nature neuroscience

Review article

https://doi.org/10.1038/s41593-024-01814-0

Implementation and validation of single-cell genomics experiments in neuroscience

Received: 7 November 2023

Accepted: 15 October 2024

Published online: 3 December 2024

Check for updates

Marco Colonna[1] ✉, Genevieve Konopka[2] ✉, Shane A. Liddelow[3,4,5,6] ✉, Tomasz Nowakowski[7,8,9,10] ✉, Rajeshwar Awatramani[11], Helen S. Bateup[12,13,14], Cathryn R. Cadwell[7,9,15,16], Emre Caglayan[2], Jerry L. Chen[17,18,19,20], Jesse Gillis[21,22], Martin Kampmann[23,24], Fenna Krienen[25], Samuel E. Marsh[26,27,28], Michelle Monje[29,30], Michael R. O'Dea[3], Rickie Patani[31,32], Alex A. Pollen[33,34], Francisco J. Quintana[35,36], Marissa Scavuzzo[37,38], Matthew Schmitz[33], Steven A. Sloan[39], Paul J. Tesar[37,46], Jessica Tollkuhn[22], Maria Antonietta Tosches[40], Madeleine E. Urbanek[41], Jonathan M. Werner[21,22], Omer A. Bayraktar[42] ✉, Ozgun Gokce[43,44] ✉ & Naomi Habib[45] ✉

Single-cell or single-nucleus transcriptomics is a powerful tool for identifying cell types and cell states. However, hypotheses derived from these assays, including gene expression information, require validation, and their functional relevance needs to be established. The choice of validation depends on numerous factors. Here, we present types of orthogonal and functional validation experiment to strengthen preliminary findings obtained using single-cell and single-nucleus transcriptomics as well as the challenges and limitations of these approaches.

Single-cell or single-nucleus RNA sequencing (sc/snRNA-seq) is a powerful tool for identifying cell types and cell states, detecting gene expression and epigenetic changes in disease and dysfunction, inferring developmental trajectories and cell-state transitions, predicting gene regulatory mechanisms, and comparing evolutionary modifications in specific tissues across species. The high-throughput nature of these experiments and the ever-growing array of computational tools for analyzing the data they produce also increases the risk of false discoveries, or discoveries that do not manifest with functional phenotypes—which confounds interpretation of these data[1,2]. To confirm preliminary findings from single-cell genomics experiments, validation experiments using orthogonal and functional methods are required (Fig. 1 and Table 1).

Here we discuss several general use cases and examples of these validation approaches[1] (Box 1 and Table 1). In Box 2, we discuss why evolutionary comparisons are important for both basic and translational neuroscientists. We also discuss challenges and limitations of sc/snRNA-seq for evolutionary comparisons (for example, disentangling homology versus convergent evolution), relating in vitro models to in vivo biology and accounting for technical and biological variability. We present specific examples of sc/snRNA-seq findings that require validation experiments, covering a variety of common scenarios and methods. Although the selected examples are not exhaustive, they illustrate how different validation steps complement and confirm sc/snRNA-seq results.

Some layer of confirmation is essential, with additional validation being subjective and depending on the nuances of each biological system. The simplest validation experiments are those which validate expression of a small number of genes. Such an approach using in situ hybridization (ISH) with probes targeting cell-type-specific genes combined with differentially expressed genes (DEGs) of interest can provide confirmation of the sc/snRNA-seq findings. Protein-level validation can be examined using immunostaining. However, these methods of validation are not suitable for validation of large numbers of DEGs, or when several genes are necessary for identification of a cellular state or multiple states (Box 1, third use case). To validate such claims, multiplex in situ methods or genome-wide spatial sequencing methods may be necessary (see below). With a growing array of spatial transcriptomics and proteomics methods, careful consideration of the pros and cons of each method is warranted.

Finally, many sc/snRNA-seq findings reach beyond description of gene expression and infer putative function and gene regulatory

A full list of affiliations appears at the end of the paper. ✉e-mail: mcolonna@wustl.edu; genevieve.konopka@utsouthwestern.edu; shane.liddelow@nyulangone.org; tomasz.nowakowski@ucsf.edu; ob5@sanger.ac.uk; ozgun.goekce@ukbonn.de; naomi.habib@mail.huji.ac.il

https://doi.org/10.1038/s41593-024-01814-0

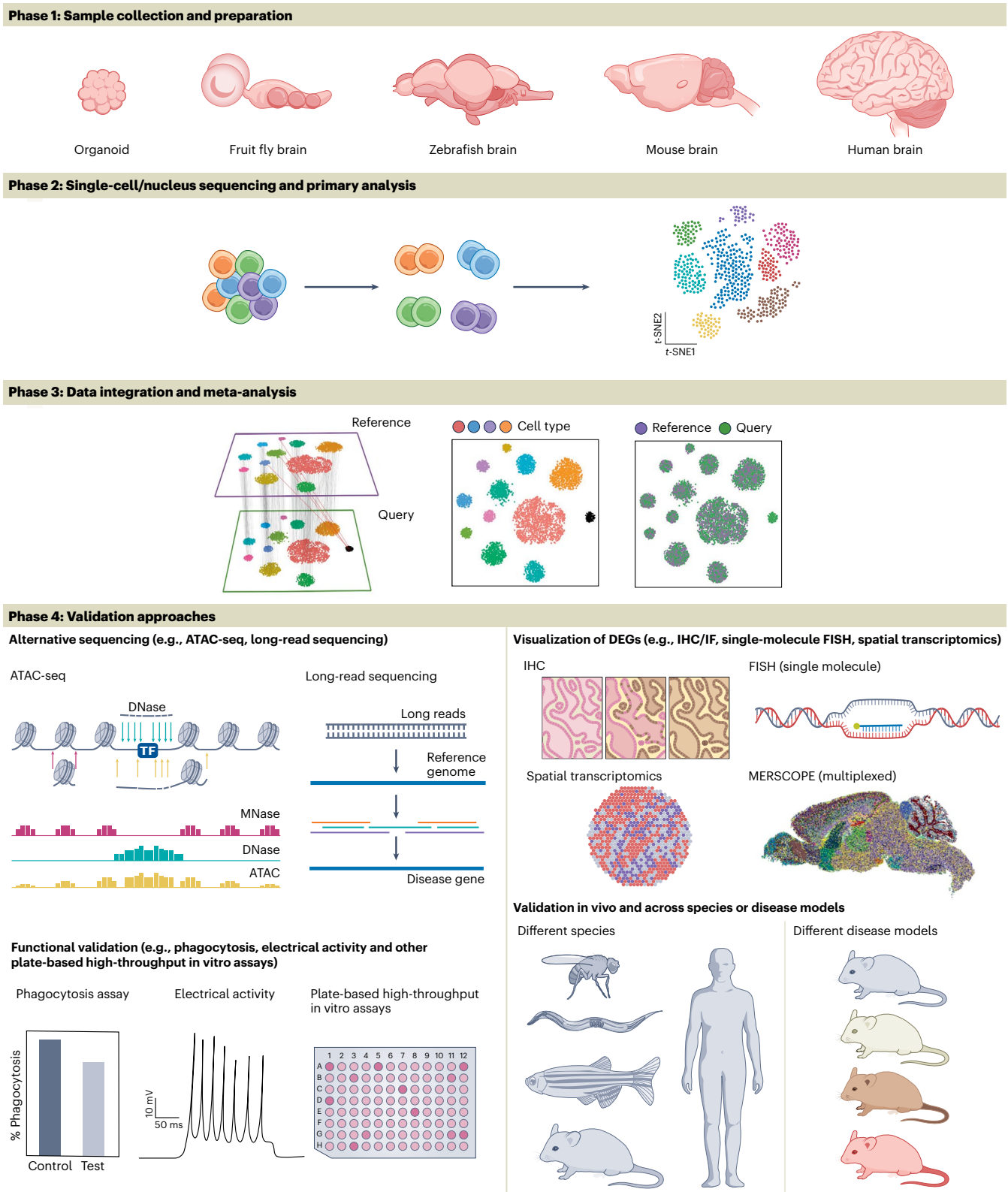

Fig. 1 | Considerations for orthogonal and functional validation of sc/snRNA-seq data. sc/snRNA-seq data, regardless of the tissue originally collected from, require multiple validation steps to ensure their biological validity (see also refs. 1,2). In addition to ensuring proper powering of cell types of interest, additional steps should be applied for best practice, including several phases: Phase 1, experimental design (for example, choice of model species) and sample preparation (for example, enzymatic versus mechanical digestion—which is important for collection of immune cells[153] or fresh versus frozen samples); Phase 2, Single-cell or single-nucleus sequencing, quality control, and primary analysis; Phase 3, integration with other datasets (across disease models, species or laboratories); Phase 4, alternative sequencing methods (for example, to access chromatin accessibility, or long-read sequencing to detect isoform abundance), visualization (using IHC, FISH or spatial transcriptomics), functional validation to ensure subtypes or substates of cells are terminal and not transitory and, finally, cross-species validation (of particular importance when using animal models of disease to ensure relevance to humans). One or all of these methods, among others, may be required to validate a number of DEGs identified in initial sc/snRNA-seq experiments.

https://doi.org/10.1038/s41593-024-01814-0

Table 1 | Overview of common validation methods

Orthogonal validation approach	Readout	What is being validated	Pros	Cons
RNAscope ISH	Molecular	RNA transcripts	Single-cell resolution spatial validation; relatively inexpensive	Low throughput
MERFISH (Vizgen)	Molecular	RNA transcripts	Single-cell resolution spatial validation	Costly; requires specialized equipment and reagents
Visium (10x)	Molecular	RNA transcripts	High-throughput and anatomical validation	Not single-cell resolution; requires specialized reagents
IHC	Molecular	Protein	Easily accessible with no specialized reagents required	Need validated antibodies; low throughput
Flow cytometry	Molecular	Protein	Quantitative readouts at protein level	Validating translation potential, which may be discordant from RNA findings; requires validated antibodies
CyTOF	Molecular	Protein	Quantification of multiple cellular components simultaneously (high throughput)	Validating translation potential, which may be discordant from RNA findings
CRISPR knockout	Functional	Gene function	Test necessity of candidate genes; relatively standardized workflows across model systems	Can be low throughput; can be costly to test multiple genes
CRISPRi/a	Functional	Gene function	Manipulate expression of endogenous genes and monitor phenotypic consequences; can be multiplexed or performed in pooled screens	Variability in degree of interference or activation from gene to gene; susceptible to epigenetic or *trans*-acting regulatory environment
Perturb-seq	Functional	Gene function	Massively parallel functional readouts of gene perturbation phenotypes by single-cell transcriptomics and individual cell resolution; can be used with traditional Cas9 or CRISPRa/i	Not trivial to design, execute and interpret; costly; may require robust selective challenge
CROP-seq	Functional	Gene function	Massively parallel functional readouts of gene perturbation phenotypes by single-cell transcriptomics and individual cell resolution; can be used with traditional Cas9 or CRISPRa/i	Not trivial to design, execute, and interpret; costly; may require robust selective challenge
ECCITE-seq	Functional	Gene function	An extension of Perturb-seq/ CROP-seq to multimodal readouts	Challenging to implement for intracellular antigens
RABID-seq	Functional connections	Cell–cell interactions	High-throughput approach to validate physical cell–cell interactions	Requires specialized reagents and bioinformatic pipelines
Circuit tracing	Functional connections	Cell–cell interactions	Can be used to identify short-range and long-range neuronal connections	May be difficult to label deep brain regions
SPEAC-seq	Functional connections	Cell–cell interactions	Allows screening for non-cell-autonomous phenotypes of gene perturbations in one cell type on another cell type in individual droplets	Requires specialized reagents and bioinformatic pipelines
Physiological readouts (calcium imaging, electrophysiology, transporter activity)	Functional	Physical properties of cells	Can match biophysical properties of cells to their transcriptional identities; powerful tools available	Requires specialized skillsets (electrophysiology); may require live intact tissue sections, cell-type-specific genetic labeling, or robust purification strategies to target cell types of interest
Live imaging (migration, proliferation)	Functional	Physical properties of cells	Can be performed in high throughput (multiple cells per image); provides input on cellular behavior	Requires specialized microscopes (two-photon, light-sheet) and live cell labeling tools
Dye filling for morphological readouts	Morphological	Cell morphology	Can provide morphological information that is far more detailed than IHC	Low throughput; requires specialized equipment
Viral targeting	Morphological	Cell morphology	High-fidelity morphological information, can provide sparse labeling for ease of reconstruction	May be difficult to label deep brain regions
Fluorescent protein expression (driver line)	Morphological	Cell morphology	Can label all cells of one type or subtype across the entire CNS	Depending on driver, labeled cell density could be too high to identify individual complex cells

CRISPR, clustered regularly interspaced short palindromic repeats; CROP-seq, CRISPR droplet sequencing; ECCITE-seq, expanded CRISPR-compatible cellular indexing of transcriptomes and epitopes by sequencing.

BOX 1

Questions that can be addressed using sc/snRNA-seq

- scRNA-seq data indicate gene X is expressed by cell type A.
- scRNA-seq data indicate gene Y is upregulated in cell type B during disease or pathology.
- scRNA-seq data indicate cell type C is composed of three substates characterized by expression of gene X, gene Y and gene Z, respectively.
- Compositional analysis indicates cell type or state D increases or decreases in abundance in disease or pathology.
- Trajectory inference or RNA velocity suggests that gene X is upregulated or downregulated as cells differentiate or respond to insult or pathology.
- Gene regulatory network inference and peak-gene linkage analysis, for example, suggest TF1 (or enhancer E1/repressor R1) modulates expression of gene A.
- Cell–cell communication analysis suggests cell type A modulates cell type B through the interactions of ligand L1 and receptor R1.

BOX 2

The importance of evolutionary studies

- Evolution is a general biological principle. Thus, understanding the contribution of evolution to nervous system function provides important foundational basic science knowledge. In addition, understanding the evolutionary constraints and opportunities that have occurred in many organisms informs our understanding of the relevance of these changes in humans.
- Understanding the similarities or differences between cell types helps us better translate our findings from one organism to another (for example, from other mammals to humans).
- Convergent evolution informs us about the constraints that shape brain evolution in terms of plasticity and functional organization of the tissue. In this manner, we can focus on the potential cellular and molecular mechanisms that correlate with convergent behaviors (for example, direct corticospinal connections onto lower motor neurons and fine motor control).
- The implementation of evolutionary approaches can result in adaption of new model systems that may offer some technical advantages for studying a general problem (for example, the evolution of sleep[154]).
- Evolved nervous system function may be directly linked to the emergence of many types of nervous system disorder in humans that are not observable in other species.
- It can be valuable to include multiple species (for example, nonhuman primates or rodents) rather than simply completing pairwise comparisons, as it is important to show that gene expression signatures track with the biology being studied and do not present irrelevant species-specific signals.

mechanisms, as in the cases of cell–cell communication analysis and gene regulatory network inference[3]. While spatial visualization of gene expression or protein abundance is useful for demonstrating colocalization of a ligand and receptor, visualization alone is insufficient to demonstrate biological function, bona fide cell–cell communication or transcription factor (TF)–enhancer–promoter interactions. Validating these mechanistic and functional inferences requires perturbation experiments (Fig. 2) and further functional studies.

Primary orthogonal validation approaches

Validation of clusters versus validation of individual DEGs

By facilitating simultaneous comparison of transcriptomic profiles from a collection of cells, sc/snRNA-seq can define clusters based on DEGs[4–6]. However, algorithms used to process and analyze these datasets are designed to detect transcriptional differences and, depending on the resolution specified, they will continue to subset cells even if the differences are so minute that they constitute noise[7]. It is essential to set several resolutions in clustering followed by iterations of computational post hoc validation before biological validation (see ref. 2 on sequencing technologies). The inherent features of clustering algorithms that provide modularity of the system also reinforce the importance of post hoc validation methods to ensure clusters reflect biology.

Integration and clustering across studies. Integration is a powerful tool for cross-dataset comparison, allowing joint assessment of cells in large-scale atlases. While this does remove cross-dataset independence, it is frequently worth that price to jointly apply cell-type calling methods. These tools can be either classical integration methods with external dataset(s) of interest (to understand whether the cell type in question has an exact correspondence to published cell types), or transcriptomic similarity approaches (to understand whether the newly identified cell type is similar to a previously described cell type). The latter can be accomplished by using AddModuleScore() in Seurat[8], Celltypist[9] or CellHint[10]. Further computational methods to validate clustering numbers include visualization of the most highly enriched transcripts per cluster (top 10 or top 100) and projection across all cells by cluster to create a feature plot of gene expression[11]. This method allows visualization of transcriptional discernment between cells in each group: the best clustering illustrates a diagonal line where each transcript has relatively uniform enrichment across all cells in a cluster, but it is lowly expressed in all other cells across other clusters. High expression of cluster-enriched genes outside the candidate cluster indicates overclustering and too high resolution—suggesting clusters likely represent the same cells or subtle changes between cell states. When sub-clustering a cell type into distinct cell states, differences between clusters are expected to be more subtle, reflecting shared and distinct expression programs between cell states. An emerging alternative approach to identify these subtle continuous differences in gene expression that drive cell states is to identify gene programs coexpressed in subsets of cells (for example, by topic modeling[12]). It may, therefore, be best to initially set the resolution high to overcluster before modifying the resolution, lowering it until robust cluster separation is visualized.

An additional computational method to ensure clustering is not an algorithmic artifact is to pulse in a known disparate cell type to see whether clusters collapse. For example, many groups subcluster cells from larger datasets to allow identification of more subtle cellular subtypes or states that are otherwise hidden in larger datasets owing to divergent cell types, or missing from smaller datasets owing to underpowering of low-abundance populations. While this method is standard and acceptable, adding in a different cell type is a stringent method to determine whether these subtypes or substates are truly different. Unlike cell types, distinguishing between cell states is more complex, as cell-state transition might require only subtle differences in expression programs and involve a continuous transition between

https://doi.org/10.1038/s41593-024-01814-0

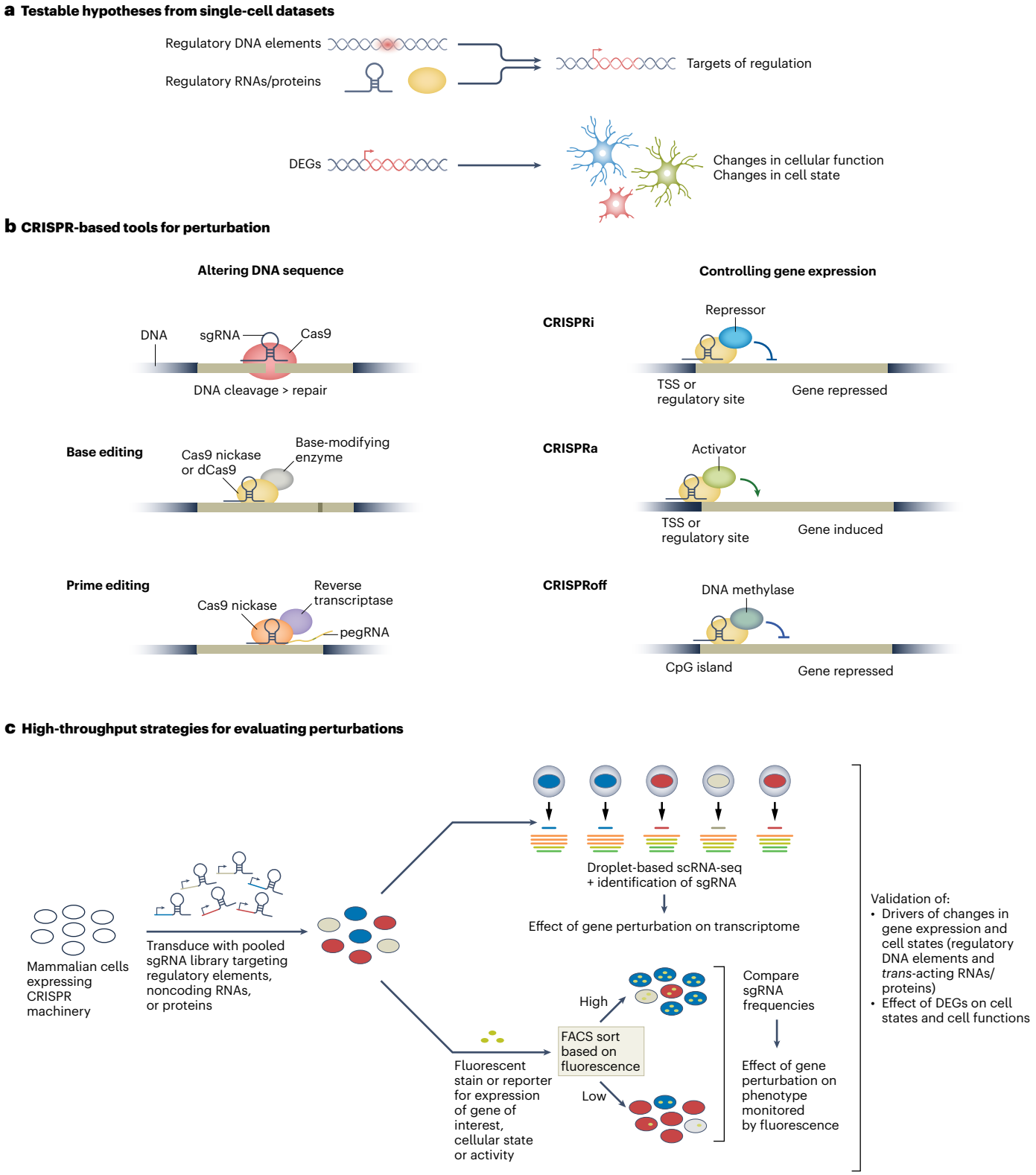

Fig. 2 | Overview of perturbation-based validation approaches. a, sc/snRNA-seq datasets can generate different types of functional or mechanistic hypothesis. Arrows mark hypotheses for causal relationships, which are generated from single-cell datasets and can be tested by CRISPR-based validation tools. **b**, Examples of CRISPR-based tools to perturb genome sequence and gene expression. **c**, Experimental strategies for high-throughput CRISPR-based perturbation experiments to validate and test functional or mechanistic hypotheses from sc/snRNA-seq datasets. TSS, transcriptional start state.

states. Moreover, the correct or most appropriate resolution of the diversity of cell states is less well defined, and different resolutions could correctly capture true biological differences in cell states. We will further discuss this issue in depth below.

Most importantly, all clustering of sc/snRNA-seq data necessitates post hoc validation[13]. Single-cell or single-nucleus data can be viewed as a prediction of biology that allows for guided hypothesis generation to gain biological insights[14].

DEGs as markers of individual cell types and clusters. DEGs can be extrapolated as rudimentary markers of cell types or an indication of changes in expression programs between cell states. However, transcriptomic profiles provide a snapshot of cellular behaviors. Given that DEG identification is inherently comparative, additional genes playing a pivotal role in driving cell identity may escape such analysis. For example, an essential interaction between two genes may be necessary for supporting a cell's function. In the absence of one of those genes, a population of cells is actually losing the combined effect of both; while this first gene may appear in a DEG analysis between these two clusters, the second gene can go unnoticed[14–16].

Herein lies the need for orthogonal validation, as the complex interplay of DEGs remains largely hidden within transcriptomic data. Orthogonal validation is a crucial intermediate in determining whether a DEG derived from a cluster is recapitulated within a cellular context relevant to the biological question at hand. Depending on parameters used for DEG discovery, as well as the condition of samples before sequencing, genes defining a cluster may not translate back to a tissue setting. The statistical method applied relies on an assumption of distribution of noise, and inferred DEGs could vary with the choice of method. For example, frequently used tests such as a generalized linear model assume a negative-binomial or Poisson distribution, while a Wilcoxon rank-sum test assumes a non-parametric distribution, and should be followed by multiple-hypothesis correction[2]. The expression level of DEGs may also fluctuate during a cell's normal state; thus, cell types or identities cannot be declared in lieu of lineage tracing and developmental analysis. It then becomes important to demonstrate that any genes of interest are truly and consistently expressed at different levels between cell states of interest before moving toward mechanistic interrogation. This orthogonal evidence includes experiments that are integral to validation of epigenomic, morphological, spatial and biophysical properties or function of cells that are predicted from transcriptomic data[13].

Prior cell classifications as a scaffold for analysis. Understandably, prior cell-type classification schemes have been biased toward molecular assays owing to the unprecedented scale and throughput of sc/snRNA-seq[17]. However, there are a number of issues that can arise when such data are interpreted in a vacuum: (1) it is difficult to distinguish molecular features that define stable cell types from transient cell states; (2) resulting cell-type atlases may vary depending on sample size and analytical parameters used for clustering, leading to lack of reproducibility with no clear ground truth; and (3) functional relevance of molecularly defined cell types is unclear.

Recent multimodal single-cell analyses call into question the notion of discrete cell types, suggesting that continuous and correlated variation in cellular morphology, biophysical properties and molecular features contributes substantially to cellular diversity within broad transcriptomic classes[18]. Validation of transcriptomic clusters, that is, confirming they have biological relevance, requires orthogonal and functional validation. One major reason for requiring such validation of in silico clusters is that cell atlas studies can be underpowered for rare cells or cell states, often causing clustering artifacts. Some methods have been developed to overcome this caveat, such as FIND-seq (focused interrogation of cells by nucleic acid detection and sequencing)[19]—developed to study rare astrocyte populations isolated on the basis of expression of a few mRNA markers. Additional validation steps, including alternative sequencing efforts, multi-dataset integration and meta-analyses and visualization (for example, in situ, MERSCOPE), are integral to validate the biological truth behind computational analyses.

Validation of transcriptomic clusters rests on the hypothesis that bona fide cell types should form discrete entities. In other words, if a group of cells segregates as a distinct population using multiple assays, this would support its designation as a valid cell type or state. The number and type of assays needed to validate a newly identified cell type or state remain unclear, but at a minimum it is recommended that findings from sc/snRNA-seq be validated using at least one independent assay (for example, visualization using in situ or spatial transcriptomics, single-nucleus assay for transposase-accessible chromatin (ATAC) to highlight chromatin accessibility for DEGs, or functional assays). However, it is also advisable to move beyond validation of individual transcripts and make every attempt to validate targets at the level of proteins, cellular physiology, anatomical distribution, developmental lineage, morphology, connectivity and/or function. These additional phenotypic validation steps establish the robustness of the cell type or state under investigation and provide mechanistic understanding of its role in the nervous system.

Methods of validation by visualization

Spatial transcriptomics. Spatial transcriptomics, in conjunction with ISH, including single-molecule ISH or RNAscope ISH, immunofluorescence (IF) and immunohistochemistry (IHC), represents a powerful combination of techniques for comprehensive characterization of gene expression and protein localization[20–22]. While sc/snRNA-seq provides information on transcriptomes at the single-cell or nucleus level, ISH enables visualization of specific RNA molecules within intact tissues, confirming their spatial distribution. IF and IHC allow detection and localization of proteins, providing additional information on cell types, protein–protein interactions and putative cellular functions. Integration of these complementary techniques verifies sc/snRNA-seq findings and allows researchers to study coexpression of genes and proteins in the context of tissue architecture, providing a more comprehensive understanding of cellular behavior and molecular interactions within complex biological systems.

By leveraging the strengths of spatial transcriptomics, ISH, IF and IHC, researchers can unravel intricate spatial dynamics of gene expression and protein localization, advancing understanding of tissue development and disease pathogenesis. These methods require a priori knowledge of DEGs from sc/snRNA-seq experiments. Whereas ISH, IF and IHC have low throughput, allowing the validation of a few genes at a time, spatial transcriptomics has a throughput of up to thousands of genes. Sequencing-based methods such as Visium and Slide-seq[23] enable simultaneous capture of gene expression information from multiple spatially defined regions within a single sample. These genome-wide technologies enable localization of groups of DEGs (often called 'gene modules') in tissue sections. By combining spatially resolved gene expression profiling at 50–100 µm resolution (updated to 8 µm resolution with Visium HD and 10 µm with Slide-tags[24]) with high-throughput sequencing, these methods provide comprehensive validation and complement sequencing data. In situ-based methods, such as multiplexed error-robust fluorescence in situ hybridization (MERFISH)[25], STARmap[26] and in situ sequencing[27], provide high-throughput direct identification of RNA transcripts at subcellular resolution of panels of several hundred genes by single-molecule fluorescence in situ hybridization (FISH) with sequential imaging and signal amplification techniques. See ref. 2 for a more in-depth discussion of these methods.

Genes versus proteins. Use of imaging-based RNA visualization to validate sc/snRNA-seq results, and/or to spatially resolve sc/snRNA-seq data, provides important context and orthogonal validation. However, when inferring potential functional consequences of gene expression changes, it is critical to also consider protein-level validation. Although transcript and protein levels are generally correlated, there are several factors that can drive dichotomy in transcript:protein ratios, including regulatory relationships and mechanisms regulating protein localization, activation and turnover[28–30]. The nonlinear relationship between gene expression and protein levels is particularly noticeable when comparing sc/snRNA-seq data with TRAP-seq or proteomics data,

where ribosomes and proteins may be in soma-distant processes—particularly common in CNS cells.

One area in which protein-level validation provides critical information is in the inference of cell–cell communication. The power of analysis at the single-cell or nucleus level has led to a rapid expansion of methods to assess putative cell–cell communication through ligand–receptor interactions from scRNA-seq and/or spatial transcriptomics data[31–33]. Results from these tools, however, should be considered hypothesis generating and not hypothesis validating. Beyond assessing whether ligands and receptors truly interact, a lack of confirmation that ligands and receptors are present at the protein level and in appropriate spatial context represents a critical gap to fill. Expression and/or spatial colocalization can be accomplished by relatively straightforward techniques such as IHC or flow cytometry/cytometry by time of flight (cyTOF), and physical interactions can be examined via co-immunoprecipitation or newer techniques such as multistage native mass spectrometry (nativeomics)[34].

Methods of validation by interrogation

Single-cell or single-nucleus transcriptomic approaches uncover DEG signatures between different cell types and states. While powerful, these data are descriptive and do not establish causality or provide mechanistic insights. However, DEGs can be leveraged to generate additional hypotheses to test using perturbation-based approaches to interrogate causality and link gene expression to cellular function (Fig. 2). We cover specific functional validation approaches in the next section. We will first describe additional gene editing approaches to validate gene expression data. It remains true, however, that even perturbation experiments must be validated at the functional level. For instance, if a perturbation experiment removes (putatively) phagocytic pathway genes, the true final validation of this is to test the functionality of phagocytosis itself.

Several perturbation methods exist in cultured cells and model organisms. RNA interference technology enables knockdown of mRNAs by synthetic short interfering RNAs or transgenically expressed short hairpin RNAs, although it suffers from off-target effects. CRISPR technology, which enables gene knockout to achieve a complete loss of function, has improved our ability to interrogate gene function in a scalable and precise manner[35]. CRISPR interference (CRISPRi) and CRISPR activation (CRISPRa) approaches use a catalytically inactive Cas9 protein to recruit transcriptional regulators to genomic sites of interest, enabling modulation of expression levels[36] and thereby providing a strategy to directly model changes in expression levels of specific DEGs identified from sc/snRNA-seq. CRISPRi/a also can target distal regulatory elements, such as enhancers, to establish their function in controlling gene expression[37,38].

CRISPR-based gene perturbation can target genes of interest in individual experiments and in massively parallel screens using pooled single guide RNA libraries targeting multiple genes of interest. Pooled screens can be conducted for a large range of phenotypes, including cell survival or specific cellular functions, and states can be read out by fluorescent markers or reporters or additional sc/snRNA-seq (for example, Perturb-seq[39–42] or CROP-seq[41]). They can also be used for screening of long noncoding RNAs[43] and *cis*-regulatory regions (for example, in induced pluripotent stem (iPS) cell-derived neurons[44] and microglia[44,45]). As reviewed in ref. 2, the development of genome editing technologies, exemplified by CRISPRi/a, CRISPR deletion and CRISPR indel screens, has facilitated large-scale perturbation and assessment of DNA sequences. CRISPR-based screens have been successfully deployed to study cell types relevant to neuroscience, including human iPS cell-derived neurons[46], microglia[47], astrocytes[48], neuron–glial co-culture systems[49], brain organoids[50], and in mouse brains in vivo[51–53]. To minimize artifacts linked to some in vitro cell isolated systems (see below), care should be taken in choice of cell culture systems, or in vivo CRISPR assays should be considered[54]. However, in vivo screens are challenging to perform, because not all cell types are currently easy to manipulate (for example, microglia), or target (for example, substates of reactive cells). In vitro screens using iPS cell-derived models and primary isolated cells are an alternative, and have been predictive of cell states found in vivo—in particular for microglia[47] and astrocytes[48,55]. These in vitro CRISPR validation steps can predict functional consequences of changes in gene expression (Fig. 2).

Functional validation

Functional validation studies are needed, as the physiological relevance of different cell clusters cannot be inferred from their transcriptional state, but they can serve as a cell typing guide. For cell types described by sc/snRNA-seq, it is important to perform functional validation to (1) ensure transcriptionally defined clusters represent a true cell type and/or subtype or substate, and (2) understand their properties in order to place them into the context of the larger circuit or brain region. This is particularly important when defining 'states' of cells. What 'functional validation' entails is still under debate. Broadly, it links the molecular profile (measured through sequencing data) with a phenotype as a way to confirm that genes or gene modules identified in sc/snRNA-seq are biologically meaningful. Such inference is not trivial, and transcriptionally distinct clusters should not be assumed to always associate with functional differences. Indeed, owing to the heuristic nature of clustering tools, some separation of cell clusters in sc/snRNA-seq data is almost inevitable and so must be interpreted with appropriate caution in terms of its biological relevance. It is noteworthy that fleeting changes in cell state, such as stage within the cell cycle, may have greater transcriptomic impact than cell type or substate[56]. The degree of functional plasticity in this context should also not be underestimated[57]. These issues notwithstanding, sc/snRNA-seq has revealed relatively consistent signatures across adult differentiated tissues, predictably correlating with changes in cellular morphology[58].

Using cell culture systems for validation

Many protocols exist for purification and culture of CNS cells. Originally involving enrichment (rather than purification) of rodent cells from embryos or neonates and inclusion of serum in media to provide trophic support, these cultures have provided understanding of basic functions of CNS cells. However, recent advances in these techniques, combined with expanded interest in studying disease states of individual cell types, requires that these methods be questioned—namely as serum is not a normal component of the CNS. While methods for the culture of CNS cells in the absence of serum have emerged[59–61], these are not always used—raising concern about in vitro functional validation experiments, in particular those of CNS glia and immune cells[62]. This does not negate the power of cell culture as a validation tool; instead it highlights the importance of choosing the appropriate in vitro model to use for replication of populations of cells of interest identified from sc/snRNA-seq studies. Such methods have been discussed extensively for astrocytes and microglia elsewhere[62], but similar points exist for all CNS cells, and will not be covered here.

Strengths of in vitro functional validation include the ability to study a homogeneous and pure population of cells of a single type or state, and the ability to validate human sc/snRNA-seq data using human cells (Fig. 2). This involves use of (induced or embryonic) pluripotent stem cell models[63,64], two-dimensional or three-dimensional culture systems[65–67] including organoids (see below), or using primary tissue explants. Astrocyte reactive states can be used as an example, with the comparison of untreated pluripotent stem cell-derived astrocytes versus an induced reactive state following treatment with tumor necrosis factor, interleukin-1α and C1q[68,69]. Clusters of interest can be isolated through live cell sorting (for example, fluorescence-activated cell sorting or magnetic-activated cell sorting) and, after revalidation of transcriptional state, consensus homeostatic functional attributes

can be interrogated (for example, glutamate reuptake, synaptogenesis capacity, neurotrophic capacity) and separately the gain of entirely new functions (for example, neurotoxicity).

Deploying these approaches means generating and/or isolating cell cultures with particular transcriptomic signatures that are viable for such functional characterization assays, which may present challenges for different cell types and/or states. Single genes or gene sets implicated by statistical prioritization (for example, by significance and effect size) or necessity and sufficiency for a particular function (or set of functions) can be genetically interrogated using these approaches. However, if a functional difference is not detected in vitro, the negative predictive value is somewhat limited because there may be specific functions that are difficult to reproduce in in vitro conditions.

Additional hypotheses and validation can be seeded by results from methods for the study of cell–cell interactions. Emerging technologies that have recently been applied to astrocytes and other glial cells include RABID-seq (rabies barcode interaction detection followed by sequencing[70]), SPEAC-seq (systematic perturbation of encapsulated associated cells followed by sequencing), which enables droplet-based culture of putative interactive cells[42] and LIPSTIC (labeling immune partnerships by SorTagging intercellular contacts[71]) for cell–cell interactions (first used to uncover interactions between T cells and dendritic cells).

Functional validation at the organismal level

Here, we aim to provide general workflows to test the functional relevance of molecularly identified cell clusters. To perform functional validation studies in animal models, it may be necessary to generate genetic reporter animals that label a given cell type based on its expression of a marker/DEG[72,73]. Genetic reporter mice can also be useful for mapping the in vivo properties of cell types in the nervous system including their activity patterns and contributions to behavior. This strategy has been valuable to catalog properties of CNS neurons[74] (with the caveat that some features may be species specific (for example, human) and thus escape validation in a different species (for example, a mouse model)). In many cases, cell types cannot be selectively labeled by expression of one marker gene alone, as other cell types may also express that gene. In that case, it can be useful to use an intersectional strategy whereby cells expressing two genes (or expressing one gene but not another) can be selectively targeted[75]. This strategy has been used to study different types of neuromodulatory neurons, including dopamine and serotonin neurons, which have been shown by sc/snRNA-seq to be highly heterogeneous[76–79].

Functional validation of neuron types has been performed in several large-scale cellular taxonomy papers[80]. A more general strategy moving forward could be to start with cell types defined by sc/snRNA-seq, then perform spatial transcriptomics to demonstrate anatomical localization, then record from cells to examine intrinsic membrane (or synaptic) properties and perform morphological reconstruction (for example, measure properties of dendrites, spines or axons), or perform Patch-seq to pair electrophysiological recordings with gene expression data from the same cell[81,82]. A further step includes generating a reporter animal that labels the population of interest, then testing in vivo circuit connectivity and behavioral relevance of that specific cell population. This strategy can also be applied for functional validation of cell types in brain organoids.

Recent work in human iPS cell-derived organoids containing both neurons and astrocytes provided functional validation of astrocytes from organoids at several developmental stages, performing a variety of assays that probe known astrocyte functions[83]. In brain organoids, cell morphology, protein levels, progenitor differentiation potential and intrinsic physiological properties can be measured functionally. Recently, in vivo imaging of neuronal activity in the mouse brain has been combined with spatial transcriptomics[84–86] to functionally validate molecularly defined cell types.

Similarly, functional validation of astrocytes in different reactive and disease states has been successful at adding biological value to transcriptomically defined populations. Starting with transcriptomic data as a roadmap for reactive astrocyte substates in vivo, researchers can isolate primary rodent or iPS cell-derived human astrocytes, recapitulate the original gene expression signature, then continue with functional validation to determine whether any loss-of-function or gain-of-function changes are present across substates. Several pipelines exist for producing high-throughput and controllable cell culture platforms for validating astrocyte disease biology[55,68,69,87,88].

Cross-species comparisons

sc/snRNA-seq provides a molecular common language definition of cell types across any species with a quality genome. While the definition of a cell type typically invokes other features (for example, connectivity, function, morphology), these features can be difficult or impossible to acquire across species[17,89]. Moreover, cell-autonomous gene expression programs are the foundation on which many (but not all) structural and functional features of a cell are built[90]. This foundation of shared gene expression programs and functional properties across species enables an inference process termed homology mapping. Properties such as connectivity and physiology are far easier to study in genetically tractable and experimentally accessible animal models (for example, *Mus musculus* or *Drosophila*), and then can be transferred by anchoring to homologous cell types in other species, including human[91].

Identifying homologous cell types across species ideally involves identifying sets of cells in each species that access similar regulatory programs for their differentiation[92]. Single-cell or single-nucleus sequencing combined with lineage tracing or fate mapping is powerful for reconstructing developmental histories of cell types[93], and hence their relationships across species. We still have a fragmented understanding of the lineages of transcriptionally defined cell types in any one species[94–97]. Few comparative studies have matched progenitor classes across species[91,98]; although, even if such information is available, shared developmental history is neither necessary nor sufficient to infer cell-type homology. The challenge with matching homologous types from adult data alone is distinguishing shared evolutionary history from phenotypic convergence[94,99], but with large enough sets of species this can be alleviated[100]. Transcription factors (TFs) potently specify cell-type identity, suggesting that prioritizing TFs in cross-species cell-type mapping may improve homology assignments. A single TF or small set of factors can be sufficient to switch the fate and ultimate cellular identity. However, TFs are developmentally regulated and may not be conserved between the early stages of specification and adulthood.

Across species, cell types change abundance, gene expression profiles and spatial distribution. Each of these differences carries its own challenges for cross-species comparisons. In general, cell-type similarity decreases, and homology mapping becomes less accurate, with increasing evolutionary distance[91,100–104]. Remarkably, sc/snRNA-seq can be used to reveal conservation and novelty of brain cell types across 500 million years of evolution, suggesting that core transcriptional programs that define cell types are often deeply conserved[99,100,105,106]. Integrating transcriptional profiles from single cells or nuclei across species can be difficult, because approaches may rely on assumptions of 1:1 orthologous genes[107], and gene duplications, losses and sequence-level divergence increase with evolutionary distance. Evolutionary modifications of cell types may result from neutral drift, physical constraints associated with brain reorganization or new functional requirements. Overall patterns of transcriptional divergence have been linked to neutral drift among primates, coupled with stabilizing selection over longer timescales at the cell type or tissue level[108].

Understanding the neuroanatomical and/or physical constraints that drive evolutionary features such as proportional shifts in cell types across species remains challenging. Increases in cell-type

abundance across species can be intuitively linked to species-specific adaptations, such as the proportional increase in a retinal ganglion cell subtype in primates that may relate to neocortically driven adaptations to high visual acuity[100]. However, sometimes the mechanism driving the differential modification of each cell type is difficult to resolve. For example, reduced proportions of subcortically projecting cortical neurons in larger mammalian brains[109] may relate to functional requirements to maintain scaling relationships between upper and lower motor neurons despite disproportionate cortical expansion, or to shifts in the migration of homologous types during development, with either mechanism leading to different anatomical distributions. Another example is the recently observed relative increase in the proportion of oligodendrocyte progenitor cells compared to mature oligodendrocytes in the human versus nonhuman primate brain[110]. This difference might support enhanced neuronal or myelin plasticity in humans, or some other phenotype that has not yet been linked to oligodendrocyte function. Increased sampling through sc/snRNA-seq across species and individuals within a species helps to distinguish processes of drift and selection, while further analysis of scaling relationships and functional changes are required to resolve contribution of physical constraints and new cellular specializations. Arbitrating between such possibilities improves understanding of the targets of evolutionary selection. Finally, issues such as overfitting to a single species reference dataset, and differences in genome quality across species, add technical complexity to cross-species comparisons.

When and how does validation fail?

Validation may fail owing to poorly designed initial sequencing experiments. This can be due to contamination of cell types of interest in bulk sequencing efforts, or, more common with the uptick in sc/snRNA-seq studies, a failure to properly power the study for the cell type of interest (see ref. 2 for a review of study design considerations). Similar artifacts arise when insufficient biological replicates are sequenced—often owing to high costs of sc/snRNA-seq experiments[111–113]. Although challenges exist, emerging technologies such as RNA editing and cell-based assays offer promise for improved validation. Careful consideration and exploration of validation options are necessary to ensure the reliability and robustness of sc/snRNA-seq findings. A brief overview of common validation difficulties from a computation standpoint are covered elsewhere[1]; here we cover other validation considerations.

Conceptual limitations of transcriptomic-based homology inference

Homology describes phylogenetic relationships and is not a synonym of similarity. Cell types may have similar transcriptomes because they descend from a common ancestor, or because they acquired these properties by convergence after evolving under similar selective pressures (Fig. 3). The ideal way to discriminate between these two possibilities is to sample many species and reconstruct ancestral states using the principle of parsimony (convergent characters tend to lack phylogenetic continuity). Because this is not always feasible for detailed sc/snRNA-seq characterization, here we highlight some observations that may help with comparing transcriptomic data across species.

Data from neuronal cell types for which homologies were established by independent criteria (morphology, input–output connectivity) offer two key insights. First, although transcriptomic divergence of homologous neurons is generally a function of phylogenetic distance[100,104,109], rates of transcriptomic divergence are cell-type specific: in primate cerebral cortex, for example, nonneuronal cells diverged more rapidly than neurons[104]. Second, transcriptomic divergence is not even across gene families. TFs known for specifying cell identity have conserved expression in homologous neuron types, whereas the expression of terminal markers or effector genes may switch more rapidly[104]. This indicates that homologous neurons may acquire species-specific functions, such as new electrophysiological properties[109], without changes to their core genetic identity.

These observations are in line with an evolutionary definition of cell type, whereby homologous cells share expression of TFs that establish and maintain their genetic identity[114]. Comparing TF expression can help disambiguate homology and convergent evolution in cell-type comparisons of distantly related vertebrate species[99,102,105,115–117]. For example, distinct classes of cortical GABAergic interneurons in amphibians, reptiles, birds and mammals express the same TFs defining class identity; however, expression of certain effector genes, such as the calcium-binding protein parvalbumin, which marks a class of mammalian GABAergic interneurons, is not conserved across species[99,102,105,115].

Are all genes equally informative?

Whether and how to give different weights to genes for homology inference remains an open question both conceptually and algorithmically. While TFs seem to carry a higher weight for homology inference, TF combinatorial codes may themselves drift (for example, by paralog switching[118]). This is particularly relevant for comparing distantly related species. It should be noted that many comparative studies are carried out in lower-quality assemblies, which may not have completed enough annotation for the 3′-biased data of sc/snRNA-seq. Standard approaches use one-to-one orthologs, assuming that genes carry similar functions, but this assumption cannot be made for paralogs, as gene duplication may be followed by sub-functionalization or neo-functionalization. However, limiting analysis to one-to-one orthologs filters out a considerable fraction of the transcriptome when the species compared are separated by large phylogenetic distances. Computational solutions to solve this problem have been proposed recently[107,119].

Finally, limiting cross-species comparisons to TFs comes with the risk of providing an oversimplified representation of cellular diversity. As the field defines subtler distinctions of subtypes within given cell classes, TF-level gradients and/or post-translational modifications may be identified in eliciting distinct transcriptional programs, making cross-species comparisons more complicated. Potential developmental regulation of TF expression makes their use to define identity challenging from a temporal perspective, even within a species. For example, the comparison of TFs may not be powered to identify cell types that have diverged recently (sister cell types[114]), which, by virtue of their recent diversification, share a large fraction of their transcriptomes. It should be noted that such post-translational modification considerations are not unique to TFs, and should be considered for all proteins.

Are cellular transcriptomes enough to infer neuronal homologies?

As described above, homology inference becomes harder with increasing phylogenetic distance, especially when there are large branch lengths between clades with no extant species, for example, comparing mammals to reptiles. Natural selection acts on the output of brain activity, that is, the ability of the brain to support adaptive behaviors in an organism's environment. The substrate of selection is the frequency of allelic variants in the population, yet resulting changes in gene transcription that do not lead to functional changes at the individual cell level would not be under selection pressure. This mapping between genotype and phenotypes under selection in the brain is nontrivial: genes do not control function directly (with a few exceptions); rather, they affect behavior by instructing cell-type identity, neuronal wiring, activity, spatial allocation and several other complex biological variables. Transcriptomes alone might be insufficient to infer homology when we do not know the traits under selection or how genes are related to those traits. Other comparisons that can help with homology inference are defining the developmental origin and neuronal connectivity of a given cell, although these criteria have their own caveats. The concordance of developmental origin, transcriptomic similarity and input–output connectivity is ideal for a solid homology inference.

https://doi.org/10.1038/s41593-024-01814-0

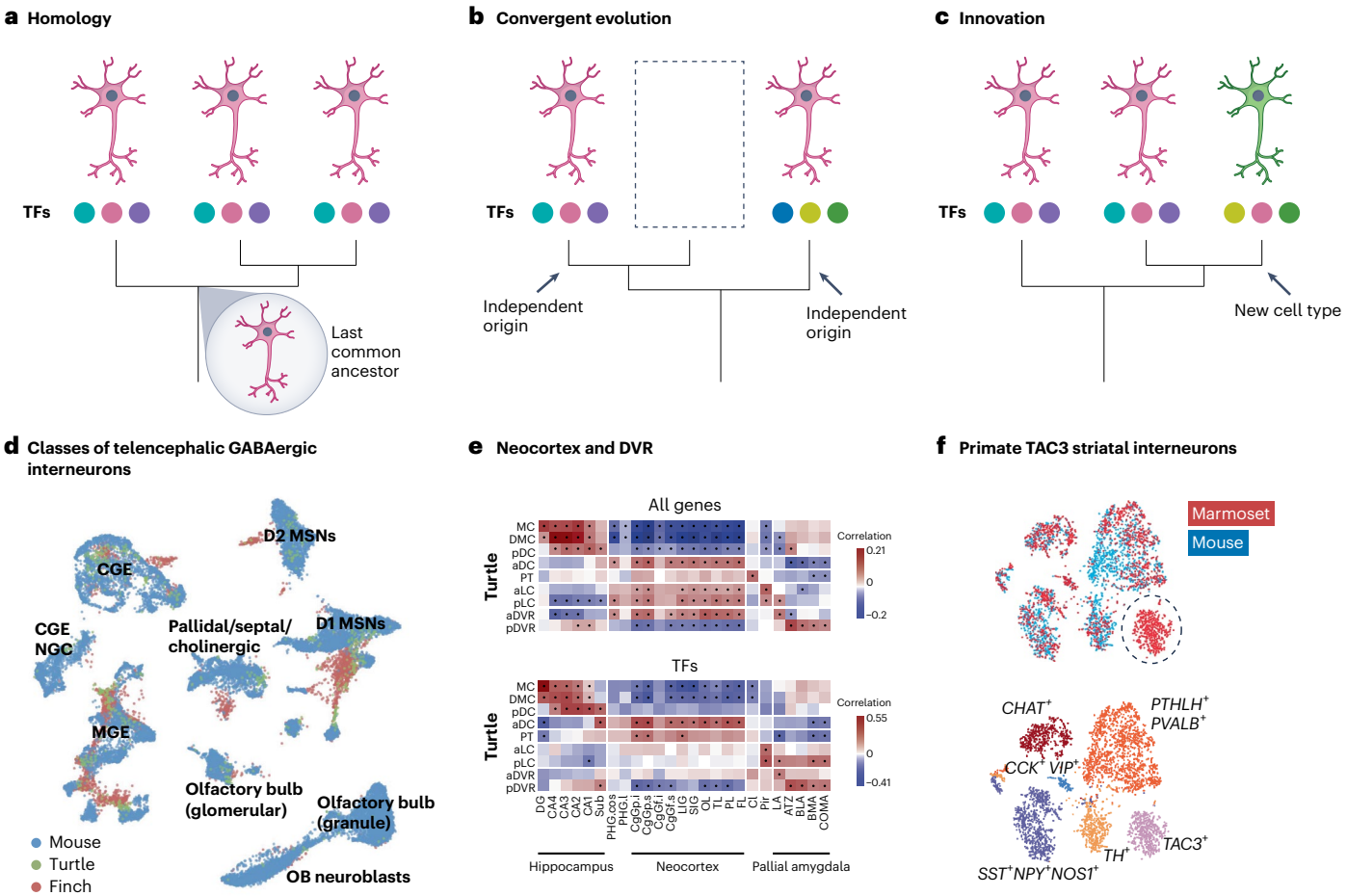

Fig. 3 | Illustration of cell-type homology, convergence and innovation. **a**–**c**, Schematics of cell-type evolution. Circles indicate TFs. **d**–**f**, Examples from the literature using single-cell genomics to address each type of cell-type evolution. CGE, caudal ganglionic eminence; DVR, dorsodorsal ventricular ridge; MGE, medial ganglionic eminence; MSN, medium spiny neuron; NGC, neurogliaform cell; OB, olfactory bulb; TF, transcription factor. Panel **d** adapted with permission from ref. 115, AAAS; panel **e** adapted with permission from ref. 102, AAAS; panel **f** reproduced from ref. 101, Springer Nature Limited.

Consideration for modeling function using in vitro systems

Inaccessibility of neural tissues in humans and other species can limit the types of functional validation that can be performed, making in vitro models necessary tools for validating sc/snRNA-seq data generated from nonexperimental species[120,121]. The experimental tractability of in vitro models makes them an attractive model for the functional characterization of transcriptomics states and changes through perturbation experiments (Fig. 2). Organoids[121] are experimentally tractable systems to model the cell-type heterogeneity[121–127] and spatial organization[128,129] of neural tissues. A striking example of functional investigations made possible include the generation of human neural organoids with or without a single amino acid change found in Neanderthals, enabling study of the neurobiological consequences of genetic variation found in an extinct species[130,131]. Single-cell or single-nucleus dissection of neural organoids can provide cell-type resolution of developmental trajectories[132,133], enable perturbation of dynamic gene regulatory networks[134], model neurological disease mechanisms[135–137] and support neurodevelopmental cross-species comparisons[126,132,138]. The inaccessibility of non-postmortem human neural tissues positions organoids and/or monoculture systems (often) as the only available option for functional investigations, demanding that the limitations of these models be acutely understood.

sc/snRNA-seq comparisons demonstrate the capacity of neural organoids to model broad in vivo neural cell types across numerous genomic modalities[122–124,133,139]. However, these single-cell data comparisons typically use a singular in vivo dataset as a reference, which ignores potential variability within the reference data that organoids may not recapitulate. Especially with sparse and noisy sc/snRNA-seq data, any individual dataset carries error. A neural organoid model that recapitulates signal from a single in vivo dataset may in fact be a poor general model if the reference signal is of low quality and fails to replicate. Therefore, to avoid overfitting, it is useful to incorporate cross-validation of in vivo signal among in vivo datasets[140]. Quantifying DEG statistics across cell types and collating the *P* values and fold changes of genes derived from individual in vivo datasets establishes a benchmark of reference signal for interpreting organoid differential expression statistics.

For cross-species comparisons, sc/snRNA-seq dissection of neural organoids can resolve key developmental differences across species, such as molecular mechanisms underlying neural progenitor variation across human and primate organoids[126,132]. However, observations are robust only to the class of variability sampled, and assessments should be applied to diverse genetic backgrounds (cell lines) or differentiation protocols to identify signals that are not specific to an individual cell line or protocol. As examples, different organoid protocols aiming to derive similar neural lineages (cortical organoids) have reported that biases in differentiation patterns[139] and cell-line-specific effects[141,142] can obscure disease phenotypes in organoid models. Sampling increased genetic and/or technical variability increases the likelihood of replicable signal and may buffer against overfitting.

https://doi.org/10.1038/s41593-024-01814-0

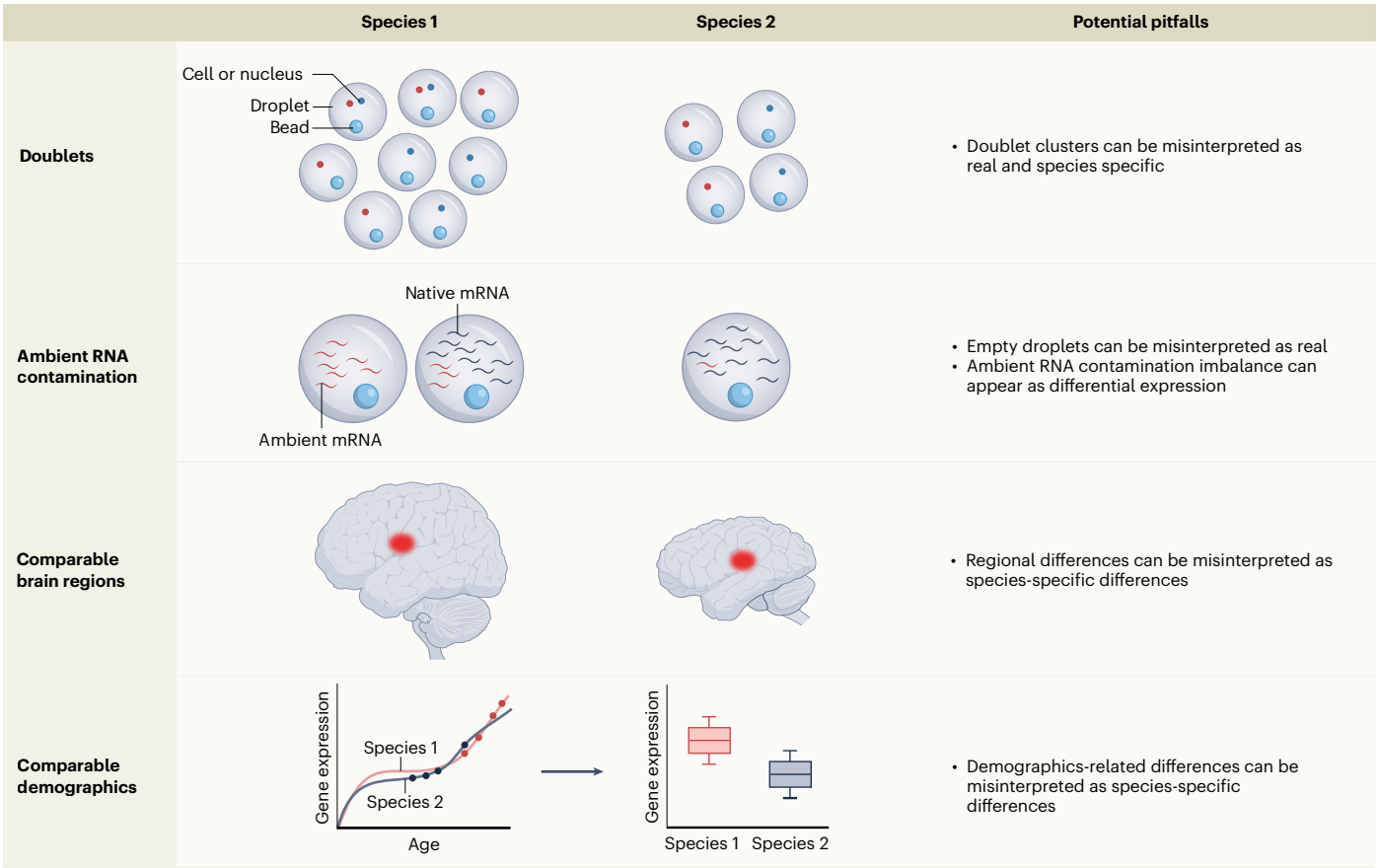

Fig. 4 | Illustration of technical and biological artifacts. Schematics of how evolutionary comparisons using single-cell genomics could be vulnerable to misinterpretations due to either biological (for example, brain region selection/dissection or demographics such as age) and/or technical (for example, doublet or ambient RNA) artifacts. It should be noted that these comparisons and errors could also occur in comparison of other variables: samples of different ages, at different stages of disease, across different CNS regions or following different treatment paradigms. As always, analysis should be considered in the context of the underlying biology being interrogated.

Technical limitations of transcriptomic-based homology inference

While we have attempted to provide conceptual criteria above, there are computational challenges to defining cell-type homologies because there are no formal or uniformly accepted criteria. Multiple methods predict homologous cell types[107,143–148], but integrating across species can be difficult because identification of homologous cell types often relies on heuristics such as shared nearest neighbors and nonlinear data transformations, rather than formal models of gene expression divergence and cell-type evolution[149]. As such, inclusion or exclusion of cell types within a given dataset can alter which cell types appear homologous. For example, a putative primate-specific cell type thought to be most similar to other striatal interneurons[101] was determined to actually be more similar to diencephalic neurons when such cell types were further included in the analysis[150]. This issue is a potential caveat for any type of comparison, whether it is between species, regions or developmental time periods—and is particularly important for the deployment of functional orthogonal validation experiments. Compositional concerns are especially pressing in the context of in vitro studies, in which different iPS cell lines respond divergently to patterning factors and generate cultures with variable compositions. It is also important to consider that conserved populations can be repurposed to different brain structures over development. Recent work highlights that classes of inhibitory neurons that migrate to rodent olfactory bulb have been redirected to the expanded primate white matter[98], and a mammalian conserved interneuron type is most numerous in the mouse hippocampus but more abundant in the primate neocortex[101].

The challenges in using sc/snRNA-seq approaches to study cell types across species are multiplicative. Even with reliable in vivo data, spatiotemporal context and biological variation must be considered when modeling homology. In vitro studies have the same challenges amplified: cell-type distributions are untethered to the anatomy that is reproducibly generated in vivo, with the added concern that the observed cell states approximate those seen in vivo, heavily layered with various sources of technical variation. Despite these challenges, existing data and tools wielded with perspicacious judgment have enabled the discovery of new cell types and shared features and principles of vertebrate brain development and function.

Technical and biological artifacts

As evolutionary findings may be challenging to experimentally validate, it is important to consider experimental factors that could lead to erroneous interpretations. Some of these are pertinent to evolutionary comparisons, but most are generalizable to other types of comparison (and can be mitigated by careful orthogonal validation). Recent studies have shown that technical artifacts such as doublets and ambient RNA contamination can lead to misinterpretations[151]. This issue is exacerbated when datasets are compared without properly adjusting for sequencing artifacts. For example, if datasets for one species contained more artifacts (for example, higher doublet rate, greater ambient RNA contamination), the result could be misinterpreted as a species-specific effect (Fig. 4). Equally, it is crucial to obtain demographically and spatiotemporally similar brain tissues from all species for a proper evolutionary comparison (or samples for within-species

comparisons). If regional boundaries are not rigorously considered during dissection, it is possible to compare improperly matched brain regions, which can lead to misclassification of region-specific cellular and molecular features as species- or sample-specific results. Although spatial transcriptomics may alleviate this problem for species with small brain sizes. In addition to brain regions, developmental time points should be matched to prevent misinterpreting age-specific effects as species-specific effects (Fig. 4). However, matching developmental time points in distantly related species might be impossible, and heterochrony should also be considered as a mechanism for evolutionary change. Finally, it is important to consider that age matching often depends on an estimate based on life history traits, and some cell types may be more sensitive to the effects of age than others (for example, glia change more than neurons in very old age[152]). Thus, interpretation should consider the age bracket of the samples.

Conclusion

We have highlighted the importance of validating sc/snRNA-seq data using in-depth data analysis, functional characterization, cross-validation, multi-omics integration and follow-up validation experiments. We also emphasize the need for specific practices to handle confounds in cross-system analyses, such as sampling broadly within each system, measuring variance, assessing similarity without merging and reporting robustness with effect sizes (see also refs. 1,2). These practices can help avoid overfitting and bias, provide meaningful cross-system assessments, and reveal the molecular mechanisms of brain evolution, disease responses and adaptive phenotypes. By applying sc/snRNA-seq approaches with careful consideration of their inherent challenges and limitations, researchers can advance our understanding of cellular heterogeneity and evolution across different biological systems.

References

1. Adameyko, I. et al. Applying single-cell/nucleus genomics to studies of cellular heterogeneity and cell fate transitions in the nervous system. *Nat. Neurosci.* https://doi.org/10.1038/s41593-024-01827-9 (2024).
2. Bonev, B. et al. Opportunities and challenges of single-cell and spatially resolved genomics methods for neuroscience discovery. *Nat. Neurosci.* https://doi.org/10.1038/s41593-024-01806-0 (2024).
3. Luecken, M. D. & Theis, F. J. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol. Syst. Biol.* **15**, e8746 (2019).
4. Trapnell, C. Defining cell types and states with single-cell genomics. *Genome Res.* **25**, 1491–1498 (2015).
5. Kiselev, V. Y. et al. SC3: consensus clustering of single-cell RNA-seq data. *Nat. Methods* **14**, 483–486 (2017).
6. Hao, Y. et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2024).
7. Nayak, R. & Hasija, Y. A hitchhiker's guide to single-cell transcriptomics and data analysis pipelines. *Genomics* **113**, 606–619 (2021).
8. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
9. Dominguez Conde, C. et al. Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science* **376**, eabl5197 (2022).
10. Xu, C. et al. Automatic cell-type harmonization and integration across Human Cell Atlas datasets. *Cell* **186**, 5876–5891 (2023).
11. Menon, V. Clustering single cells: a review of approaches on high-and low-depth single-cell RNA-seq data. *Brief. Funct. Genomics* **17**, 240–245 (2018).
12. Pancheva, A., Wheadon, H., Rogers, S. & Otto, T. D. Using topic modeling to detect cellular crosstalk in scRNA-seq. *PLoS Comput. Biol.* **18**, e1009975 (2022).
13. Sokol, L. et al. Prioritization and functional validation of target genes from single-cell transcriptomics studies. *Commun. Biol.* **6**, 648 (2023).
14. Lahnemann, D. et al. Eleven grand challenges in single-cell data science. *Genome Biol.* **21**, 31 (2020).
15. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587 (2021).
16. Ding, J. et al. Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. *Nat. Biotechnol.* **38**, 737–746 (2020).
17. Zeng, H. & Sanes, J. R. Neuronal cell-type classification: challenges, opportunities and the path forward. *Nat. Rev. Neurosci.* **18**, 530–546 (2017).
18. Scala, F. et al. Phenotypic variation of transcriptomic cell types in mouse motor cortex. *Nature* **598**, 144–150 (2021).
This paper applies Patch-seq to explore the connection between single-cell gene expression, morphology and electrophysiological properties in neurons of the mouse motor cortex, illustrating the complex relationship between a cell's transcriptional subtype and its functional properties.
19. Clark, I. C. et al. Identification of astrocyte regulators by nucleic acid cytometry. *Nature* **614**, 326–333 (2023).
20. Jung, N. & Kim, T. K. Spatial transcriptomics in neuroscience. *Exp. Mol. Med* **55**, 2105–2115 (2023).
21. Fangma, Y., Liu, M., Liao, J., Chen, Z. & Zheng, Y. Dissecting the brain with spatially resolved multi-omics. *J. Pharm. Anal.* **13**, 694–710 (2023).
22. Hartman, A. & Satija, R. Comparative analysis of multiplexed in situ gene expression profiling technologies. *eLife* **13**, RP96949 (2024).
23. Rodriques, S. G. et al. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).
24. Russell, A. J. C. et al. Slide-tags enables single-nucleus barcoding for multimodal spatial genomics. *Nature* **625**, 101–109 (2024).
This paper describes a new approach for spatial single-nucleus genomic profiling in frozen tissue sections using spatial barcoding, which can be applied to multiple modalities including snRNA-seq, ATAC-seq and T cell antigen receptor sequencing. This tool can be applied for multimodal, orthogonal discovery and validation experiments.
25. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
26. Wang, X. et al. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**, eaat5691 (2018).
27. Ke, R. et al. In situ sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods* **10**, 857–860 (2013).
28. Liu, Y., Beyer, A. & Aebersold, R. On the dependency of cellular protein levels on mRNA abundance. *Cell* **165**, 535–550 (2016).
29. Wegler, C. et al. Global variability analysis of mRNA and protein concentrations across and within human tissues. *NAR Genom. Bioinform* **2**, lqz010 (2020).
30. Buccitelli, C. & Selbach, M. mRNAs, proteins and the emerging principles of gene expression control. *Nat. Rev. Genet.* **21**, 630–644 (2020).
31. Almet, A. A., Cang, Z., Jin, S. & Nie, Q. The landscape of cell-cell communication through single-cell transcriptomics. *Curr. Opin. Syst. Biol.* **26**, 12–23 (2021).
32. Dimitrov, D. et al. Comparison of methods and resources for cell-cell communication inference from single-cell RNA-seq data. *Nat. Commun.* **13**, 3224 (2022).
33. Armingol, E., Officer, A., Harismendy, O. & Lewis, N. E. Deciphering cell-cell interactions and communication from gene expression. *Nat. Rev. Genet.* **22**, 71–88 (2021).

34. Gault, J. et al. Combining native and 'omics' mass spectrometry to identify endogenous ligands bound to membrane proteins. *Nat. Methods* **17**, 505–508 (2020).
35. Wang, J. Y. & Doudna, J. A. CRISPR technology: a decade of genome editing is only the beginning. *Science* **379**, eadd8643 (2023).
36. Gilbert, L. A. et al. Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* **159**, 647–661 (2014).
37. Kearns, N. A. et al. Functional annotation of native enhancers with a Cas9-histone demethylase fusion. *Nat. Methods* **12**, 401–403 (2015).
38. Thakore, P. I. et al. Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat. Methods* **12**, 1143–1149 (2015).
39. Adamson, B. et al. A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell* **167**, 1867–1882 (2016).
40. Dixit, A. et al. Perturb-seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* **167**, 1853–1866 (2016).
41. Datlinger, P. et al. Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* **14**, 297–301 (2017).
42. Wheeler, M. A. et al. Droplet-based forward genetic screening of astrocyte-microglia cross-talk. *Science* **379**, 1023–1030 (2023).
43. Wu, D. et al. Dual genome-wide coding and lncRNA screens in neural induction of induced pluripotent stem cells. *Cell Genom.* **2**, 100177 (2022).
44. Cooper, Y. A. et al. Functional regulatory variants implicate distinct transcriptional networks in dementia. *Science* **377**, eabi8654 (2022).
45. Yang, X. et al. Functional characterization of Alzheimer's disease genetic variants in microglia. *Nat. Genet.* **55**, 1735–1744 (2023).
46. Tian, R. et al. CRISPR interference-based platform for multimodal genetic screens in human iPSC-derived neurons. *Neuron* **104**, 239–255 (2019).
47. Drager, N. M. et al. A CRISPRi/a platform in human iPSC-derived microglia uncovers regulators of disease states. *Nat. Neurosci.* **25**, 1149–1162 (2022).
This paper uses a multiplex CRISPRi and CRISPRa platform using iPS cell microglia to identify genes that alter microglial cell states and functions when perturbed. It also uses scRNA-seq and dataset integration to demonstrate that microglial states in this in vitro system replicate states observed in postmortem human brain tissue.
48. Leng, K. et al. CRISPRi screens in human iPSC-derived astrocytes elucidate regulators of distinct inflammatory reactive states. *Nat. Neurosci.* **25**, 1528–1542 (2022).
49. Li, E. et al. CRISPRi-based screens in iAssembloids to elucidate neuron-glia interactions. Preprint at *bioRxiv* https://doi.org/10.1101/2023.04.26.538498 (2023).
50. Esk, C. et al. A human tissue screen identifies a regulator of ER secretion as a brain-size determinant. *Science* **370**, 935–941 (2020).
51. Jin, X. et al. In vivo Perturb-seq reveals neuronal and glial abnormalities associated with autism risk genes. *Science* **370**, eaaz6063 (2020).
52. Wertz, M. H. et al. Genome-wide in vivo CNS screening identifies genes that modify CNS neuronal survival and mHTT toxicity. *Neuron* **106**, 76–89 (2020).
53. Ramani, B. et al. Scalable, cell type-selective, AAV-based in vivo CRISPR screening in the mouse brain. Preprint at *bioRxiv* https://doi.org/10.1101/2023.06.13.544831 (2023).
54. Wheeler, M. A. et al. Environmental control of astrocyte pathogenic activities in CNS inflammation. *Cell* **176**, 581–596 (2019).
55. Hasel, P., Rose, I. V. L., Sadick, J. S., Kim, R. D. & Liddelow, S. A. Neuroinflammatory astrocyte subtypes in the mouse brain. *Nat. Neurosci.* **24**, 1475–1487 (2021).
56. Buettner, F. et al. Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat. Biotechnol.* **33**, 155–160 (2015).
57. Velten, L. et al. Human haematopoietic stem cell lineage commitment is a continuous process. *Nat. Cell Biol.* **19**, 271–281 (2017).
58. Crow, M., Paul, A., Ballouz, S., Huang, Z. J. & Gillis, J. Characterizing the replicability of cell types defined by single cell RNA-sequencing data using MetaNeighbor. *Nat. Commun.* **9**, 884 (2018).
59. Brunner, G., Lang, K., Wolfe, R. A., McClure, D. B. & Sato, G. H. Selective cell culture of brain cells by serum-free, hormone-supplemented media: a comparative morphological study. *Brain Res.* **254**, 563–575 (1981).
60. Foo, L. C. et al. Development of a method for the purification and culture of rodent astrocytes. *Neuron* **71**, 799–811 (2011).
61. Bohlen, C. J., Bennett, F. C. & Bennett, M. L. Isolation and culture of microglia. *Curr. Protoc. Immunol.* **125**, e70 (2019).
62. Guttenplan, K. A. & Liddelow, S. A. Astrocytes and microglia: models and tools. *J. Exp. Med.* **216**, 71–83 (2019).
63. Hall, C. E. et al. Progressive motor neuron pathology and the role of astrocytes in a human stem cell model of VCP-related ALS. *Cell Rep.* **19**, 1739–1749 (2017).
64. de Majo, M. et al. Granulin loss of function in human mature brain organoids implicates astrocytes in TDP-43 pathology. *Stem Cell Rep.* **18**, 706–719 (2023).
65. Cuomo, A. S. E. et al. Single-cell RNA-sequencing of differentiating iPS cells reveals dynamic genetic effects on gene expression. *Nat. Commun.* **11**, 810 (2020).
66. Virdi, G. S. et al. Protein aggregation and calcium dysregulation are hallmarks of familial Parkinson's disease in midbrain dopaminergic neurons. *NPJ Parkinsons Dis.* **8**, 162 (2022).
67. Szebenyi, K. et al. Human ALS/FTD brain organoid slice cultures display distinct early astrocyte and targetable neuronal pathology. *Nat. Neurosci.* **24**, 1542–1554 (2021).
68. Liddelow, S. A. et al. Neurotoxic reactive astrocytes are induced by activated microglia. *Nature* **541**, 481–487 (2017).
One of the first studies to develop an in vitro functional testing platform to interrogate and validate a transcriptomically defined reactive astrocyte substate.
69. Barbar, L. et al. CD49f is a novel marker of functional and reactive human iPSC-derived astrocytes. *Neuron* **107**, 436–453 (2020).
70. Clark, I. C. et al. Barcoded viral tracing of single-cell interactions in central nervous system inflammation. *Science* **372**, abf1230 (2021).
71. Pasqual, G. et al. Monitoring T cell-dendritic cell interactions in vivo by intercellular enzymatic labelling. *Nature* **553**, 496–500 (2018).
72. Gerfen, C. R., Paletzki, R. & Heintz, N. GENSAT BAC cre-recombinase driver lines to study the functional organization of cerebral cortical and basal ganglia circuits. *Neuron* **80**, 1368–1383 (2013).
73. Gong, S. et al. A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. *Nature* **425**, 917–925 (2003).
74. Gertler, T. S., Chan, C. S. & Surmeier, D. J. Dichotomous anatomical properties of adult striatal medium spiny neurons. *J. Neurosci.* **28**, 10814–10824 (2008).
75. Fenno, L. E. et al. Targeting cells with single vectors using multiple-feature Boolean logic. *Nat. Methods* **11**, 763–772 (2014).
76. Kramer, D. J. et al. Generation of a DAT-P2A-Flpo mouse line for intersectional genetic targeting of dopamine neuron subpopulations. *Cell Rep.* **35**, 109123 (2021).

77. Poulin, J. F. et al. Mapping projections of molecularly defined dopamine neuron subtypes using intersectional genetic approaches. *Nat. Neurosci.* **21**, 1260–1271 (2018).
78. Ren, J. et al. Single-cell transcriptomes and whole-brain projections of serotonin neurons in the mouse dorsal and median raphe nuclei. *Elife* **8**, e49424 (2019).
79. Okaty, B. W. et al. A single-cell transcriptomic and anatomic atlas of mouse dorsal raphe Pet1 neurons. *Elife* **9**, e55523 (2020).
80. BRAIN Initiative Cell Census Network. A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature* **598**, 86–102 (2021).
81. Cadwell, C. R. et al. Electrophysiological, transcriptomic and morphologic profiling of single neurons using Patch-seq. *Nat. Biotechnol.* **34**, 199–203 (2016).
A technical improvement is provided with the advent of PATCH-seq—enabling the electrophysiological recording and transcriptomic measurement in individual neurons.
82. Fuzik, J. et al. Integration of electrophysiological recordings with single-cell RNA-seq data identifies neuronal subtypes. *Nat. Biotechnol.* **34**, 175–183 (2016).
83. Sloan, S. A. et al. Human astrocyte maturation captured in 3D cerebral cortical spheroids derived from pluripotent stem cells. *Neuron* **95**, 779–790 (2017).
84. Xu, S. et al. Behavioral state coding by molecularly defined paraventricular hypothalamic cell type ensembles. *Science* **370**, eabb2494 (2020).
85. Condylis, C. et al. Dense functional and molecular readout of a circuit hub in sensory cortex. *Science* **375**, eabl5981 (2022).
86. Bugeon, S. et al. A transcriptomic axis predicts state modulation of cortical interneurons. *Nature* **607**, 330–338 (2022).
87. Guttenplan, K. A. et al. Knockout of reactive astrocyte activating factors slows disease progression in an ALS mouse model. *Nat. Commun.* **11**, 3753 (2020).
88. Guttenplan, K. A. et al. Neurotoxic reactive astrocytes induce cell death via saturated lipids. *Nature* https://doi.org/10.1038/s41586-021-03960-y (2021).
89. Zeng, H. What is a cell type and how to define it? *Cell* **185**, 2739–2755 (2022).
90. Yuste, R. et al. A community-based transcriptomics classification and nomenclature of neocortical cell types. *Nat. Neurosci.* **23**, 1456–1468 (2020).
91. Hodge, R. D. et al. Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61–68 (2019).
92. Arendt, D., Bertucci, P. Y., Achim, K. & Musser, J. M. Evolution of neuronal types and families. *Curr. Opin. Neurobiol.* **56**, 144–152 (2019).
93. Wagner, D. E. & Klein, A. M. Lineage tracing meets single-cell omics: opportunities and challenges. *Nat. Rev. Genet.* **21**, 410–427 (2020).
94. Bandler, R. C. et al. Single-cell delineation of lineage and genetic identity in the mouse brain. *Nature* **601**, 404–409 (2022).
95. Cadwell, C. R. et al. Cell type composition and circuit organization of clonally related excitatory neurons in the juvenile mouse neocortex. *Elife* **9**, e52951 (2020).
96. Mayer, C. et al. Developmental diversification of cortical inhibitory interneurons. *Nature* **555**, 457–462 (2018).
97. Nowakowski, T. J. et al. Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* **358**, 1318–1323 (2017).
98. Schmitz, M. T. et al. The development and evolution of inhibitory neurons in primate cerebrum. *Nature* **603**, 871–877 (2022).
This paper uses comparative single-cell genomics in mammalian brains over developmental stages, identifying the evolutionary changes in developmental specification of a class of primate-specific interneurons.
99. Woych, J. et al. Cell-type profiling in salamanders identifies innovations in vertebrate forebrain evolution. *Science* **377**, eabp9186 (2022).
100. Hahn, J. et al. Evolution of neuronal cell classes and types in the vertebrate retina. *Nature* **624**, 415–424 (2023).
101. Krienen, F. M. et al. Innovations present in the primate interneuron repertoire. *Nature* **586**, 262–269 (2020).
102. Tosches, M. A. et al. Evolution of pallium, hippocampus, and cortical cell types revealed by single-cell transcriptomics in reptiles. *Science* **360**, 881–888 (2018).
103. Peng, Y. R. et al. Molecular classification and comparative taxonomics of foveal and peripheral cells in primate retina. *Cell* **176**, 1222–1237 (2019).
104. Jorstad, N. L. et al. Comparative transcriptomics reveals human-specific cortical features. *Science* **382**, eade9516 (2023).
105. Hain, D. et al. Molecular diversity and evolution of neuron types in the amniote brain. *Science* **377**, eabp8202 (2022).
This paper shows that generating single-cell genomics brain atlases from non-model organisms (that is, an exemplar lizard) and comparing to mammals can identify conserved neuronal cell types throughout the brains of both species across hundreds of million years of evolution while also showing how evolution has modified this conservation based on connectivity needs.
106. Lust, K. et al. Single-cell analyses of axolotl telencephalon organization, neurogenesis, and regeneration. *Science* **377**, eabp9262 (2022).
107. Tarashansky, A. J. et al. Mapping single-cell atlases throughout metazoa unravels cell type evolution. *Elife* **10**, e66747 (2021).
108. Chen, J. et al. A quantitative framework for characterizing the evolutionary history of mammalian gene expression. *Genome Res.* **29**, 53–63 (2019).
109. Bakken, T. E. et al. Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature* **598**, 111–119 (2021).
This paper used multimodal single-cell approaches (snRNA-seq, snATAC-seq, single-cell methylation, morphology and Patch-seq) across mammalian species to determine the species-specific adaptiations of cell types in one brain region and focuses on differences in Betz cells.
110. Caglayan, E. et al. Molecular features driving cellular complexity of human brain evolution. *Nature* **620**, 145–153 (2023).
This paper demonstrates the power of comparing proportions of nonneuronal cell types (that is, the oligodendrocyte lineage) across closely related primate species and the integration of ancient human genomes to identify modern human-specific variants associated with human-specific chromatin accessibility.
111. Jeon, H. et al. Statistical power analysis for designing bulk, single-cell, and spatial transcriptomics experiments: review, tutorial, and perspectives. *Biomolecules* **13**, 221 (2023).
112. Svensson, V. et al. Power analysis of single-cell RNA-sequencing experiments. *Nat. Methods* **14**, 381–387 (2017).
113. Su, K., Wu, Z. & Wu, H. Simulation, power evaluation and sample size recommendation for single-cell RNA-seq. *Bioinformatics* **36**, 4860–4868 (2020).
114. Arendt, D. et al. The origin and evolution of cell types. *Nat. Rev. Genet.* **17**, 744–757 (2016).
115. Colquitt, B. M., Merullo, D. P., Konopka, G., Roberts, T. F. & Brainard, M. S. Cellular transcriptomics reveals evolutionary identities of songbird vocal circuits. *Science* **371**, eabd9704 (2021).
116. Kebschull, J. M. et al. Cerebellar nuclei evolved by repeatedly duplicating a conserved cell-type set. *Science* **370**, abd5059 (2020).
117. Lamanna, F. et al. A lamprey neural cell type atlas illuminates the origins of the vertebrate brain. *Nat. Ecol. Evol.* **7**, 1714–1728 (2023).

118. Shafer, M. E. R., Sawh, A. N. & Schier, A. F. Gene family evolution underlies cell-type diversification in the hypothalamus of teleosts. *Nat. Ecol. Evol.* **6**, 63–76 (2022).
119. Tarashansky, A. J., Xue, Y., Li, P., Quake, S. R. & Wang, B. Self-assembling manifolds in single-cell RNA sequencing data. *Elife* **8**, e48994 (2019).
120. Agboola, O. S., Hu, X., Shan, Z., Wu, Y. & Lei, L. Brain organoid: a 3D technology for investigating cellular composition and interactions in human neurological development and disease models in vitro. *Stem Cell Res. Ther.* **12**, 430 (2021).
121. Lancaster, M. A. et al. Cerebral organoids model human brain development and microcephaly. *Nature* **501**, 373–379 (2013).
122. Amiri, A. et al. Transcriptome and epigenome landscape of human cortical development modeled in organoids. *Science* **362**, eaat6720 (2018).
123. Bhaduri, A. et al. Cell stress in cortical organoids impairs molecular subtype specification. *Nature* **578**, 142–148 (2020).
124. Camp, J. G. et al. Human cerebral organoids recapitulate gene expression programs of fetal neocortex development. *Proc. Natl Acad. Sci. USA* **112**, 15672–15677 (2015).
125. Gordon, A. et al. Long-term maturation of human cortical organoids matches key early postnatal transitions. *Nat. Neurosci.* **24**, 331–342 (2021).
126. Pollen, A. A. et al. Establishing cerebral organoids as models of human-specific brain evolution. *Cell* **176**, 743–756 (2019).
127. Velasco, S. et al. Individual brain organoids reproducibly form cell diversity of the human cerebral cortex. *Nature* **570**, 523–527 (2019).
128. Qian, X. et al. Sliced human cortical organoids for modeling distinct cortical layer formation. *Cell Stem Cell* **26**, 766–781 (2020).
129. Xiang, Y. et al. Fusion of regionally specified hPSC-derived organoids models human brain development and interneuron migration. *Cell Stem Cell* **21**, 383–398 (2017).
130. Pinson, A. et al. Human TKTL1 implies greater neurogenesis in frontal neocortex of modern humans than Neanderthals. *Science* **377**, eabl6422 (2022).
131. Trujillo, C. A. et al. Reintroduction of the archaic variant of NOVA1 in cortical organoids alters neurodevelopment. *Science* **371**, eaax2537 (2021).
132. Kanton, S. et al. Organoid single-cell genomic atlas uncovers human-specific features of brain development. *Nature* **574**, 418–422 (2019).
133. Uzquiano, A. et al. Proper acquisition of cell class identity in organoids allows definition of fate specification programs of the human cerebral cortex. *Cell* **185**, 3770–3788 (2022).
134. Fleck, J. S. et al. Inferring and perturbing cell fate regulomes in human brain organoids. *Nature* https://doi.org/10.1038/s41586-022-05279-8 (2022).
135. Bershteyn, M. et al. Human iPSC-derived cerebral organoids model cellular features of lissencephaly and reveal prolonged mitosis of outer radial glia. *Cell Stem Cell* **20**, 435–449 (2017).
136. Trevino, A. E. et al. Chromatin accessibility dynamics in a model of human forebrain development. *Science* **367**, eaay1645 (2020).
137. Li, C. et al. Single-cell brain organoid screening identifies developmental defects in autism. *Nature* **621**, 373–380 (2023).
138. Mora-Bermudez, F. et al. Differences and similarities between human and chimpanzee neural progenitors during cerebral cortex development. *Elife* **5**, e18683 (2016).
139. Tanaka, Y., Cakir, B., Xiang, Y., Sullivan, G. J. & Park, I. H. Synthetic analyses of single-cell transcriptomes from multiple brain organoids and fetal brain. *Cell Rep.* **30**, 1682–1689 (2020).
140. Werner, J. M. & Gillis, J. Preservation of co-expression defines the primary tissue fidelity of human neural organoids. Preprint at *bioRxiv* https://doi.org/10.1101/2023.03.31.535112 (2023).
141. Czerminski, J. T., King, O. D. & Lawrence, J. B. Large-scale organoid study suggests effects of trisomy 21 on early fetal neurodevelopment are more subtle than variability between isogenic lines and experiments. *Front. Neurosci.* **16**, 972201 (2022).
142. Lee, J. H. et al. Cell-line dependency in cerebral organoid induction: cautionary observations in Alzheimer's disease patient-derived induced pluripotent stem cells. *Mol. Brain* **15**, 46 (2022).
143. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
144. Haghverdi, L., Lun, A. T. L., Morgan, M. D. & Marioni, J. C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* **36**, 421–427 (2018).
145. Johansen, N. & Quon, G. scAlign: a tool for alignment, integration, and rare cell identification from scRNA-seq data. *Genome Biol.* **20**, 166 (2019).
146. Polanski, K. et al. BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics* **36**, 964–965 (2020).
147. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
148. Welch, J. D. et al. Single-cell multi-omic integration compares and contrasts features of brain cell identity. *Cell* **177**, 1873–1887 (2019).
149. Mostajo-Radji, M. A., Schmitz, M. T., Montoya, S. T. & Pollen, A. A. Reverse engineering human brain evolution using organoid models. *Brain Res.* **1729**, 146582 (2020).
150. Krienen, F. M. et al. A marmoset brain cell census reveals regional specialization of cellular identities. *Sci Adv.* **9**, eadk3986 (2023).
151. Caglayan, E., Liu, Y. & Konopka, G. Neuronal ambient RNA contamination causes misinterpreted and masked cell types in brain single-nuclei datasets. *Neuron* **110**, 4043–4056 (2022).
152. Soreq, L. et al. Major shifts in glial regional identity are a transcriptional hallmark of human brain aging. *Cell Rep.* **18**, 557–570 (2017).
153. Norimoto, H. et al. A claustrum in reptiles and its role in slow-wave sleep. *Nature* **578**, 413–418 (2020).
154. Marsh, S. E. et al. Dissection of artifactual and confounding glial signatures by single-cell sequencing of mouse and human brain. *Nat. Neurosci.* **25**, 306–316 (2022).

Acknowledgements

We thank I. Adameyko, T. Bakken, A. Bhaduri, B. Bonev, G. Castelo-Branco, F. Chen, C. Chhatbar, S. Codeluppi, R. Corces, J. Fan, M. G. Filbin, D. Gate, G. Green, M. Heiman, K. Harris, H. Hochgerner, F. Inoue, M. Kellis, C. N. Kim, J. Krull, G. L. Manno, A. Levine, Q. Li, S. Linnarsson, M. Lotfollahi, C. Luo, Q. Ma, E. Macosko, C. Mayer, K. R. Maynard, V. Menon, P. Nano, M. Nitzan, M. Prinz, S. Quake, V. Ramani, R. Satijia, L. Schirmer, Y. Shen, N. Sun, F. Theis, C. Walsh, X. Wang, J. D. Welch and J. Yang for the insightful feedback on this work. M.C. thanks A. U. Antonova for helpful additional discussions. G.K. is a Jon Heighten Scholar in Autism Research and Townsend Distinguished Chair in Research on Autism Spectrum Disorders at UT Southwestern, and was partially supported by the James S. McDonnell Foundation 21st Century Science Initiative in Understanding Human Cognition Scholar Award (220020467), Simons Foundation for Autism Research Award (947591) and National Institutes of Health (NIH) grants (NS126143, HG011641, MH126481, NS115821). S.A.L. is supported by the Carol and Gene Ludwig Family Foundation, The Cure Alzheimer's Fund, the National Multiple Sclerosis Society and NIH grants (R01EY033353, R03NS127079). C.R.C. is supported by the Weill Neurohub, the Shurl and Kay Curci Foundation, the CURE Epilepsy Taking Flight Award and grants from the NIH (K08NS126573,

U01NS132353). E.C. is a Neural Scientist Training Program Fellow in the Peter O'Donnell Brain Institute at UT Southwestern. J.L.C. is supported by NIH grant U01MH10907. J.G. is supported by NIH grant R01MH113005. M.S. is supported by a HHMI Hanna H. Gray Fellowship, a SFARI Pilot Award and a NYSCF Druckenmiller Fellowship. J.W. is supported by NIH grant R01MH113005. M.E.U. is supported by the National Science Foundation Graduate Research Fellowship Program and a Hertz Foundation Fellowship Program.

Author contributions

All authors contributed to the writing of this Review.

Competing interests

M.C. is a member of the Scientific Advisory Board of Vigil, NGMBio, Cartesian and Halyardtx. M.C. receives research support from Ono Pharmaceutical, is a consultant for CST and has patents pending on LILRB4 and TREM2. S.A.L. maintains a financial interest in AstronauTx and Synapticure and is on the Scientific Advisory Board of the Global BioAccess Fund. S.A.L. is an inventor on US Patents WO2018081250A1 and WO2022187517A1. M.K. is a co-scientific founder of Montara Therapeutics and serves on the Scientific Advisory Boards of Engine Biosciences, Casma Therapeutics, Cajal Neuroscience, Alector and Montara Therapeutics, and is an advisor to Modulo Bio and Recursion Therapeutics. M.K. is an inventor on US Patent 11,254,933 related to CRISPRi and CRISPRa screening, and on a US Patent Application on in vivo screening methods. All other authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Marco Colonna, Genevieve Konopka, Shane A. Liddelow, Tomasz Nowakowski, Omer A. Bayraktar, Ozgun Gokce or Naomi Habib.

Peer review information *Nature Neuroscience* thanks Renzo Mancuso and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

[1]Department of Pathology and Immunology, Washington University School of Medicine in St. Louis, St. Louis, MO, USA. [2]Department of Neuroscience, Peter O'Donnell Jr. Brain Institute, UT Southwestern Medical Center, Dallas, TX, USA. [3]Neuroscience Institute, NYU Grossman School of Medicine, New York, NY, USA. [4]Department of Neuroscience & Physiology, NYU Grossman School of Medicine, New York, NY, USA. [5]Department of Ophthalmology, NYU Grossman School of Medicine, New York, NY, USA. [6]Parekh Center for Interdisciplinary Neurology, NYU Grossman School of Medicine, New York, NY, USA. [7]Department of Neurological Surgery, University of California, San Francisco, San Francisco, CA, USA. [8]Eli and Edythe Broad Center for Regeneration Medicine and Stem Cell Research, University of California, San Francisco, San Francisco, CA, USA. [9]Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA, USA. [10]Department of Anatomy, University of California, San Francisco, San Francisco, CA, USA. [11]Department of Microbiology and Immunology, Northwestern University, Chicago, IL, USA. [12]Department of Molecular and Cellular Biology, University of California, Berkeley, Berkeley, CA, USA. [13]Department of Neuroscience, University of California, Berkeley, Berkeley, CA, USA. [14]Chan Zuckerberg Biohub, San Francisco, CA, USA. [15]Department of Pathology, University of California, San Francisco, San Francisco, CA, USA. [16]Kavli Institute for Fundamental Neuroscience, University of California, San Francisco, San Francisco, CA, USA. [17]Department of Biomedical Engineering, Boston University, Boston, MA, USA. [18]Center for Neurophotonics, Boston University, Boston, MA, USA. [19]Department of Biology, Boston University, Boston, MA, USA. [20]Center for Systems Neuroscience, Boston University, Boston, MA, USA. [21]Department of Physiology and Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario, Canada. [22]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA. [23]Institute for Neurodegenerative Diseases, University of California, San Francisco, San Francisco, CA, USA. [24]Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, CA, USA. [25]Princeton Neuroscience Institute, Princeton University, Princeton, NJ, USA. [26]F.M. Kirby Neurobiology Center, Boston Children's Hospital, Boston, MA, USA. [27]Harvard Medical School, Boston, MA, USA. [28]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [29]Department of Neurology and Neurological Sciences, Stanford University, Stanford, CA, USA. [30]Howard Hughes Medical Institute, Stanford University, Stanford, CA, USA. [31]Department of Neuromuscular Disease, UCL Queen Square Institute of Neurology, London, UK. [32]The Francis Crick Institute, Human Stem Cells and Neurodegeneration Laboratory, London, UK. [33]Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, University of California, San Francisco, San Francisco, CA, USA. [34]Department of Neurology, University of California, San Francisco, San Francisco, CA, USA. [35]Ann Romney Center for Neurologic Diseases, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [36]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [37]Department of Genetics and Genome Sciences, Case Western Reserve University School of Medicine, Cleveland, Ohio, OH, USA. [38]Institute for Glial Sciences, Case Western Reserve University School of Medicine, Cleveland, OH, USA. [39]Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA. [40]Department of Biological Sciences, Columbia University, New York, NY, USA. [41]Biomedical Sciences Graduate Program, University of California, San Francisco, San Francisco, CA, USA. [42]Wellcome Sanger Institute, Cambridge, UK. [43]Department of Old Age Psychiatry and Cognitive Disorders, University Hospital Bonn, Bonn, Germany. [44]German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany. [45]Edmond & Lily Safra Center for Brain Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel. [46]Present address: Institute for Glial Sciences, Case Western Reserve University School of Medicine, Cleveland, OH, USA.
✉e-mail: mcolonna@wustl.edu; genevieve.konopka@utsouthwestern.edu; shane.liddelow@nyulangone.org; tomasz.nowakowski@ucsf.edu; ob5@sanger.ac.uk; ozgun.goekce@ukbonn.de; naomi.habib@mail.huji.ac.il