see Supplementary Table 3 in Yao et al., 2020a), and between mammalian species (see Supplementary Table 3 in Bakken et al., 2020). In particular, the multiple taxonomies presented as part of the cross-species comparison (Bakken et al., 2020) present a good example of its utility. This study included single cell data from three separate “omics” modalities (transcriptomics, epigenetics, and methylation), for mouse, non-human primate, and human, for a total of ten distinct datasets. Datasets were integrated in two ways: first, epigenetics and methylation datasets were integrated with snRNA-seq data within mouse, marmoset, and human independently (as shown in Figure 6A for human); and second, snRNA-seq from each species (plus macaque) were aligned into a single integrated reference. A total of eleven taxonomies were generated (Fig. 6B), and this nomenclature convention was applied to all eleven.

Figure 6: Series of multimodal, cross-species taxonomies in primary motor cortex demonstrates utility of nomenclature schema. A) Taxonomies based on transcriptomic (top), open chromatin (middle), and DNA methylation in human M1. Epigenomic clusters (in rows) aligned to RNA-seq clusters as indicated by horizontal black bars and are also assigned matching cell sets in the relevant taxonomies. Adapted from
This application highlights several strategies to compare taxonomies without needing to look at gene expression (or other quantitative features). First, each taxonomy draws upon a common set of preferred alias terms, which allows for immediate linking of common provisional cell types between species and across modalities (in cases where such information can be reliably assigned). Furthermore, since this study included such a diverse representation of data types, species, and author institutions, this preferred alias list represents a reasonable starting point for any future cell type classification studies in mammalian cortex (Table 3). Second, each taxonomy includes cell sets for every link with parent integrated taxonomies (arrows and grey box in Figure 6B) and are aliased with relevant cell sets. Table 4 shows an example for all cell sets associated with Chandelier cells, which all have “Chandelier” in the preferred alias. In human, all three modalities have a single cell type corresponding to Chandelier cells, which are all given the cell set alias of “Inh L1-6 PVALB COL15A1”. The cell sets from the methylation- and epigenetics-based taxonomies additionally have a cell set alternative alias that corresponds to the cell set label in transcriptomics taxonomy, directly linking these cell types. This transcriptomic cell type is similarly linked to the Chandelier cell type in the integrated transcriptomic taxonomy. The same is seen in marmoset, but in mouse Chandelier cells are found in two transcriptomic cell types and therefore the other taxonomies link to a cell set containing both of these cell types (Neuron 056-057). While such linkages may be redundant in this case, for cell sets that have been quantitatively aligned but that don’t have extensive prior knowledge, they may be the most convenient option. A final strategy for linking cell sets between taxonomies is through inclusion of common cells; for example, many of the same cells cluster in the Chandelier cell sets from taxonomies CS1912131 and CS2002270 which both include the same cells from human M1. Tables for building these cell linkages are output as part of our nomenclature GitHub repo, but are of limited use without a corresponding cell database.
Table 3: A set of aliases in mammalian primary motor cortex, reproduced from (Bakken et al., 2020) with minor modifications (*). These terms are also applicable to other cortical areas, representing a starting point for future cell type classification efforts and for ontology curation.

| Preferred aliases | Alternative aliases / Description / Notes |
|-------------------|------------------------------------------|
| Lamp5             |                                          |
| Sncg              |                                          |
| Vip               |                                          |
| Sst Chodl         | *A very distinct Sst type expressing the gene Chodl in mouse; the only reported long-range projecting GABAergic type |
| Sst               |                                          |
| Pvalb             | Includes basket and chandelier cells     |
| Chandelier        |                                          |
| Meis2             | A very distinct GABAergic type expressing the gene Meis2 in mouse |
| CR                | Cajal Retzius                            |
| L2/3 IT           | Intratelencephalic                      |
| L4                | *Intratelencephalic; sparsely present in M1 |
| L5 IT             | Intratelencephalic                      |
| L6 IT             | Intratelencephalic                      |
| L6 IT Car3        | Intratelencephalic, a specific cell type expressing the gene Car3 in mouse |
| L5 ET             | Extratelencephalic; also known as CF (corticofugal), PT (pyramidal tract), or SC (subcortical) |
| L5/6 NP           | Near-projecting                         |
| L6 CT             | Corticothalamic                         |
| L6b               |                                          |
| OPC               | Oligodendrocyte precursor cell          |
| Astro             | Astrocyte                               |
| Oligo             | Oligodendrocyte                         |
| Endo              | Endothelial cell                        |
| VLMC              | Vascular leptomeningeal cell            |
| SMC               | Smooth muscle cell                      |
| Peri              | Pericyte                                |
| Micro-PVM         | Microglia / Perivascular macrophage     |
| GABAergic         | Typically inhibitory                    |
| Glutamatergic     | Typically excitatory                    |
| Non-neuronal                                      |
|-------------------------------------------------|
| CGE/PoA                                         |
| Caudal ganglionic eminence / Preoptic area       |
| MGE                                             |
| Medial ganglionic eminence                       |
| IT projecting                                   |
| Non-IT projecting                               |
| Oligo-OPC                                       |
| Other NN                                        |
| Cells other than neurons, astrocytes, OPCs, or oligodendrocytes |

**Table 4: Nomenclature schema for chandelier cells in Bakken et al 2020.** Table showing relevant nomenclature schema columns and taxonomy metadata, including the cell set alternative alias which links to cell set labels from other taxonomies. Abbreviations: Sp., species; Mod., Modality; Hu., Human; Ma., Marmoset; Mo., Mouse; T, Transcriptomics; M, Methylation; E., Epigenetics. Modified from (Bakken et al., 2020).

| Cell set accession ID | Cell set label | Cell set alias | Cell set preferred alias | Cell set alternate alias | Taxonomy ID | Sp. | Mod. |
|-----------------------|----------------|----------------|--------------------------|--------------------------|-------------|-----|------|
| CS1912131072          | Neuron 072     | Inh L1-6 PVALB | Chandelier               | Neuron 24 (in CS2002270) | CS1912131   | Hu. | T    |
| CS2002272017          | Neuron 17      | Inh L1-6 PVALB | Chandelier               | Neuron 072 (in CS1912131) | CS2002272   | Hu. | M    |
| CS2002273018          | Neuron 18      | Inh L1-6 PVALB | Chandelier               | Neuron 072 (in CS1912131) | CS2002273   | Hu. | E    |
| CS1912132052          | Neuron 52      | Inh PVALB     | Chandelier               | Neuron 24 (in CS2002270) | CS1912132   | Ma. | T    |
| CS2002274016          | Neuron 16      | Inh PVALB     | Chandelier               | Neuron 52 (in CS1912132) | CS2002274   | Ma. | M    |
| CS2002275006          | Neuron 06      | Inh PVALB     | Chandelier               | Neuron 52 (in CS1912132) | CS2002275   | Ma. | E    |
| CS1912133056          | Neuron 056     | Pvalb Vipr2_1 | Chandelier 1             | [none]                   | CS1912133   | Mo. | T    |
| CS1912133057          | Neuron 057     | Pvalb Vipr2_2 | Chandelier 2             | [none]                   | CS1912133   | Mo. | T    |
| CS1912133159          | Neuron 056-057 | Chandelier    | Chandelier               | Neuron 24 (in CS2002270) | CS1912133   | Mo. | T    |
| CS2002276019          | Neuron 19      | Pvalb Vipr2   | Chandelier               | Neuron 056-057 (in CS1912133) | CS2002276 | Mo. | M    |
Databasing and ontology curation

The current proposal allows scientists to organize their cell types within a single dataset, and to link taxonomies using “preferred alias” and “alias” terms. However, extension of this mechanism to allow for programmatic linking of taxonomies would dramatically increase the utility of this convention. Two strategies for addressing this shortcoming would be ontology curation and databasing. Ontology curation would allow users to link data-derived cell sets to common usage terms derived from prior knowledge and that connect directly with the well-annotated ontology tools that are available for many broader cell types (e.g., the Cell Ontology). In addition, preferred aliases defined in reference taxonomies could represent a starting point for expansion of existing ontologies to higher-resolution cell types defined using data-driven approaches (such as the terms in Table 3 for cortical neurons). Databasing addresses the issue at the opposite end, allowing matching of cell sets across taxonomies based on inclusion of the same cells. Such a database would include a module for ingesting cells and cell sets from multiple taxonomies, as well as a centralized location for storing this information. Databases provide a number of specific advantages to a nomenclature schema, including: 1) general data organization; 2) tracking individual cells across taxonomies and associated cell sets; 3) matching cell sets across taxonomies based on inclusion of the same cells; and 4) matching cell sets using a combination of multiple strategies and probabilistic inference.

Combination of these strategies for ontology and databasing with the proposed nomenclature convention would expand its utility for centralized, community-based cell typing efforts. For example, one could imagine using such a community-based tool to quickly collate information about a single cell type (e.g., chandelier cells). In this tool, a search for “chandelier cell” would return an entry in a cell type ontology, which would show its place in a hierarchy (“is a GABAergic interneuron”) along with some curated marker genes and other distinctive features. This would also link to the corresponding cell type database to pull out any cell sets tagged as “Chandelier cells”. Such a search could provide information about gene expression, electrophysiology, and morphology characteristics of these cells in mouse (see Figure 5), as well as potential selective areas of open chromatin and DNA-methylation marks, and links to corresponding cell types and their associated feature sets in human and non-human primate (see Figure 6 and Table 4). Finally, such a database could identify tagged cells from modalities entirely lacking a gene component. For example, a recent study using large-scale volumetric electron microscopy and functional imaging characterized the synaptic and connectional properties of chandelier cells in layer 2/3 of mouse visual cortex (Schneider-Mizell et al., 2020), which could be returned alongside the above information using this tool.

As a step in this direction, a recent paper (Yao et al., 2020b) presents a static browser of more than 370 transcriptomically-defined cell types in mouse hippocampus and cortex, with many of the features mentioned above, for a single taxonomy: https://taxonomy.shinyapps.io/ctx_hip_browser/. More ambitious, the Human Cell Atlas (HCA) has proposed a tool to facilitate convergence on common cell type nomenclature through a combination of automated suggestions and other strategies, and that could potentially address
some of these use cases directly. This Cell Annotation Platform is a centralized ecosystem for defining cell types and states for the Human Cell Atlas, and beyond, and is available on GitHub (https://github.com/hms-dbmi/cap-example). Whatever strategy is chosen could directly access and improve upon the proposed nomenclature convention.

Community input and governance

For this or any other schema to be useful and adopted, a governing body that will be respected by an international scientific user community needs to assemble and take on several key tasks. This governing body would need to decide which datasets to include in any reference, find a place and standard format to store the reference, provide tools for visualization of the reference data alone and mapping of data onto the reference, and potentially select a framework for annotating cell types. This group would also be responsible for vetting a standard ontology for organizing data, along with a controlled vocabulary for assigning cell type nomenclature and would need to ensure that critical data and metadata can be stored in the database. Among others, potential entities that may provide guidance for this governing body include the BICCN (https://biccn.org/), whose charter is to provide researchers and the public with a comprehensive reference of the diverse cell types in human, mouse, and marmoset brain, and the HCA (https://www.humancellatlas.org/), whose goal is to build an atlas of all cells in the human body, and which already has many of the required mechanisms in place for governance of this cell type classification workflow (Rozenblatt-Rosen et al., 2017). Expertise in the field of information management will also be necessary, hence representation from organizations like Open Biological and Biomedical Ontology (OBO) Foundry (Smith et al., 2007), the National Center for Biomedical Ontology (https://ncbo.bioontology.org/) (Musen et al., 2012), and Uberon (http://uberon.github.io/) (Haendel et al., 2014) is desired. The Neuroinformatics Information Framework (NIF) (Gardner et al., 2008) and International Neuroinformatics Coordinating Facility (INCF) (https://www.incf.org/) are other specialized communities of neuroinformaticists to consider.

This proposal is a small step in a long and iterative process involving many constituents. With cross-disciplinary partnership and ever-increasing data, refinement of the proposed convention will occur. The Allen Brain Map Community Forum (https://community.brain-map.org/c/cell-taxonomies/) has a dedicated space for discussion related to cell taxonomy refinement, to promote open and accessible opportunity for exchanging ideas and suggesting improvements. The authors look forward to engagement here, or through other open forums that are embraced by the scientific community.

Acknowledgements

The authors would like to acknowledge general input and considerations on aspects of cell nomenclature from attendees and affiliates of the workshop, “Defining an Ontological Framework for a Brain Cell Type Taxonomy: Single-Cell -omics and Data-Driven Nomenclature,” held in Seattle WA, June 2019, including Alex Pollen, Alex Wiltschko, Alexander Diehl, Andrea Beckel-Mitchener, Angela Pisco, Anna Maria Masci, Anna-Kristin Kaufmann, Anton Arkhipov, Aviv Regev, Becky Steck, Bishen Singh, Brad Spiers, Chris Mungall, Christophe Benoist, Cole Trapnell, Dan Geschwind, David Holmes, David Osumi-Sutherland, Davide Risso, Deep Ganguli, Detlev Arendt, Ed Callaway, Eran Mukamel, Evan Mascalco, Fenna Krienen, Gerald Quon, Giorgio Ascoli, Gordon Shepherd, Guoping Feng, Hangjing Liu, Jay Shendure, Jens Hjerling-Leffler, Jessica Peterson, Joe Ecker, John Feo, John Marioni, John Ngai, Jonah Cool, Josh
Huang, Junhyong Kim, Kelly Street, Kelsey Montgomery, Kara Woo, Lindsay Cowell, Lucy Wang, Luis De La Torre Ubieta, Mark Musen, Maryann Martone, Michele Solis, Mike Hawrylycz, Ming Zhan, Nicole Vasilevsy, Olga Botvinik, Olivier Bodenreider, Owen White, Peter Hunter, Peter Kharchenko, Rafael Yuste, Rahul Satija, Richard Scheuermann, Samuel Kerrien, Sean Hill, Sean Mooney, Shoaib Mufti, Sten Linnarsson, Tim Linnarsson, Tim Jacobs, Tim Tickle, Tom Nowakowski, Uygar Sümbül, Vilas Menon, and Yong Yao. We thank the NIH and CZI for generous co-sponsorship of this workshop. We would also like to acknowledge the many members of the Allen Institute, past and present, who contribute to or support the development of data and analysis of brain cell types - and the organization of this information, particularly Christof Koch, Kimberly Smith, Zizhen Yao, Carol Thompson, Rebecca Hodge, Lucas Graybuck, Thuc Nguyen, Jim Berg, Jonathan Ting, Staci Sorensen, Nik Jorstad, Susan Sunkin, Stefan Mihalas, Lydia Ng, Shoaib Mufti, and Stephanie Mok. Research reported in this publication was supported by the Allen Institute, and the National Institute of Mental Health of the National Institutes of Health under award numbers U19MH114830 (to H.Z.) and U01MH114812 (to E.S.L.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors would like to thank the Allen Institute founder, Paul G. Allen, for his vision, encouragement and support.

References

Aevermann BD, Novotny M, Bakken T, Miller JA, Diehl AD, Osumi-Sutherland D, Lasken RS, Lein ES, Scheuermann RH. 2018. Cell type discovery using single-cell transcriptomics: implications for ontological representation. Hum Mol Genet 27:R40–R47. doi:10.1093/hmg/ddy100

Armañanzas R, Ascoli GA. 2015. Towards the automatic classification of neurons. Trends Neurosci 38:307–318. doi:10.1016/j.tins.2015.02.004

Bakken TE, Jorstad SA, Hu Q, Lake BB, Tian W, Kalmbach BE, Crow M, Hodge RD, Krienen FM, Sorensen SA, Eggermont J, Yao Z, Aevermann BD, Aldridge AI, Bartlett A, Bertagnolli D, Casper T, Castanon RG, Crichton K, Daigle TL, Dalley R, Dee N, Dembrow N, Diep D, Ding S-L, Dong W, Fang R, Fischer S, Goldman M, Goldy J, Graybuck LT, Herb BR, Hou X, Kancherla J, Kroll M, Lathia K, van Lew B, Li YE, Liu CS, Liu H, Mahurkar A, McMillen D, Miller JA, Moussa M, Nery JR, Orvis J, Owen S, Palmer CR, Pham T, Phongthongkum N, Poirion O, Reed NM, Rimorin C, Rivkin A, Romanow WJ, Sedeño-Cortés AE, Siletti K, Somasundaram S, Sulc J, Tieu M, Torkelson A, Tung H, Wang X, Xie Y, Yanny AM, Zhang R, Ament SA, Bravo HC, Chun J, Dobin A, Gillis J, Hertzano R, Hof PR, Höltt T, Horwitz GD, Dirk Keene C, Kharchenko PV, Ko AL, Lelièvedt BP, Luo C, Mukamel EA, Preissl S, Regev A, Ren B, Scheuermann RH, Smith K, Spain WJ, White OR, Koch C, Hawrylycz M, Tasic B, Macosko EZ, McCarthy SA, Ting JT, Zeng H, Zhang K, Feng G, Ecker JR, Linnarsson S, Lein ES. 2020. Evolution of cellular diversity in primary motor cortex of human, marmoset monkey, and mouse. bioRxiv. doi:10.1101/2020.03.31.016972

Barkas N, Petukhov V, Nikolaeva D, Lozinsky Y, Demharter S, Khodosevich K, Kharchenko PV. 2018. Wiring together large single-cell RNA-seq sample collections. bioRxiv. doi:10.1101/460246

Batiuk MY, Martirosyan A, Wahis J, de Vin F, Marneffe C, Kusserow C, Koeppen J, Viana JF, Oliveira JF, Voet T, Ponting CP, Belgard TG, Holt MG. 2020. Identification of region-specific astrocyte subtypes at single cell resolution. Nat Commun 11:2200. doi:10.1038/s41467-019-14198-8

Berg J, Sorensen TA, Ting JT, Miller JA, Chartrand T, Buchin A, Bakken TE, Budzillo A, Dee N, Ding S-L, Gouwens NW, Hodge RD, Kalmbach B, Lee C, Lee BR, Alfiler L, Baker K, Barkan E, Beller A, Berry K, Bertagnolli D, Bickley K, Bomhen J, Braun T, Brouner K, Casper T, Chong P, Crichton K, Dalley R, de Frates R, Desta T, Lee SD, D’Orazio F, Dotson N, Eduard T, Enstrom R, Farrell C, Feng D, Fong O, Furdan S, Galakhova AA, Gamlin C, Gary A, Glandon A, Goldy J, Gorham M, Goriounova NA, Gratiy S, Graybuck L, Gu H, Hadley K, Hansen N, Heistek TS, Henry AM, Heyer DB, Hill D, Hill C, Hupp M, Jarsky T, Kebede S, Keene L, Kim L, Kim M-H, Kroll M, Latimer C, Levi BP, Link KE, Mallyor M, Mann R, Marshall D, Maxwell M, McGraw M, McMillen D, Melief E, Mertens EJ, Mezei L, Mihut N, Mok S,
Molnar G, Mukora A, Ng L, Ngo K, Nicovich PR, Nyhus J, Olah G, Oldre A, Omstead V, Ozsche A, Park D, Peng H, Pham T, Pom CA, Potekhina L, Rajanabu R, Ransford S, Reid D, Rimorin C, Ruiz A, Sandman D, Sulc J, Sunkin SM, Szafer A, Szemenyei V, Thomsen ER, Tieu M, Torkelson A, Trinh J, Tung H, Wakeman W, Ward K, Wilbers R, Williams G, Yao Z, Yoon J-G, Anastassiou C, Arkhipov A, Barzo P, Bernard A, Cobbs C, de Witt Hamer PC, Ellenbogen RG, Esposito L, Ferreira M, Gwinn RP, Hawrylycz MJ, Hpr PR, Idema S, Jones AR, Dirk Keene C, Ko AL, Murphy GJ, Ng L, Ojemann JG, Patel AP, Phillips JW, Silbergeld DL, Smith K, Tasic B, Yuste R, Segev I, de Kock CPJ, Mansvelder HD, Tamas G, Zeng H, Koch C, Lein ES. 2020. Human cortical expansion involves diversification and specialization of supragranular intratelencephalic-projecting neurons. bioRxiv. doi:10.1101/2020.03.31.018820

Betz W. 1874. Anatomischer Nachweis zweier GehirncentraCentralblatt Für Die Medizinischen Wissenschaften. pp. 12:578–580, 595–599.

Boldog E, Bakken TE, Hodge RD, Novotny M, Aevermann BD, Baka J, Bordé S, Close JL, Diez-Fuertes F, Ding S-L, Faragó N, Kocsis ÁK, Kovács B, Maltzer Z, McCorrisson JM, Miller JA, Molnár G, Oláh G, Oszvár A, Rózsa M, Shehata SI, Smith KA, Sunkin SM, Tran DN, Veneprally P, Wall A, Puskás LG, Barzo P, Steemers FJ, Schork NJ, Schueermann RH, Lasken RS, Lein ES, Tamás G. 2018. Transcriptomic and morphophysiological evidence for a specialized human cortical GABAergic cell type. Nat Neurosci 21:1185–1195. doi:10.1038/s41593-018-0205-2

Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. 2018. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol 36:411–420. doi:10.1038/nbt.4096

Cadwell CR, Palassanza A, Jiang X, Berens P, Deng Q, Yilmaz M, Reimer J, Shen S, Bethge M, Tolias KF, Sandberg R, Tolias AS. 2016. Electrophysiological, transcriptomic and morphologic profiling of single neurons using Patch-seq. Nat Neurosci 19:203. doi:10.1038/nn.3445

Cajal R. 1899. La textura del sistema nervioso del hombre y los vertebrados. 1st Moya.

Cembrowski MS, Menon V. 2018. Continuous Variation within Cell Types of the Nervous System. Trends Neurosci 41:337–348. doi:10.1016/j.tins.2018.02.010

Chen X, Sun Y-C, Zhan H, Kebschull JM, Fischer S, Matho K, Huang ZJ, Gillis J, Zador AM. 2019. High-Throughput Mapping of Long-Range Neuronal Projection Using In Situ Sequencing. Cell 179:772–786.e19. doi:10.1016/j.cell.2019.09.023

DeFelipe J, López-Cruz PL, Benavides-Piccione R, Bielza C, Larrañaga P, Anderson S, Burkhalter A, Caulli B, Fairén A, Feldmeyer D, Fishell G, Fitzpatrick D, Freund TF, González-Burgos G, Hestrin S, Hill S, Hof PR, Huang J, Jones EG, Kawaguchi Y, Kisvárdai Z, Kubota Y, Lewis DA, Marín O, Markram H, McBain CJ, Meyer HS, Monyer H, Nelson SB, Rockland K, Rossier J, Rubenstein JLR, Rudy B, Scanziani M, Shepherd GM, Sherwood CC, Staiger JF, Tamás G, Thomson A, Wang Y, Yuste R, Ascoli GA. 2013. New insights into the classification and nomenclature of cortical interneurons. Nat Rev Neurosci 14:202–216. doi:10.1038/nrn3444

Diehl AD, Meehan TF, Bradford YM, Brush MH, Dahul WM, Dougall DS, He Y, Osumi-Sutherland D, Ruttenberg A, Sabatier AF, Van Slyke CE, Vasisilevsky NA, Haendel MA, Blake JA, Mungall CJ. 2016. The Cell Ontology 2016: enhanced content, modularization, and ontology interoperability. J Biomed Semantics 7:44. doi:10.1186/s13326-016-0088-7

Frankish A, Diekhans M, Ferreira A-M, Johnson R, Jungreis I, Loveland J, Mudge JM, Sisu C, Wright J, Armstrong J, Barnes I, Berry A, Bignell A, Carbonell Sala S, Chrust J, Cunningham F, Di Domenico T, Donaldson S, Fiddes IT, García Girón C, González JM, Grego T, Hardy M, Hourlier T, Hunt T, Izuogu C, Kaneko Y, Kihara Y, Kikuno T, Knapstein EM, Korf IR, Leung KA, Mackay TF, Mair M, McCashland T, Meuth M, Middendorf M, Minden MD, Monzón-Torres CI, Mostafavi S, Müschen M, Nakamura S, Natarajan P, Notredame C, Ostermann K, Ozsolak F, Parra M, Pascual J, Patro R, Patwa D, Pei B, Perez-Perez GI, Piehler KM, Poustka A, Querfurth H, Rane J, Ratan P, Ratan T, Reimann M, Ren Y, Richardson A, Rinaldi M, Rives AD, Roy S, Schuler G, Schuw K, Schmidt D, Schmidt E, Schuler-González J, Schulerick M, Schulte C, Schuster B, Seo H, Simister N, Sinha S, Sivaramakrishnan R, Sjöblom T, Smit A, Smit B, Sonstegard TS, Spikes AB, Steiner R, Steinmetz LM, Stolnicov A, Stojanovic M, Stojanovic S, Thayer S, Tolarova M, Tran C, Trifkilis P, Turner S, van der Meer M, Wang X, Weng Z, Wu Y, Yaffe D, Yu W, Yu Y, Zdobnov EM, Zhang J, Zhang Y, Zheng D, Zerbino D, Zhang X, Zhang Y, Zylber AJ, Zylber A. 2019. GENCODE reference annotation for the human and mouse genomes. Nature 577:560–565. doi:10.1038/s41587-019-0477-4

Fuzik J, Zeisel A, Máté Z, Calvigioni D, Yanagawa Y, Szabó G, Linnarsson S, Harkany T. 2016. Integration of electrophysiological recordings with single-cell RNA-seq data identifies neuronal subtypes. Nat Biotechnol 34:175–183. doi:10.1038/nbt.3443

Gala R, Gouwens N, Yao Z, Budzillo A, Penn O, Tasic B, Murphy G, Zeng H, Sümbül U. 2019. A coupled autoencoder approach for multi-modal analysis of cell types In: Wallach H, Larochelle H, Beygelzimer A, dtextquotesingle Alché-Buc F, Fox E, Garnett R, editors. Advances in Neural Information Processing Systems 32. Curran Associates, Inc. pp. 9267–9276.
Gardner D, Aki H, Ascoli GA, Bowden DM, Bug W, Donohue DE, Goldberg DH, Grafstein B, Grethe JS, Gupta A, Halavi M, Kennedy DN, Marenco L, Martone ME, Miller PL, Müller H-M, Robert A, Shepherd GM, Sternberg PW, Van Essen DC, Williams RW. 2008. The neuroscience information framework: a data and knowledge environment for neuroscience. *Neuroinformatics* 6:149–160. doi:10.1007/s12021-008-9024-z

Geirsdottir L, David E, Keréns-Shaul H, Weiner A, Bohlen SC, Neuber J, Balic A, Giladi A, Sheban F, Dutertre C-A, Pfeifle C, Peri F, Raffo-Romero A, Vizioli J, Matiassek K, Scheiwe C, Meckel S, Mätz-Rensing K, van der Meer F, Thormodsson FR, Stadelmann C, Zilkha N, Kimchi T, Ginhoux F, Ulitsky I, Enry D, Amit I, Prinz M. 2019. Cross-Species Single-Cell Analysis Reveals Divergence of the Primate Microglia Program. *Cell* 179:1609–1622.e16. doi:10.1016/j.cell.2019.11.010

Gillespie TH, Tripathy S, Sy FM, Martone ME, Hill SL. 2019. A FAIR Knowledge Model for Classifying Neuronal Types. *Neuroinformatics*. doi:10.12751/incl.ni2019.0001

Gouwens NW, Sorensen SA, Baftizadeh F, Budzillo A. 2020. Toward an integrated classification of neuronal cell types: morphoelectric and transcriptomic characterization of individual GABAergic cortical neurons. *BioRxiv*.

Gouwens NW, Sorensen SA, Berg J, Lee C, Jarsky T, Tiensuu H, Lemus R, Slaughterbeck CR, Smith SC, Sodt AJ, Sunkin SM, Swanson BE, Vawter MP, Williams D, Wohnoutka P, Zielke HR, Geschwind DH, Hof PR, Smith SM, Koch C, Grant SGN, Jones AR. 2012. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* 489:391–399. doi:10.1038/nature11405
Hodge RD, Bakken TE, Miller JA, Smith KA, Barkan ER, Graybuck LT, Close JL, Long B, Johansen N, Penn O, Yao Z, Eggermont J, Höltt V, Levi BP, Shehata SI, Avermann B, Beller A, Bertagnolli D, Brouner K, Casper T, Cobbs C, Dalley T, Dee N, Ding S-L, Ellenbogen RG, Fong O, Garren E, Goldy J, Gwinn RP, Hirschstein D, Keene CD, Keshk M, Ko AL, Laithia K, Mahfouz A, Maltzer Z, McGraw M, Nguyen TN, Nyhus J, Ojemann JG, Oldre A, Parry S, Reynolds S, Rimorin C, Shapovalova NV, Somasundaram S, Szafar A, Thomsen ER, Tieru M, Quon G, Scheuermann RH, Yuste R, Sunkin SM, Lelienveldt B, Feng D, Ng L, Bernard A, Hawrylycz M, Phillips JW, Tasic B, Zeng H, Jones AR, Koch C, Lein ES. 2019. Conserved cell types with divergent features in human versus mouse cortex. Nature 573:61–68. doi:10.1038/s41586-019-1506-7

Hodge RD, Miller JA, Novotny M, Kalmbach BE, Ting JT, Bakken TE, Avermann BD, Barkan ER, Berkowitz-Cer盆ano ML, Cobbs C, Diez-Fuentes F, Ding S-L, McCorrison J, Schork NJ, Shehata SI, Smith KA, Sunkin SM, Tran DN, Vepally P, Yanny AM, Steemers FJ, Phillips JW, Bernard A, Koch C, Lasken RS, Scheuermann RH, Lein ES. 2020. Transcriptomic evidence that von Economo neurons are regionally specialized extratelencephalic-projecting excitatory neurons. Nat Commun 11:1172. doi:10.1038/s41467-020-14952-3

Johansen N, Quon G. 2019. scAlign: a tool for alignment, integration, and rare cell identification from scRNA-seq data. Genome Biol 20:166. doi:10.1186/s13059-019-1766-4

Jonas E, Kording K. 2015. Automatic discovery of cell types and microcircuitry from neural connectomics. Elife 4:e04250. doi:10.7554/eLife.04250

Kim D-W, Yao Z, Graybuck LT, Kim TK, Nguyen TN, Smith KA, Fong O, Yi L, Koulena N, Pierson N, Shah S, Lo L, Pool A-H, Oka Y, Pachter L, Cai L, Tasic B, Zeng H, Anderson DJ. 2019. Multimodal Analysis of Cell Types in a Hypothalamic Node Controlling Social Behavior. Cell 179:713–728.e17. doi:10.1016/j.cell.2019.09.020

Kitts PA, Church DM, Thibaud-Nissen F, Choi J, Hem V, Sapoajnikov V, Smith RG, Tatusova T, Xiang C, Zherikov A, DiCuccio M, Murphy TD, Pruitt KD, Kimchi A. 2016: Assembly for resource assembled genomes at NCBI. Nucleic Acids Res 44:D73–80. doi:10.1093/nar/gkv1226

Kotliar D, Veres A, Nagy MA, Tabrizi S, Hodis E, Melton DA, Sabeti PC. 2019. Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-Seq. Elife 8:e43803. doi:10.7554/eLife.43803

Lake BB, Ai R, Kaeser GE, Salathia NS, Yung YC, Liu R, Wildberg A, Gao D, Fung H-L, Chen S, Vijayaraghavan R, Wong J, Chen A, Sheng X, Kaper F, Shen R, Ronaghi M, Fan J-B, Wang W, Chun J, Zhang K. 2016. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. Science 352:1586–1590. doi:10.1126/science.aaf1204

Larson SD, Martone ME. 2009. Ontologies for Neuroscience: What are they and What are they Good for? Front Neurosci 3:60–67. doi:10.3389/neuro.01.007.2009

Lees JR, Azimzadeh AM, Ding Y, Webb TJ, Bromberg JS. 2015. Cells of the immune system In: Li XC, Jevnikar AM, editors. Transplant Immunology. Chichester, UK: John Wiley & Sons, Ltd. pp. 25–47. doi:10.1002/9781119072997.ch2

Li Q, Cheng Z, Zhou L, Darmanis S, Neff NF, Okamoto J, Gulati G, Bennett ML, Sun LO, Clarke LE, Marschallinger J, Yu G, Quax SR, Wyss-Coray T, Barres BA. 2019. Developmental Heterogeneity of Microglia and Brain Myeloid Cells Revealed by Deep Single-Cell RNA Sequencing. Neuron 101:207–223.e10. doi:10.1016/j.neuron.2018.12.006

Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tiros h I, Bialas AR, Kamitaki N, Martersteck EM, Trombek JJ, Weitz DA, Sanes JR, Shalek AK, Regev A, McCarroll SA. 2015. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell 161:1202–1214. doi:10.1016/j.cell.2015.05.002

Madisen L, Garner AR, Shimada D, Chuong AS, Klapoetke NC, Li L, van der Bourg A, Niino Y, Egolf L, Monetti C, Gu H, Mills M, Cheng A, Tasic B, Nguyen TN, Sunkin SM, Benucci A, Nagy A, Miyawaki A, Helmchen F, Empson RM, Knöpfel T, Boyden ES, Reid RC, Carandini M, Zeng H. 2015. Transgenic mice for intersectoral targeting of neural sensors and effectors with high specificity and performance. Neuron 85:942–958. doi:10.1016/j.neuron.2015.02.022

Madisen L, Zwingman TA, Sunkin SM, Oh SW, Zariwala HA, Gu H, Ng LL, Palmiter RD, Hawrylycz MJ, Jones AR, Lein ES, Zeng H. 2010. A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. Nat Neurosci 13:133–140. doi:10.1038/nn.2467

Markram H, Toledo-Rodriguez M, Wang Y, Gupta A, Silberberg G, Wu C. 2004. Interneurons of the neocortical inhibitory system. Nat Rev Neurosci 5:793–807. doi:10.1038/nrn1519
Marques, S., Zeisel, A., Codeluppi, S., van Bruggen, D., Mendanha, Falcão, A., Xiao, L., Li, H., Häring, M., Hochgener, H., Romanov, R.A., Gyllibrand, D., Muñoz, Manchado, A., La Manno, G., Lönnberg, P., Floridiana, E.M., Rezayee, F., Ernfors, P., Arenas, E., Hjerling-Leffler, J., Harkany, T., Richardson, W.D., Linnarsson, S., Castelo-Branco, G. 2016. Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system. Science 352:1326–1329. doi:10.1126/science.aaf6463

Mathys, H., Davalila-Velderrain, J., Peng, Z., Gao, F., Mohammad, S., Young, J.Z., Menon, M., He, L., Abdurrob, F., Jiang, X., Martorell, A.J., Ransoho, R.F., Hafler, B.P., Bennett, D.A., Kellis, M., Tsai, L.-H. 2019. Single-cell transcriptomic analysis of Alzheimer’s disease. Nature 570:332–337. doi:10.1038/s41586-019-1195-2

McInnes, L., Healy, J., Melville, J. 2018. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv [statML].

Musen, M.A., Noy, N.F., Shah, N.H., Whetzel, P.L., Chute, C.G., Story, M.A., Smith, B., NCBO team. 2012. The National Center for Biomedical Ontology. J Am Med Inform Assoc. 19:190–195. doi:10.1136/amiajnl-2011-000523

Nowakowski, T.J., Bhaduri, A., Pollen, A.A., Alvarado, B., Mostajo-Radji, M.A., Di Lullo, E., Haeussler, M., Sandoval-Espinosa, C., Liu, S.J., Velmeshev, D., Ounadjela, J.R., Shuga, J., Wang, X., Lim, D.A., West, J.A., Leyrat, A.A., Kent, W.J., Kriegstein, A.R. 2017. Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. Science 358:1318–1323. doi:10.1126/science.aap8809

Packer, J.S., Zhu, Q., Huynh, C., Sivaramakrishnan, P., Preston, E., Dueck, H., Stefanik, D., Tan, K., Trapnell, C., Kim, J., Waterston, R.H., Murray, J.I. 2019. A lineage-resolved molecular atlas of C. elegans embryogenesis at single-cell resolution. Science 365. doi:10.1126/science.aax1971

Rakic, P. 1984. Organizing Principles for Development of Primate Cerebral Cortex In: Sharma SC, editor. Organizing Principles of Neural Development. Boston, MA: Springer US. pp. 21–48. doi:10.1007/978-1-4684-4802-3_2

Rozenblatt-Rosen, O., Stubbington, M.J.T., Regev, A., Teichmann, S.A. 2017. The Human Cell Atlas: from vision to reality. Nature 550:451–453. doi:10.1038/n550451a

Scalá, F., Kobak, D., Bernabucci, M., Bernaerts, Y., Cadwell, C.R., Castro, J.R., Hartmanis, L., Jiang, X., Laturnus, S., Miranda, E., Mulherkar, S., Tan, Z.H., Yao, Z., Zeng, H., Sandberg, R., Berens, P., Tolias, A.S. 2020. Phenotypic variation within and across transcriptomic cell types in mouse motor cortex. bioRxiv. doi:10.1101/2020.02.03.929158

Schneider-Mizell, C.M., Bodor, A., Collman, F., Brittain, D., Bleckert, A.A., Dorkenwald, S., Turner, N.L., Macrina, T., Lee, K., Lu, R., Wu, J., Zhuang, J., Nandi, A., Hu, B., Buchanan, J., Takeno, M.M., Torres, R., Mahalingam, G., Bumbarger, D.J., Li, Y., Charttrand, T., Kemnitz, N., Silversmith, W.M., Ih, D., Zung, J., Zlateski, A., Tartavull, I., Popovych, S., Wong, W., Castro, M., Jordan, C.S., Froudarakis, E., Becker, L., Luckow, S., Reimer, J., Tolias, A.S., Anastassiou, C., Sebastian, Seung, H., Clay, Reid, R., da Costa, N.M. 2020. Chandelier cell anatomy and function reveal a variably distributed but common signal. bioRxiv. doi:10.1101/2020.03.31.018952

Seung, H.S., Sümbül, U. 2014. Neuronal Cell Types and Connectivity: Lessons from the Retina. Neuron 83:1262–1272. doi:10.1016/j.neuron.2014.08.054

Shepherd, G.M., Marron, L., Hines, M.L., Migliore, M., McDougal, R.A., Carnevale, N.T., Newton, A.J.H., Surles-Zeigler, M., Ascoli, G.A. 2019. Neuron Names: A Gene- and Property-Based Name Format, With Special Reference to Cortical Neurons. Front Neuroanat 13:25. doi:10.3389/fnana.2019.00025

Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L.J., Eilbeck, K., Ireland, A., Mungall, C.J., OBI Consortium, Leontis, N., Rocca-Serra, P., Ruttenberg, A., Sansone, S-A., Scheuermann, R.H., Shah, N., Whetzel, P.L., Lewis, S. 2007. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol 25:1251–1255. doi:10.1038/nbt1346

Szentesi-Gothai, J., Arbib, M.A. 1974. Conceptual models of neural organization. Neurosci Res Program Bull 12:305–510.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, BB., Siddiqui, A., Lao, K., Surani, M.A. 2009. mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods 6:377–382. doi:10.1038/nmeth.1315

Tasic, B., Menon, V., Nguyen, T.N., Kim, T.K., Jarsty, T., Yao, Z., Levi, B., Gray, L.T., Sorensen, S.A., Dolbeare, T., Bertagnolli, D., Goldy, J., Shapovalova, N., Parry, S., Lee, C., Smith, K., Bernard, A., Madisen, L., Sunkin, S.M., Hawrylycz, M., Koch, C., Zeng, H. 2016. Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. Nat Neurosci 19:335–346. doi:10.1038/nn.4216

Tasic, B., Yao, Z., Graybuck, L.T., Smith, K.A., Nguyen, T.N., Bertagnolli, D., Goldy, J., Garren, E., Economou, M.N., Viswanathan, S., Penn, O., Bakken, T., Menon, V., Miller, J., Fong, O., Hirokawa, K., Lathia, K., Rimorin, C.,
Tieu M, Larsen R, Casper T, Barkan E, Kroll M, Parry S, Shapovalova NV, Hirschstein D, Pendergraft J, Sullivan HA, Kim TK, Szafer A, Dee N, Groblewski P, Wickersham I, Cetin A, Harris JA, Levi BP, Sunkin SM, Madisen L, Daigle TL, Looger L, Bernard A, Phillips J, Lein E, Hawrylycz M, Svoboda K, Jones AR, Koch C, Zeng H. 2018. Shared and distinct transcriptomic cell types across neocortical areas. *Nature* **563**:72–78. doi:10.1038/s41586-018-0654-5

The Tabula Muris consortium, Pisco AO, Schaun N, McGeever A, Karkaniajs J, Neff NF, Darmanis S, Wyss-Coray T, Quake SR. 2019. A Single Cell Transcriptomic Atlas Characterizes Aging Tissues in the Mouse. *bioRxiv*. doi:10.1101/661728

Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, González-Galindo A, Gray AJG, Groth P, Goble C, Grethe JS, Heringa J, t’Hoen PAC, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone S-A, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagepetersen S, Wolstencroft K, Zhao J, Mons B. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**:160018. doi:10.1038/sdata.2016.18

Wu YE, Pan L, Zuo Y, Li X, Hong W. 2017. Detecting Activated Cell Populations Using Single-Cell RNA-Seq. *Neuron* **96**:313–329.e6. doi:10.1016/j.neuron.2017.09.026

Yao Z, Liu H, Xie F, Fischer S, Sina Booeeshaghi A, Adkins RS, Aldridge AL, Ament SA, Pinto-Duarte A, Bartlett A, Margarita Behrens M, Van den BERGE K, Bertagnolli D, Bialcalski T, Bravo HC, Casper T, Colantuoni C, Creasy H, Crichton K, Crow M, Dee N, Dougerty EL, Doyle WI, Dudoit S, Fang R, Félix V, Fong O, Giglio M, Goldy J, Hawrylycz M, de Bézieux HR, Herb BR, Hertzano R, Hou X, Hu Q, Crabtree J, Kancherla J, Kroll M, Lathia K, Li YE, Lucero JD, Luo C, Mahurkar A, McMillen D, Nadaf N, Nery JR, Niu S-Y, Orvis J, Osteen JK, Pham T, Poisson R, Preissl S, Purdom E, Rimorin C, Risso D, Rivkin AC, Smith K, Street K, Suls J, Nguyen TN, Tieu M, Torkelson A, Tung H, Vaishnav ED, Svensson V, Vanderburg CR, Ntranos V, van Velthoven C, Wang X, White OR, Josh Huang Z, Kharchenko PV, Pachter L, Ngai J, Regev A, Tasic B, Velch JD, Gillis J, Macosko EZ, Ren B, Ecker JR, Zeng H, Mukamel EA. 2020a. An integrated transcriptomic and epigenomic atlas of mouse primary motor cortex cell types. *bioRxiv*. doi:10.1101/2020.02.29.970558

Yao Z, Nguyen TN, van Velthoven CTJ, Goldy J, Sedeno-Cortes AE, Baitzideh F, Bertagnolli D, Casper T, Crichton K, Ding S-L, Fong O, Garren E, Glandon A, Gray J, Graybuck LT, Hirschstein D, Kroll M, Lathia K, Levi B, McMillen D, Mok S, Pham T, Ren Q, Rimorin C, Shapovalova N, Suls J, Sunkin SM, Tieu M, Torkelson A, Tung H, Ward K, Dee N, Smith KA, Tasic B, Zeng H. 2020b. A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *bioRxiv*. doi:10.1101/2020.03.30.015214

Yuste R, Hawrylycz M, Aalling N, Arendt D, Armananzas R, Ascoli G, Bokharaie C, Bergmann T, Bystron I, Capogna M, Chang Y, Clemens A, de Kock C, DeFelipe J, Dos Santos S, Dunville K, Feldmeyer D, Fiath R, Fishell G, Foogetti A, Gao X, Ghaderi P, Guntrukon O, Hall VJ, Helmsstreiter M, Herculeano-Houzel S, Hilscher M, Hirase H, Hjerling-Leffler J, Hodge R, Josh Huang Z, Huda R, Juan Y, Khodosevich K, Kiehn O, Koch H, Kuebler E, Kuhnenmund M, Larrañaga P, Lelieveldt B, Louth EL, Lui J, Mansvelder H, Marin O, Martinez-Trujillo J, Moradi H, Giornouinova N, Mohapatra A, Nedergaard M, Némec P, Ofer N, Pfisterer U, Pontes S, Redmond W, Rossier J, Sanes J, Scheuermann R, Saiz ES, Somogyi P, Tamás G, Tolias A, Tosches M, Garcia MT, Aguilar-Valles A, Munguba H, Wozny C, Wuttke T, Yong L, Zeng H, Lein ES. 2019. A community-based transcriptomics classification and nomenclature of neocortical cell types. *arXiv [q-bio.GN]*.

Zappia L, Phipson B, Oshlack A. 2018. Exploring the single-cell RNA-seq analysis landscape with the scRNA-tools database. *PLoS Comput Biol* **14**:e1006245. doi:10.1371/journal.pcbi.1006245

Zeilis A, Muñoz-Manchado AB, Codeluppi S, Lönnherberg P, Manno GL, Juréus A, Marques S, Munguba H, He L, Betsholtz C, Rolny C, Castelo-Branco G, Hjerling-Leffler J, Linnarsson S. 2015. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**:1138–1142.
Zeng H, Sanes JR. 2017. Neuronal cell-type classification: challenges, opportunities and the path forward. Nat Rev Neurosci 18:530–546. doi:10.1038/nrn.2017.85

Zeng H, Shen EH, Hohmann JG, Oh SW, Bernard A, Royall JJ, Glattfelder KJ, Sunkin SM, Morris JA, Guillozet-Bongaarts AL, Smith KA, Ebbert AJ, Swanson B, Kuan L, Page DT, Overly CC, Lein ES, Hawrylycz MJ, Hof PR, Hyde TM, Kleinman JE, Jones AR. 2012. Large-scale cellular-resolution gene profiling in human neocortex reveals species-specific molecular signatures. Cell 149:483–496. doi:10.1016/j.cell.2012.02.052

Zheng Z, Lauritzen JS, Perlman E, Robinson CG, Nichols M, Milkie D, Torrens O, Price J, Fisher CB, Sharifi N, Calle-Schuler SA, Kmecova L, Ali IU, Karsh B, Trautman ET, Bogovic JA, Hanslovsky P, Jefferis GSXE, Kazhdan M, Khairy K, Saalfeld S, Fetter RD, Bock DD. 2018. A Complete Electron Microscopy Volume of the Brain of Adult Drosophila melanogaster. Cell 174:730–743.e22. doi:10.1016/j.cell.2018.06.019