ASTRA: A Deep Learning Framework for Astrocyte Modeling and Therapeutic Target Discovery in Alzheimer's Disease

Soham Samanta

soham2020sam@gmail.com

Mentor: Manolis Kellis

ASTRA: A Deep Learning Framework for Astrocyte Modeling and Therapeutic Target Discovery in Alzheimer's Disease

Soham Samanta

Abstract

Alzheimer's disease (AD) progressively impairs memory, cognition and independence. Astrocytes support neuronal metabolism, synaptic function and brain homeostasis. Identifying ways to reverse their disease-associated changes is an important step toward astrocyte-directed treatments. Single-cell foundation models learn reusable expression features from large cell atlases, creating opportunities to connect cell states to disease biology. Applying these features to therapeutic discovery requires models that retain the disease-associated differences within a specific cell population. Here, we present ASTRA, a deep learning framework that models human astrocyte states and identifies candidate targets for AD reversal. We analyze 40 000 astrocyte nuclei from 422 donors. Three neural-network encoders transform each expression profile into a shared 608-dimensional embedding: a vector of learned features that summarizes the astrocyte's expression state. We evaluate nine prediction tasks covering expression state, astrocyte subtype, and donor-level pathology or cognitive decline, using 25 splits with separate training and test donors. ASTRA achieves the best aggregate mean rank (3.18) among the compared representations. It also exceeds four pretrained single-cell models—Geneformer, scGPT, Nicheformer and SCimilarity—in mean performance on every endpoint when we keep their feature-extracting encoders fixed and train prediction models on the resulting features. We predict a reference score that measures how closely an astrocyte's expression resembles a disease-associated pattern. ASTRA achieves Spearman $\rho = 0.9252$, compared with 0.8382 for the count-based model scVI. We develop a target-discovery method that uses predictions for held-out donors to identify genes associated with this score. We account for differences in average scores between donors and in how strongly genes are expressed. We prioritize genes that consistently rise or fall with the score across donor groups, identifying candidates for reduced or increased expression. We evaluate the rankings across six broad astrocyte populations and ten brain regions in an independent atlas. ASTRA-derived rankings exceed the reference ranking in 57 of 60 comparisons and the case–control ranking in 60 of 60. We prioritize *GFAP* and *PLCE1* for reduced expression and *RERG* and *WIF1* for increased expression. We examine these targets using regional expression, pathway and regulatory analyses, published responses to interventions, and protein structures. Together, the findings identify routes to lower target RNA, alter enzyme activity, increase target expression or supply a target protein. Our work connects astrocyte-state prediction to specific targets and intervention directions for drug discovery. The framework could also support therapeutic research in other neurodegenerative diseases using suitable cohorts and disease-specific expression patterns.

Keywords: Alzheimer's disease; astrocytes; deep learning; single-cell foundation models; therapeutic target discovery; gene-expression reversal.

Table of Contents

1 Introduction **4**

2 Methods **5**

3 Results **14**

4 Discussion **27**

Data availability **28**

Acknowledgments **28**

References **28**

1 Introduction

Alzheimer's disease (AD) progressively disrupts memory, reasoning and the ability to carry out everyday activities, with consequences for patients' independence and the well-being of their families. An estimated 57 million people were living with dementia worldwide in 2021; AD is its most common cause and contributes to approximately 60–70% of cases. Dementia also imposes substantial demands on health systems and informal caregivers, with an estimated global economic cost of US$1.3 trillion in 2019 [1]. Developing interventions that preserve cognition and daily function is therefore a major biomedical priority. This requires identifying molecular processes that treatments can alter to preserve brain function.

AD develops through interacting changes in neurons, glia and their surrounding tissue environment. We use the term astrocyte state to describe a pattern of gene expression reflecting the cell's molecular activities. Single-nucleus profiling measures RNA from individual nuclei and allows these patterns to be compared within and across donors. Human single-nucleus studies reveal considerable variation in disease-associated expression across cell types, brain regions and individuals, including molecular features associated with pathology and cognitive resilience [2, 3, 4, 5]. These differences help identify which cellular processes a treatment should address. Disease-associated expression patterns reveal altered processes within a cell population. Comparing donors and regions helps distinguish recurring changes from responses specific to a particular setting. Models that capture these relationships can identify molecular changes to test in intervention studies.

Astrocytes are particularly important to this approach because their activities support multiple aspects of neuronal function. They supply metabolic substrates, regulate ions and neurotransmitters outside cells, support synapses and interact with the brain's blood vessels [6, 7]. These functions link astrocyte biology to the maintenance of neural circuits. Altered astrocyte states can change the balance of neuronal support, inflammatory signaling and tissue repair. Experimental work has shown that microglial inflammatory signals can induce astrocyte states with impaired neuronal support and neurotoxic activity [8]. Restoring beneficial astrocyte functions could therefore also benefit the neurons and other glial cells that interact with them.

Single-nucleus transcriptomics has established a detailed view of astrocyte diversity in aging and AD. Disease-associated astrocytes emerge early and increase with progression in an AD mouse model, with related populations observed in aged mice and human brains [9]. Human atlases describe how astrocyte expression varies with pathology, cognitive function and brain region [3, 4, 5]. These studies motivate the next step: identifying genes whose manipulation could change an astrocyte's state. A target nomination should identify both the gene and whether to reduce or increase its expression. This information can shape the choice of perturbation, the expected molecular response and the design of subsequent functional experiments.

Therapeutic development also requires connecting a nominated gene to an appropriate intervention mechanism. Depending on the target, an intervention could lower its RNA, change its enzyme activity, increase its expression or supply the protein directly. Published gene-expression responses can identify interventions that oppose a disease-associated pattern. Protein structures and experimental studies can then help explain how to act on a particular target. Combining this evidence connects a ranked gene list to practical routes for developing and testing interventions. For astrocytes, this strategy can focus therapeutic research on reversing AD-associated expression programs while assessing restoration of the functions that support neurons.

Single-cell foundation models offer a way to learn from the growing scale and diversity of cell atlases. These models undergo an initial training stage, called pretraining, on large collections of expression profiles. These training collections can contain millions of cells from many tissues. A common learning task is to hide part of a cell's expression profile and ask the model to recover the missing information. This encourages it to learn relationships among genes. The resulting numerical features summarize patterns across genes and cells. Researchers can then reuse these features, or embeddings, in new studies rather than learn every relationship from the new dataset alone. This transfer of information supports cell-type annotation, comparison of cells across studies and prediction of cellular responses to perturbations [10, 11, 12]. For therapeutic research, the opportunity is to connect a cell's expression pattern to the processes that distinguish disease-associated states and to genes that may alter those states.

The models we compare illustrate several approaches to this problem. Geneformer learns from ranked gene-expression profiles and uses the relationships among genes to support predictions about gene networks [10]. scGPT learns from gene identities and expression values, with applications including cell annotation, integration of datasets and perturbation-response prediction [11]. Both use transformers, neural networks that learn which parts of an input are informative in relation to one another. Nicheformer extends this approach to human and mouse data from both dissociated cells and spatial assays, which measure expression within tissue. Its training supports predictions about a cell's surrounding tissue context [13]. SCimilarity uses a different strategy: it learns a measure of similarity that places transcriptionally related cells near one another, enabling searches for comparable cell states across studies [12]. Together, these developments broaden single-cell modeling from recognizing cell identity to studying gene relationships, tissue context and shared disease-associated states.

How a pretrained model is used also matters. In a frozen-encoder analysis, we keep the parameters that convert expression into features unchanged. We then train a separate prediction model on those features for each task. This tests how much useful information the pretrained representation already contains. Fine-tuning instead updates the pretrained encoder itself using data from the new task.

These settings answer different questions about transfer to a new study [11, 13]. Here, we evaluate Geneformer, scGPT, Nicheformer and SCimilarity as fixed feature extractors, with downstream prediction models fitted using training donors. We compare this use of pretrained features with ASTRA features learned from the study's training donors and with established expression representations, including scVI, which models gene counts probabilistically [14].

Large-scale pretraining provides a valuable starting point, but its usefulness depends on the biological question and evaluation setting. Independent evaluations of Geneformer and scGPT report variable performance across datasets and tasks, with simpler methods sometimes performing better when pretrained features are used without further adaptation [15, 16]. Differences in the training data, expression measurements and learning objectives can affect which biological signals a representation preserves. For AD, distinguishing astrocytes from other cell types is only one requirement. Target discovery also requires resolving expression differences among astrocytes and relating them to disease-associated outcomes. We therefore evaluate state prediction, subtype discrimination and donor outcomes separately, then test the resulting gene rankings in an independent atlas. This connects model evaluation to the biological decisions that the target-discovery workflow is designed to support.

Here, we introduce ASTRA, which combines three variational and deterministic encoder blocks to model human astrocyte states. ASTRA uses deep representation learning: successive neural-network layers learn nonlinear combinations of gene-expression values. The three encoder blocks produce 128, 256 and 224 features, which we concatenate into one 608-dimensional embedding for each astrocyte. Each feature summarizes a learned pattern across the input genes. This shared embedding connects the deep learning model to the downstream analyses: prediction models use its features to estimate cell states, distinguish subtypes and predict donor outcomes. We then use the predicted cell-state score to rank candidate therapeutic targets.

We evaluate 40 000 astrocyte nuclei from 422 donors across nine endpoints and 25 donor-disjoint splits. These tasks test whether the model can predict each astrocyte's expression-state score, distinguish annotated subtypes, and predict pathology or cognitive decline from a donor's astrocytes. ASTRA achieves the leading aggregate rank within this benchmark and higher mean performance on every endpoint than the four pretrained models used as fixed feature extractors. We then link the expression-state predictions to gene expression to identify intervention directions. We nominate genes that rise with the disease-aligned score for reduced expression and genes that fall with it for increased expression. Our method uses predictions for donors excluded from training. It accounts for differences in average scores between donors and for gene expression levels, then gives greater weight to associations that agree in direction across donor groups. The fixed reference score defines the expression pattern that these proposed interventions aim to oppose. Across six broad astrocyte populations and ten brain regions in an independent atlas, the resulting rankings exceed the reference in 57 of 60 comparisons.

We interpret these predictions using regulatory gene sets, published expression responses to interventions, and protein structures. Through these analyses, we prioritize *GFAP*/*PLCE1* suppression and *RERG*/*WIF1* induction and connect each nomination to a specific intervention strategy. This connects human astrocyte expression data to the selection of targets and intervention directions. ASTRA therefore combines state-of-the-art performance within the evaluated benchmark with a framework for developing treatments around specific astrocyte targets.

More broadly, our approach links disease-associated cell states to therapeutic development. Reactive astrocyte alterations are also observed in other neurodegenerative disorders, including Parkinson's disease and amyotrophic lateral sclerosis [8]. Our modeling and ranking strategy could extend to these settings through disease-specific reference programs, cohorts and independent evaluation. We focus here on AD, while developing an approach for studying how targeted perturbations might redirect cellular programs across diseases. This offers a path toward treatments that restore cellular function, starting with the reversal of disease-associated astrocyte states.

2 Methods

We organize the analysis into four stages. We first prepare astrocyte expression profiles and donor annotations. We then learn and compare cell representations using separate training and test donors. Next, we associate predicted expression states with genes and evaluate the gene rankings in an independent atlas. Finally, we interpret the nominated genes through pathway and regulatory enrichment, published perturbation responses, and target-specific experimental and structural evidence.

2.1 Cohort and donor measurements

We analyze 40 000 astrocyte nuclei from 422 participants in the Religious Orders Study and Rush Memory and Aging Project (ROSMAP), using the accompanying donor characteristics and astrocyte subtype annotations [17, 3]. Each nucleus contributes an RNA expression profile, while each donor contributes multiple nuclei and a set of clinical or pathological measurements. We use cell-level to refer to measurements from individual nuclei. We use a fixed panel of 3 000 genes as encoder inputs and normalize their counts as $x_{ig} = \log(1 + 10^4 c_{ig}/\ell_i)$, where c_{ig} is the count for gene g in nucleus i and ℓ_i is its library-size factor. We use raw counts for the negative-binomial reconstruction objective and consider a broader set of 15,051 expressed genes for target ranking. The median number of nuclei per donor is 79 (range 2–414).

Four donor-level outcomes are available for 371 of the 422 donors. Global pathology burden summarizes plaque and neurofibrillary-tangle pathology across sampled brain regions [18]. Amyloid load measures amyloid deposition, and tangle density measures tau-associated neurofibril-

lary tangles. Cognitive-decline slope describes the rate of change in cognitive performance over time [17, 3]. We use these outcomes to test whether astrocyte expression contains information about tissue pathology and cognitive change. Each outcome characterizes an individual donor and complements the cell-level expression measurements. To predict these outcomes, we combine the astrocyte embeddings into one representation per donor. The donor remains the independent sampling unit, even when many nuclei come from that individual. We use these outcomes to evaluate the representation. For gene ranking, we use the reference expression score described below.

2.2 Expression processing and analytical inputs

We retain the donor and subtype annotations alongside each expression profile so that cell-level measurements can be linked to the correct individual. The expression matrix provides the molecular input, subtype annotations define the cell-classification tasks, and donor measurements define the pathology and cognitive-outcome tasks. We use the cohort's existing astrocyte annotations and fixed gene panel. These inputs define the cell population and expression features available for representation learning.

The normalization above has two steps. Dividing a gene count by the recorded library-size factor and multiplying by 10^4 places cells on a comparable expression scale. Taking $\log(1 + x)$ compresses large values while retaining zero counts as zeros. The normalized values enter the ASTRA encoders, whereas the original counts enter the count-reconstruction objective. These two forms of the data serve different purposes: the encoder receives a numerically manageable input, and the decoder is evaluated against the observed counts.

We distinguish the 3,000 input genes used to learn the ASTRA representation from the 15,051 expressed genes considered for target discovery. Once we have predicted a score for each cell, we can compare that score with expression of genes across the broader set. A gene therefore enters target ranking through its association with the predicted state, even when it is outside the encoder input panel. For comparison with SEA-AD, we use the 14,420 genes shared by the discovery and regional analyses.

Where a transformation is learned within a training split, we estimate its parameters from the training data and apply the fitted transformation to test data. For example, predictor standardization uses the training feature means and scales. We describe the separately prepared, full-cohort comparator representations below because their access to expression data differs from ASTRA's split-specific training. Keeping these stages explicit distinguishes processing a cell's measurements from using that donor's outcomes to fit a prediction model.

2.3 Reference expression score

For target discovery, we need a score that measures how closely each astrocyte follows a specified disease-associated expression pattern. We use a fixed reference score that compares each nucleus with disease-related and control text descriptions in a shared numerical space. The target-ranking procedure seeks genes whose manipulation could oppose this pattern. A linear projection maps 50 expression principal components (PCs) to 256 dimensions, and a second projection maps 384-dimensional text embeddings to the same space. Let D and C denote the disease and control descriptions, u_i the projected expression vector, and v_p the projected vector for description p. The cosine function measures similarity between the projected expression and text vectors. We calculate the reference score as

$$r_i = |D|^{-1} \sum_{p \in D} \cos(u_i, v_p) - |C|^{-1} \sum_{p \in C} \cos(u_i, v_p).$$

Higher values indicate greater similarity to the disease descriptions than to the control descriptions. We refer to variation along this score as the *reference axis*. We construct the score from expression and text information without donor, subtype or outcome labels and hold it fixed for all subsequent comparisons. Principal component analysis (PCA) summarizes expression variation through weighted combinations of genes. Because these PCs also supply the input for constructing the reference score, predicting that score from PCA partly recovers its own construction. We therefore evaluate PCA on the other eight endpoints.

The reference score provides a continuous ordering of expression states. Cosine similarity compares the direction of two numerical vectors, so the score contrasts alignment with disease descriptions against alignment with control descriptions. We use this molecular ordering as the prediction target for the state model. The four donor-level outcomes provide separate tests of how the learned representation relates to measured pathology and cognitive change.

2.4 ASTRA implementation and model development

Encoder blocks

ASTRA implements deep learning through multilayer neural networks that transform the 3,000-gene expression input into learned features. During training, we update the network weights using the reconstruction and regularization objectives described below. The nonlinear transformations allow the features to combine expression measurements in ways that a single linear projection cannot. The final embedding is the output of this feature-learning stage; the downstream probes use it to make task-specific predictions.

The three encoder blocks learn complementary summaries of the same expression input. Each latent dimension is a learned feature that can combine information from many genes. A variational branch represents a cell through a probability distribution over features, whereas a deterministic branch produces a fixed feature vector. We build each encoder from linear layers followed by batch normalization, rectified linear activation and dropout, with hidden width 512 and dropout probability 0.1. Our variational block has one hidden layer and 128 Gaussian latent dimensions. Our split block has separate one-hidden-layer branches with 128 variational and 128 deterministic dimensions. Our deeper split block has two hidden layers per branch and supplies 96 variational and 128 determin-

istic dimensions. At inference, we concatenate the variational posterior means and deterministic features to obtain a 608-dimensional representation. The dimension follows directly from the block outputs: $128 + 256 + 224 = 608$. Concatenation places the three feature vectors next to one another, so each cell receives 608 numerical values. The network learns how these values depend on gene expression during training. Thus, 608 describes the size of the learned representation, while the original 3,000 genes define its molecular input.

A linear layer forms weighted combinations of its inputs. Rectified linear activation retains positive values and sets negative values to zero, introducing nonlinear behavior. Batch normalization rescales hidden activations using batch statistics during training. The variational posterior mean supplies a representative feature value from the distribution learned for a cell, while a deterministic branch directly returns its feature vector.

Count reconstruction and donor information

Each block has a decoder that maps its learned features back to gene counts. Reconstructing the input encourages the representation to retain information about expression, and the negative-binomial model accommodates count variability beyond a simple Poisson model. A softmax converts decoder outputs to gene proportions, which we scale by library size. The softmax produces nonnegative proportions that sum to one. The dispersion parameter controls variability of a gene's counts around its expected value. We learn gene-specific dispersion parameters and clamp them numerically for stability. We include a 16-dimensional donor embedding in the decoder to account for donor-associated variation during reconstruction and omit it from the cell representation used for prediction.

Training objectives

We use several training objectives alongside reconstruction. Kullback–Leibler regularization constrains the variational distribution. Teacher distillation encourages the learned features to agree with fixed teacher embeddings. Donor-adversarial training discourages the model from encoding donor identity, while contrastive learning encourages similar representations for two perturbed views of the same cell.

Each objective addresses a different property of the representation. Reconstruction rewards features that preserve observed expression. Kullback–Leibler regularization keeps the learned variational distribution close to its prior distribution, limiting unconstrained variation in the latent space. Distillation compares features with fixed teacher representations and encourages the encoder to retain information present in those references. In adversarial training, a donor classifier attempts to identify the donor while the encoder receives a reversed gradient that discourages this identification. Contrastive learning compares two altered views of the same cell and encourages them to remain similar. The temperature parameter controls how strongly the contrastive comparison emphasizes differences in similarity.

For all blocks, we use a Kullback–Leibler regularization weight of 0.25, increased over the first five epochs, and a teacher-distillation weight of 0.2. We perform distillation through linear prediction heads that match two fixed, standardized teacher representations. For the variational block, we additionally use donor-adversarial and contrastive weights of 0.3 and 0.25, respectively. We apply gradient reversal to posterior means to discourage donor identification [19]. We generate contrastive views by binomial thinning, which randomly retains counts with probability 0.5. We compare the resulting views at temperature 0.2 and exclude different cells from the same donor from negative pairs, the pairs that the objective otherwise pushes apart.

For the split block, we use a weight of 0.5 for linear reconstruction from its deterministic branch and disable adversarial and contrastive terms. For the deeper block, we retain linear reconstruction and donor-adversarial training and disable contrastive loss. We standardize expression for linear reconstruction using training cells.

Optimization

We optimize with Adam, using learning rate 10^{-3}, weight decay 10^{-6}, batch size 1024, gradient-norm clipping at 5 and ten training epochs.

An epoch is one pass through the training data, processed in batches. Adam adjusts the optimization step for each parameter using the history of gradients. Weight decay discourages unnecessarily large weights, and gradient clipping limits unusually large updates. Dropout randomly removes a fraction of hidden activations during training to reduce reliance on individual features. At inference, we use the fitted encoders to produce a fixed cell representation and fit the downstream prediction probes separately.

Architecture comparisons

During architecture development, we compare 64 single-branch, 48 shared-trunk and 32 independent-trunk configurations. In the initial selection, we consider reference-score prediction, within-donor subtype classification and donor mixing, measured by donor entropy. This mixing measure describes how broadly donor identities are distributed in the representation. In later rounds, we consider reference-score prediction, both within-donor subtype tasks and global-pathology prediction. We also use these benchmark results to select the combination of encoder blocks, so our evaluation includes effects of architecture selection.

To assess each block's contribution, we remove it from the representation and refit downstream probes. This block-deletion analysis asks how much predictive information the remaining blocks retain. We compare each change to the training objective with its parent configuration on the same evaluation splits. These comparisons use 24 matched splits for split-block variants and 25 for the other variants. We express effects as variant minus parent performance. For dimension-matched controls, we project ASTRA to 30, 50 or 128 dimensions using standardization and PCA fitted on training cells. These controls test whether performance persists when fewer numerical features are available to the prediction model. We use canon-

ical correlation analysis (CCA) to measure shared directions between block pairs on held-out cells, with training-fold singular-value-decomposition whitening and a 10^{-6} relative singular-value cutoff. We then use ridge regression with penalty 10 to predict each block from the other two; the remaining variance measures information that those blocks do not explain. A larger residual-variance fraction means that the other blocks explain less of that block's variation. Block deletion tests a different question: how much the block contributes to prediction. We include each block/fold once in this analysis.

CCA finds weighted combinations of features from two blocks that are maximally correlated. It therefore measures shared information even when the original feature coordinates differ. The residual-variance analysis asks how much variation remains in a block after the other two blocks have been used to predict it. We interpret these measures together with block deletion: shared variation describes similarity between representations, whereas a change in prediction after deletion describes the block's contribution to a task.

2.5 Comparison models

We evaluate each representation through the prediction tasks and donor partitions described below. Cell embeddings enter the state and subtype probes individually. For donor-level outcomes, we average the cell embeddings within each donor before fitting the outcome probe. The following model-specific descriptions distinguish the source of each representation from the prediction model that we train on top of it.

Geneformer

Geneformer encodes a cell as a sequence of genes ordered by expression relative to a reference distribution. Its transformer learns relationships among these gene tokens, where a token is the numerical identifier used to represent a gene in the model [10]. We use pretrained Geneformer cell embeddings as fixed inputs to our evaluation. We fit state and subtype probes using embeddings from training donors and apply those probes to test donors. For the four donor outcomes, we average the cell embeddings within each donor. This evaluates the information in the pretrained representation under the same downstream task definitions used for ASTRA.

scGPT

scGPT combines gene identities with expression values to learn representations of cells and genes [11]. We evaluate its pretrained cell embeddings with the encoder parameters held fixed. The cell-level embeddings enter the reference-score and subtype probes, while their donor means enter the pathology and cognitive-outcome probes. We fit the downstream models using training donors. The scGPT comparison therefore concerns transfer of its expression features to our astrocyte tasks through these fitted probes.

Nicheformer

Nicheformer learns a common expression representation from dissociated single-cell and spatially measured profiles, with training data spanning human and mouse tissues [13]. We use its pretrained cell embeddings as fixed features in the astrocyte benchmark. We fit the cell-level prediction probes on training donors and use donor-mean embeddings for donor outcomes. This asks whether features learned across expression assays retain information useful for astrocyte-state, subtype and donor-outcome prediction in our cohort.

SCimilarity

SCimilarity uses metric learning to place transcriptionally similar cells near one another in a shared numerical space. Its published training objective combines cell-type relationships with expression reconstruction [12]. We use the pretrained cell representation without updating its encoder. We then fit the same types of state, subtype and donor-outcome probes used for the other representations. In this setting, SCimilarity contributes the expression features, and our training donors determine the task-specific mapping from those features to an outcome.

scVI

scVI is a variational model of single-cell gene counts that learns a compact latent representation while accounting for count variability [14]. We use the scVI cell representation prepared from the full astrocyte cohort and fit outcome-prediction probes using training donors only. The encoder therefore has access to test-donor expression during representation fitting, while the supervised probe remains separated by donor. Cell embeddings support the state and subtype tasks, and their donor means support the four donor outcomes. We also compare scVI with reduced-dimensional ASTRA representations to examine performance at comparable representation sizes.

Principal component analysis

PCA represents expression through orthogonal directions that explain successive amounts of variation. It provides a linear baseline for the learned representations. We use the full-cohort PCA representation with prediction probes fitted to training donors. Because expression PCs also enter the construction of the reference score, we summarize PCA in the eight-endpoint comparison covering subtype and donor outcomes. The PCA transformations used for donor pseudobulk, ASTRA dimension controls and visualization have separate roles and fitting procedures, which we specify in their respective analyses.

Masked-expression representation

We include the masked-expression representation as an additional baseline for learning features from expression. We use its cell embeddings prepared from the full cohort and fit the downstream probes using training donors. Its representation fitting therefore has the same full-cohort expression access described for scVI and PCA. We evaluate it alongside the other representations in the relevant baseline and visualization comparisons.

2.6 Evaluation tasks and training–test separation

We evaluate performance using five repeats of five-fold cross-validation, with fixed random seeds determining

donor partitions. In each repeat, we divide donors into five groups, train on four groups and test on the remaining group, rotating the test group to obtain five evaluations. Repeating this procedure gives 25 training–test splits. Within each split, all nuclei from a donor belong to either training or testing. This tests prediction for donors excluded from ASTRA training. It also keeps nuclei from the same individual together, so they cannot appear in both training and testing.

An encoder converts each expression profile into a compact numerical representation, also called an embedding. After fitting ASTRA on training donors, we hold its encoder fixed and fit a downstream prediction model, or probe, to the embeddings. The encoder learns the expression features; the probe uses those features to predict a particular outcome. Logistic regression predicts subtype membership, while ridge regression predicts continuous values with a penalty that limits coefficient size. We select ridge settings by cross-validation and fit predictor standardization to training data only.

Ridge regression fits a weighted sum of the representation features while penalizing large coefficients. This stabilizes prediction when features are numerous or correlated. Logistic regression uses a weighted feature combination to estimate the probability of subtype membership. Both are linear probes: they test information that can be recovered through a relatively simple mapping from the embedding. Nearest-neighbor probes use the outcomes of similar training observations, while MLP probes allow nonlinear combinations of the learned features.

Our nine endpoints test expression-state prediction, subtype discrimination and donor-level outcome prediction. Predicting the continuous reference score of each astrocyte contributes one endpoint. Predicting global pathology burden, amyloid load, tangle density and cognitive-decline slope from a donor's aggregated astrocyte representation contributes four more.

Subtype classification contributes the remaining four endpoints. We compare *CHI3L1* and *DPP10* astrocytes separately with *GRM3* astrocytes; these names identify the cohort's gene-associated subtype annotations. Each of the two contrasts receives two evaluations: pooled performance across donors and performance calculated within each donor. Pooled performance captures overall discrimination, while the within-donor evaluation asks whether the model distinguishes subtypes among cells from the same individual. We measure discrimination by the area under the receiver operating characteristic curve (AUROC), where 0.5 represents chance-level ordering and 1 represents perfect separation. Within-donor AUROC averages values for test donors with at least ten nuclei in each relevant class.

For reference-score and donor-outcome prediction, we use Spearman correlation. It measures how closely the predicted and observed values agree in their ordering. Higher positive values indicate stronger agreement. Donor-outcome probes use mean cell embeddings and donors with available measurements. We also evaluate nearest-neighbor probes, which predict from similar embeddings, and nonlinear multilayer perceptron (MLP) probes.

We use scVI, PCA and masked-expression representations computed from cells across the full cohort, including test donors. These representations have access to test-donor expression during fitting, a setting called partly transductive. Their outcome-prediction probes use training data only. ASTRA representation learning uses training donors only. For Geneformer, scGPT, Nicheformer and SCimilarity, we use pretrained embeddings without fine-tuning. We call these representations frozen because their encoder weights remain unchanged during evaluation. We fit their downstream prediction models using the training donors, as for the other representations. Thus, the comparison tests the predictive information in fixed pretrained features; fine-tuning would additionally update the encoder for the study task. We summarize eligible evaluated settings by their aggregate mean ranks; lower values indicate better performance. Ranking allows us to summarize performance across tasks that use different measures, such as correlation and AUROC. We report the endpoint-specific scores alongside this summary so that the overall ranking remains interpretable. We report nine-endpoint rankings separately from eight-endpoint rankings that include PCA. The same donors appear across the five repeats. The 25 evaluations therefore overlap and do not represent 25 independent samples. We also use benchmark observations during architecture development, as described below.

2.7 Expression maps and visualization

We construct the expression map in Figure 1e from the first repeat and first fold, with donor-partition seed 1000. We fit centered, non-whitened randomized PCA with 50 components, four power iterations and random seed 20260913 to 31,974 training nuclei. We then fit uniform manifold approximation and projection (UMAP) with 30 neighbors, minimum distance 0.3, Euclidean distance and the same random seed [20]. We use one worker for UMAP fitting and projection and project the 8,026 test nuclei through the fitted PCA and UMAP transformations. In Figure 3k, we compare eight representations using the same 8 000 nuclei. Each representation has its own two-dimensional map, and nuclei are colored by standardized reference score. We use these maps to visualize expression structure and the prediction tasks to quantify performance.

UMAP constructs a two-dimensional display that emphasizes local relationships among cells. The neighborhood setting controls how many nearby observations contribute to this structure, while minimum distance affects how tightly points can cluster in the display. We project test cells using the transformation fitted on training cells so that both groups share the same display coordinates. A UMAP axis is a visualization coordinate; the quantitative state and outcome predictions use the underlying representations.

2.8 Donor aggregation and control models

Pseudobulk and mean pooling

The pseudobulk baseline predicts outcomes from gene expression averaged within each donor. It provides a simple comparison that uses no learned cell embeddings. We

calculate mean log expression per donor, reduce it to 50 PCs using training donors and fit ridge regression. Mean pooling of ASTRA instead averages the learned cell embeddings, giving every sampled astrocyte equal weight.

Gated-attention model

We also evaluate gated attention, which learns weights for combining cell embeddings within each donor through the product of hyperbolic-tangent and sigmoid projections [21]. We combine the resulting donor vector with ten reference-score quantiles, three subtype fractions, and the score mean and standard deviation. We cap training samples at 400 cells per donor per epoch and standardize donor outcomes using training values. We evaluate each aggregation method on the same 25 splits.

Attention pooling treats a donor's cells as a collection with a shared donor outcome. It learns how strongly each cell contributes to the donor summary. Mean pooling assigns equal weight to every included cell, providing a direct comparison with this learned weighting. The additional score quantiles summarize the distribution of cell states within a donor, and subtype fractions summarize the composition of the sampled astrocytes. A quantile marks a specified position in an ordered distribution. The hyperbolic-tangent and sigmoid transformations provide bounded values whose product acts as a gate for the attention weights.

Permutation and partition controls

In control analyses, we shuffle subtype labels globally or within donors, or fit prediction probes to fixed embeddings using random cell partitions instead of donor partitions.

Global label shuffling breaks the relationship between expression and subtype labels across the cohort. Shuffling within donors retains each donor's subtype composition while breaking the pairing between an individual cell and its label. Comparing these controls helps identify whether a probe uses within-donor expression differences or differences in subtype composition between donors. Random cell partitions address a separate issue: cells from the same donor can then contribute to both training and testing, allowing donor-associated information to influence apparent predictive performance.

2.9 Target ranking and intervention direction

Within each cross-validation split, we use ridge regression to predict the reference score for held-out donors. Each cell receives one out-of-fold prediction per repeat: both the ASTRA encoder and the prediction probe exclude that cell's donor from training. We average these five predictions to obtain one ASTRA-derived score per cell. The reference score is the fixed quantity we predict. The ASTRA-derived score is our prediction from the learned embedding. We apply the same gene-ranking method to both scores to test whether ASTRA improves the resulting rankings. For gene ranking, we divide donors into five fixed, non-overlapping partitions. Within partition k, we center the score by subtracting each donor's mean, yielding $\tilde{r}_i$, and standardize gene expression across cells, yielding $z_{ig}^{(k)}$. For n_k cells in partition k, we calculate the gene–score association and its mean across partitions as

$$a_{gk} = n_k^{-1} \sum_{i \in k} \tilde{r}_i z_{ig}^{(k)}, \qquad \bar{a}_g = 5^{-1} \sum_k a_{gk}.$$

Subtracting each donor's mean emphasizes differences among that donor's cells. This prevents differences in average scores between donors from driving the gene associations. To adjust for expression level, we remove the linear relationship between $\bar{a}_g$ and gene abundance rank. We then multiply the remaining association by the fraction of donor partitions that agree in sign. This gives greater weight to genes with more consistent association directions. We rank genes by the magnitude of the adjusted score. The sign of the original mean association determines whether we nominate reduced or increased expression. A positive association means that expression tends to rise with the disease-aligned score, so we nominate suppression. A negative association means that expression tends to fall with the score, so we nominate induction. These proposed changes in expression aim to oppose the reference-associated pattern. We apply the same procedure to ASTRA-derived and reference scores across 15,051 genes and select four targets for detailed analysis by their absolute ranks among the 250 nominations per direction.

The five partitions used for gene associations serve a different role from the repeated training–test splits. Cross-validation produces predictions for donors excluded from model training. The association partitions then ask whether the relationship between a gene and those predictions has a consistent direction across groups of donors. Centering the score, adjusting for gene abundance and weighting sign agreement each address a distinct source of variation in that ranking.

2.10 Independent regional evaluation

We use the Seattle Alzheimer's Disease Brain Cell Atlas (SEA-AD) for independent evaluation [22, 23]. It groups astrocytes into broad populations called supertypes, based on similar expression patterns. We compare 14,420 common genes across six such populations and ten brain regions, giving 60 population–region comparisons. Regional Wald statistics summarize differential expression. Each statistic compares the estimated expression difference with its standard error. Its sign gives the direction of change, and its magnitude reflects the size of that change relative to uncertainty. We compare these statistics with gene-ranking scores using Spearman correlations and resample genes for paired bootstrap comparisons. For each regional point in Figure 6c, we average the six supertype correlations. We group regions by anatomical stage: early (MEC, LEC, HIP), intermediate (ITG, MTG, STG, AnG, FI) and late (DFC, V1C). We assess spatial variation through supertype-level contrasts and permutations of region groups. Rank correlation assesses agreement across the common gene set, whereas precision assesses support among the genes selected at a fixed ranking cutoff. We calculate precision gain relative to each method's own null comparison matched for gene abundance and detection. This background contains genes with similar expression

levels and detection frequencies. Comparing against it helps distinguish the ranking signal from effects of these gene properties. We adjust the sign of the external Wald statistic to match the proposed intervention direction. In this direction-signed replication score, positive values support the nomination and negative values indicate disagreement. For example, a suppression candidate gains support when the external expression change agrees with the disease-associated increase that motivated its nomination. This comparison evaluates expression evidence in an independent cohort.

Expression detection frequency is the fraction of cells in which a gene is observed. Abundance and detection can influence which genes appear strongly supported, so the matched background provides a more comparable reference than an unrestricted gene sample. Correlation evaluates ordering across the full common gene set, while precision summarizes support within the selected portion of a ranking. These measures describe different aspects of replication and are reported separately.

2.11 Literature coverage and target annotations

We assess literature coverage using predefined searches of gene symbols and aliases and classify genes with zero to two search hits as having limited literature support. For initially zero-hit genes, we expand the searches to include up to eight approved synonyms. The expanded classification contains 415 limited-literature and 85 supported genes among the 500 nominations. We assess search coverage separately from external expression replication: a low search count describes the literature we retrieve, not proof that a gene has never been studied.

2.12 Pathway and transcription-factor enrichment

Preranked gene-set enrichment

To interpret the ranked genes biologically, we test whether genes belonging to a known pathway or regulatory program concentrate toward one end of the ranking. We perform preranked gene-set enrichment analysis using signed ranking scores, gene-set sizes of 15–500, 1,000 permutations and a false discovery rate (FDR) below 0.05 [24, 25].

A gene set is a collection of genes assigned to a common biological process or regulatory relationship. Preranked enrichment asks whether members of a set are concentrated near one end of an ordered gene list. A running enrichment statistic increases when the ordered list encounters a member of the set and decreases for genes outside the set. Its strongest departure identifies the concentration of set members. Permutation comparisons place this concentration against a null expectation. The normalized enrichment score (NES) adjusts the enrichment statistic relative to its null distribution, supporting comparisons among gene sets [24, 25].

Leading-edge genes are the genes that contribute most directly to an enrichment signal. We combine terms when their leading-edge gene sets overlap strongly. Specifically, their shared genes must make up more than half of all genes in the two sets together, giving a Jaccard overlap above 0.5. We refer to these groups of overlapping terms as modules.

This grouping reduces repeated descriptions of the same underlying signal. For example, several pathway terms may be enriched because they contain the same small group of strongly ranked genes. Their leading-edge overlap identifies that shared basis. We therefore interpret the module as a summary of related enrichment results.

Transcription-factor gene sets

We use GO Biological Process 2023, Reactome 2022, ChEA 2022, ENCODE/ChEA consensus, TRRUST 2019 and transcription-factor perturbation expression sets [26, 27].

A transcription factor is a protein that helps regulate transcription of other genes. Transcription-factor enrichment asks whether genes associated with a particular factor occur together toward one end of our ranking. The evidence underlying these gene sets includes DNA-binding experiments, curated regulatory relationships and expression responses after a factor is perturbed. ChEA and ENCODE-associated sets draw on binding evidence, while TRRUST collects regulatory relationships from the literature. Perturbation-derived sets describe genes whose expression changes after altering a factor. These sources allow us to relate the ranked genes to candidate regulatory mechanisms [26, 27].

We apply the ranked enrichment procedure to these regulatory sets and summarize significant factors separately for the two intervention directions. Thus, an enriched factor is identified through the distribution of its associated genes in the ranking. We interpret it as a regulator implicated by the expression pattern, with the source gene-set evidence defining the type of regulatory link being tested.

We report the strongest significant term for each selected factor, with six factors per intervention direction, and examine gene-expression patterns and ligand–receptor co-expression for additional context.

GO terms describe biological processes, and Reactome sets describe curated pathways. Regulatory sets instead connect genes to factors that may organize their expression. Ligand–receptor co-expression adds another layer of context: a ligand is a signaling molecule, and a receptor is a protein that can receive its signal. Here, the analysis concerns expression of the corresponding genes within the astrocyte data. We use it to describe potentially relevant signaling components alongside the pathway and regulatory results.

2.13 Perturbation signatures and reversal scoring

We analyze 3,037 perturbation expression signatures from Gene Expression Omnibus (GEO) studies curated through CREEDS, using their cross-species gene mappings and perturbation directions [28]. A perturbation signature records which genes increase or decrease after an intervention. We measure how strongly each signature opposes the ranked disease-associated pattern. The reversal score is the difference between two normalized enrichment scores.

The two components summarize where the intervention's increased and decreased genes occur in the disease-associated ranking. An opposing response places genes in-

creased by the intervention toward the disease-decreased side, and genes decreased by the intervention toward the disease-increased side. We retain the scoring direction of the analysis when combining these components. A stronger reversal match therefore describes a closer opposition between two expression patterns: the disease-associated ranking and the observed response to a perturbation. Thus, target ranking selects genes to perturb, while signature reversal ranks observed perturbations by the expression changes they produce. We summarize experiments by their reversal scores and Benjamini–Hochberg (BH)-adjusted significance. For compounds represented by multiple experiments, the median summarizes the typical reversal score and the maximum identifies the strongest observed response. A weighted Stouffer combined z statistic combines standardized evidence across experiments into a compound-level summary.

The median and maximum answer different questions: the median describes a compound's typical score across its experiments, whereas the maximum describes its strongest match. The combined z statistic summarizes evidence on a common standardized scale. We interpret these summaries together with the number and biological context of the contributing experiments, since a compound may have been measured under different doses, durations or cell conditions. We perform 200 gene-label shuffles and retain the highest compound score from each shuffle to form a null comparison for the observed maximum. This tests whether the strongest observed match exceeds the strongest matches obtained after shuffling gene identities. We consider the species and neural or non-neural cell context of each experiment when interpreting individual compounds. We also annotate targets using disease associations from Open Targets, genetic constraint (depletion of observed genetic variation), tissue expression and cancer-cell dependency (effects of gene loss on growth or survival in those cell lines) [29, 30, 31, 32].

2.14 Published intervention evidence

We obtain reviewed human protein sequences and annotations for GFAP, PLCE1, RERG and WIF1 from UniProt, a database of protein sequences and function [33]. We use these annotations to identify each target's reference sequence, protein length, alternative names and cellular location. To identify primary studies, we search target names together with terms for expression modulation, RNA targeting, enzyme or receptor activity, protein supplementation and neurological disease. We organize evidence by intervention, experimental model and measured outcome. For each study, we compare the organism, protein variant, dose, treatment duration, controls, sample size and reported effect. We extract numerical results from study text, tables or deposited data and report them in their original units.

We retain assay context when comparing studies. Where a dose, unit or assay correction is unclear, we describe the result qualitatively rather than assign an uncertain quantitative value. We apply this approach to conflicting units in some GFAP experiments and dilution handling in WIF1 enzyme-linked immunosorbent assays (ELISAs). We exclude a retracted ALKBH5/WIF1 study. We check compound identities, including distinctions between salts and parent compounds, against PubChem [34].

2.15 Protein structure mapping and molecular contacts

We select up to two structures per target from the RCSB Protein Data Bank (PDB) and AlphaFold Protein Structure Database, prioritizing human experimental structures and using predictions for full-length context [35, 36]. PDB identifiers name experimentally determined structures; AlphaFold identifiers name predicted models. We select GFAP 6A9P/AF-P14136-F1, PLCE1 2C5L/AF-Q9P212-F1, RERG 2ATV/AF-Q96A58-F1 and WIF1 2YGN/2YGQ. We use SIFTS to map each structure's residues to the full-length protein sequence. SIFTS provides residue-level correspondences between PDB structures and UniProt sequences. This distinguishes the protein fragment used in an experiment from the residues actually resolved in its structure. We consider mutations, binding partners and sequence coverage when interpreting these structures and use the rat PLCE1 catalytic-core structure 6PMP for additional context. Structural coverage is the fraction of the full protein sequence represented by resolved residues. It identifies which portions of the protein the experimental coordinates can inform. For AlphaFold models, we summarize confidence through the predicted local distance difference test score (pLDDT), a per-residue score from 0 to 100 whose higher values indicate greater confidence in local structure. Coverage describes how much sequence a structure contains, while pLDDT describes confidence in predicted geometry.

A residue is one amino-acid position in the protein. Mapping resolved residues onto the reference sequence allows us to identify which domains or sequence segments are represented in an experimental structure. The asymmetric unit contains the coordinates deposited for the crystallographic model. We retain its chain and ligand assignments so that a contact can be attributed to the correct protein or binding partner. These definitions allow the structural views to be interpreted alongside each target's molecular function.

We measure ligand contacts as Euclidean distances between non-hydrogen atoms, using 4.0 and 4.5 Å cutoffs. We use model 1, the highest-occupancy alternate atom selected by the parser, and the protein chains deposited in the asymmetric unit; we exclude waters and crystal mates. A ligand is a molecule associated with a protein in the deposited structure; it may be a natural nucleotide or lipid. We distinguish contacts to the nominated target from contacts to its binding partners, including HRAS-bound GTP in 2C5L. We reproduce all 124 target–ligand contact comparisons with an independent distance calculation. We use RDKit to assign bond orders from PDB Chemical Component Dictionary entries and check ligand stereochemistry against their InChIKeys. Bond order distinguishes single, double and other chemical bonds; stereochemistry describes their three-dimensional arrangement. An InChIKey is a standardized identifier used to compare chemical identities. We render protein coordinates with

PyMOL and chemical structures with RDKit. Through these analyses, we describe deposited molecular geometry and use AlphaFold coloring to indicate prediction confidence.

2.16 Compound data and assay selection

We search ChEMBL release 37 by human protein accession to identify target-specific assay data [37]. A dataset qualifies for target-specific modeling if it contains at least 200 distinct compounds with usable structures and 20 chemical scaffold or medicinal-series groups. It must also measure expression changes in the relevant direction or use an activity assay that is appropriate for the target. We assess assays separately by species, protein construct, endpoint and concentration units and distinguish exact measurements from thresholded values. The datasets we retrieve for the four targets do not meet these criteria. We use ChEMBL data under CC BY-SA 3.0 [38].

A chemical scaffold describes a compound's central molecular framework. Grouping related structures into scaffolds or medicinal series helps distinguish many variants of one chemical family from broad coverage of chemical space. Assay selection also preserves the meaning of the response variable: a measurement of RNA abundance, a protein-activity assay and a cellular staining response answer different experimental questions.

2.17 Chemical-response model implementation

Compound representation and random forest

For a separate chemical-response analysis, we use the analyzed compound subset from the reactive-astrocyte screen of Clayton et al. [39]. The screen measures astrocyte reactivity through staining for guanylate-binding protein 2 (GBP2). We model the reduction in the normalized fraction of GBP2-positive astrocytes relative to the reactive-astrocyte vehicle control. A larger reduction indicates a stronger decrease in this assay readout. We examine cell-number ratios alongside it to identify responses accompanied by a change in cell abundance. We encode compound structures as 2,048-bit Morgan fingerprints with radius 2, numerical descriptors of local chemical substructures. To test transfer across chemical structures, we keep chemical-series groups separate across development, calibration and testing, giving 1,180, 393 and 393 compounds, respectively. We select a 200-tree random forest with maximum depth 12 and minimum leaf size 5 using development-fold mean absolute error. As comparators, we use the training median, the training mean and a Tanimoto-similarity-weighted average of five neighboring compounds. We then evaluate the fitted model on held-out chemical series (Fig. 9a).

A Morgan fingerprint encodes local neighborhoods of atoms as a fixed-length bit vector. Radius 2 includes neighborhoods extending two bonds from each atom, and shared fingerprint bits provide a measure of structural similarity. A random forest combines predictions from many decision trees, each of which partitions compounds according to their features. Averaging their predictions produces the response estimate. Maximum depth limits the complexity of an individual tree, and minimum leaf size limits how few development observations can support a terminal prediction.

Chemical-response baselines

The mean and median baselines assign the corresponding development-data response to every test compound. They quantify how much improvement a structure-based predictor provides over a constant estimate. The nearest-neighbor baseline uses five compounds with similar fingerprints and weights their measured responses by Tanimoto similarity. For binary fingerprints, Tanimoto similarity is the number of shared on-bits divided by the number of bits present in either fingerprint. Keeping chemical series separate across datasets tests prediction beyond close analogues from the same assigned series.

Calibration and prediction intervals

The development data guide model selection, the separate calibration data determine interval widths, and the test data assess performance on held-out chemical series. We apply split-conformal calibration using the maximum residual within each calibration group to estimate prediction intervals.

A residual is the difference between a measured response and its prediction. Taking the maximum absolute residual within a calibration group summarizes the largest error for that group. The calibration scores then determine a radius around a prediction at the chosen nominal coverage. This separates selecting the response model from estimating its uncertainty and from measuring its final test error. Nominal coverage is the target coverage used to set these intervals, rather than an observed test-set success rate. Mean absolute error measures the average absolute difference between predicted and observed responses; lower values indicate smaller errors. We obtain an interval radius of 0.58 at nominal 90% coverage and a test mean absolute error of 0.226. Because this error is comparable to the simple baselines, we use the model to characterize the screen rather than infer reliable responses for new compounds. We base our biological interpretation on measured compound responses and cell-number ratios.

2.18 Statistical analysis

We calculate mean performance intervals using 2,000 bootstrap resamples of the split estimates.

A bootstrap resample draws observations from the reported analysis units with replacement and recomputes the summary. The resulting distribution describes variability under that resampling design. For paired model comparisons, we use performance differences from the same splits so that both models face the same donor partitions. A signed-rank test evaluates these paired differences using their signs and ranked magnitudes. To compare models, we use differences on matched splits, bootstrap intervals, signed-rank tests and BH adjustment within the defined comparison family. We report confidence intervals and adjusted P values separately and pool the four donor-level outcomes across 25 splits for the overall donor-outcome comparison. Bracketed intervals in Results are 95% bootstrap confidence intervals. For a paired performance dif-

ference, an interval entirely above zero supports an advantage for the first method. An interval that includes zero does not clearly separate the two methods. BH adjustment accounts for testing multiple comparisons, and we denote adjusted significance by q where shown in the figures.

The false discovery rate is the expected fraction of false discoveries among the comparisons declared significant under the procedure's assumptions. It is a property of a collection of tests, rather than the probability that an individual target is correct. Confidence intervals describe uncertainty in a reported effect, while adjusted significance accounts for the family of comparisons being tested. We preserve both summaries because they answer different statistical questions. Because repeats share donors and we use the benchmark for architecture selection, we interpret these intervals as measures of variability within the evaluated design.

We present measurements from published experiments with the uncertainty reported by their authors. In particular, the WIF1 surface plasmon resonance (SPR) study reports K_d values with plus/minus estimates without specifying whether these are standard deviations or standard errors. We retain the deposited coordinate identifiers and residue mappings for structural analyses and use consistent gene sets, donor partitions and measurement units for quantitative comparisons.

3 Results

3.1 ASTRA learns a complementary three-block representation

We analyze a cohort of 40 000 ROSMAP astrocyte nuclei from 422 donors, including 24 580 *GRM3*-, 8 755 *DPP10*- and 6 665 *CHI3L1*-annotated cells; 371 donors have measurements of global pathology burden, amyloid load, tangle density and cognitive-decline slope [17, 3]. The first three donor outcomes quantify brain pathology, while cognitive-decline slope measures change in cognitive performance over time. We average each donor's astrocyte embeddings and test whether this summary can predict the four outcomes. We retain the cohort subtype annotations and use a fixed 3 000-gene encoder panel, then consider the broader set of expressed genes when ranking targets.

The reference score compares a nucleus's expression with disease and control text descriptions (Methods). Because this score is constructed from expression PCs, its prediction from PCA uses the same underlying representation ($\rho = 0.9590$). We therefore compare PCA on the other eight tasks. Figure 1 shows the architecture and analysis workflow. We concatenate 128 features from the variational block, 256 from the split block and 224 from the deeper split block into the cell representation. These 608 features are the output of ASTRA's deep representation learning. We use the resulting cell embedding for state and subtype prediction and average its features within donors for donor-outcome prediction. Donor embeddings enter reconstruction decoders only; donor labels also participate in adversarial training where enabled. To construct the expression map, we fit transforms on 31,974 cells from 337 training donors and apply them to 8,026 cells from 85 held-out donors. The resulting map displays training and test nuclei in the same coordinate system. Each point represents one nucleus. Its position summarizes many expression measurements, while the prediction tasks test which information the representation retains.

3.2 ASTRA achieves the best aggregate rank in the evaluated benchmark

Across five repeats of five-fold donor-disjoint evaluation, we find that ASTRA achieves higher mean performance on all nine endpoints than Geneformer, scGPT, Nicheformer and SCimilarity used as fixed feature extractors (Fig. 2). The aggregate benchmark ranking gives ASTRA a mean rank of 3.18, compared with 7.84 for scVI, 10.64 for Geneformer, 9.87 for Nicheformer, 13.24 for scGPT and 14.13 for SCimilarity; lower ranks indicate stronger performance. In the eight-endpoint ranking that includes PCA, ASTRA scores 3.45, compared with 4.39 for PCA and 7.70 for scVI. These results establish ASTRA's leading aggregate performance within the evaluated benchmark. The mean rank summarizes relative performance across the evaluated settings. The correlations and AUROCs for individual tasks show the size of the predictive differences. We show split-level results for all nine endpoints in Fig. 2j–r and describe the representation fitting and ranking procedures in Methods.

We obtain an axis correlation of 0.9252 [0.9214, 0.9288] for ASTRA, compared with 0.8382 [0.8303, 0.8457] for scVI, a paired difference of 0.0870 [0.0826, 0.0917] (Fig. 3). A higher Spearman correlation indicates closer agreement between predicted and observed ordering of astrocyte scores. Table 1 compares the four donor outcomes.

Table 1: Prediction of the four donor-level outcomes. Values are mean Spearman correlations across the 25 evaluation splits; higher values indicate better prediction.

Outcome	ASTRA	scVI
Global pathology burden	0.4472	0.3665
Amyloid load	0.4271	0.3633
Tangle density	0.4269	0.3536
Cognitive decline	0.4824	0.3965

scVI has a slightly higher within-donor *CHI3L1* mean (0.9062 versus 0.9058), and PCA leads both *DPP10* contrasts (0.9928 and 0.9919, versus 0.9866 and 0.9831). We compare ASTRA and scVI on matched evaluation splits and summarize the 25 overlapping donor-disjoint splits using intervals and adjusted tests.

To predict donor outcomes, we average ASTRA features within each donor. A pseudobulk baseline, based on 50 PCs of donor-mean expression, achieves correlations of 0.4356, 0.4136, 0.4262 and 0.4806 for global pathology burden, amyloid load, tangle density and cognitive decline, respectively (Fig. 3j). This baseline tests how well donor-level outcomes can be predicted by averaging expression directly, while ASTRA first learns features at the cell level. The pooled ASTRA-minus-pseudobulk contrast is +0.0069 [-0.0020, +0.0162], $P = 0.327$. Gated attention achieves pathology correlation 0.4300, compared with 0.4472 for mean-pooled ASTRA. These comparisons support predict-

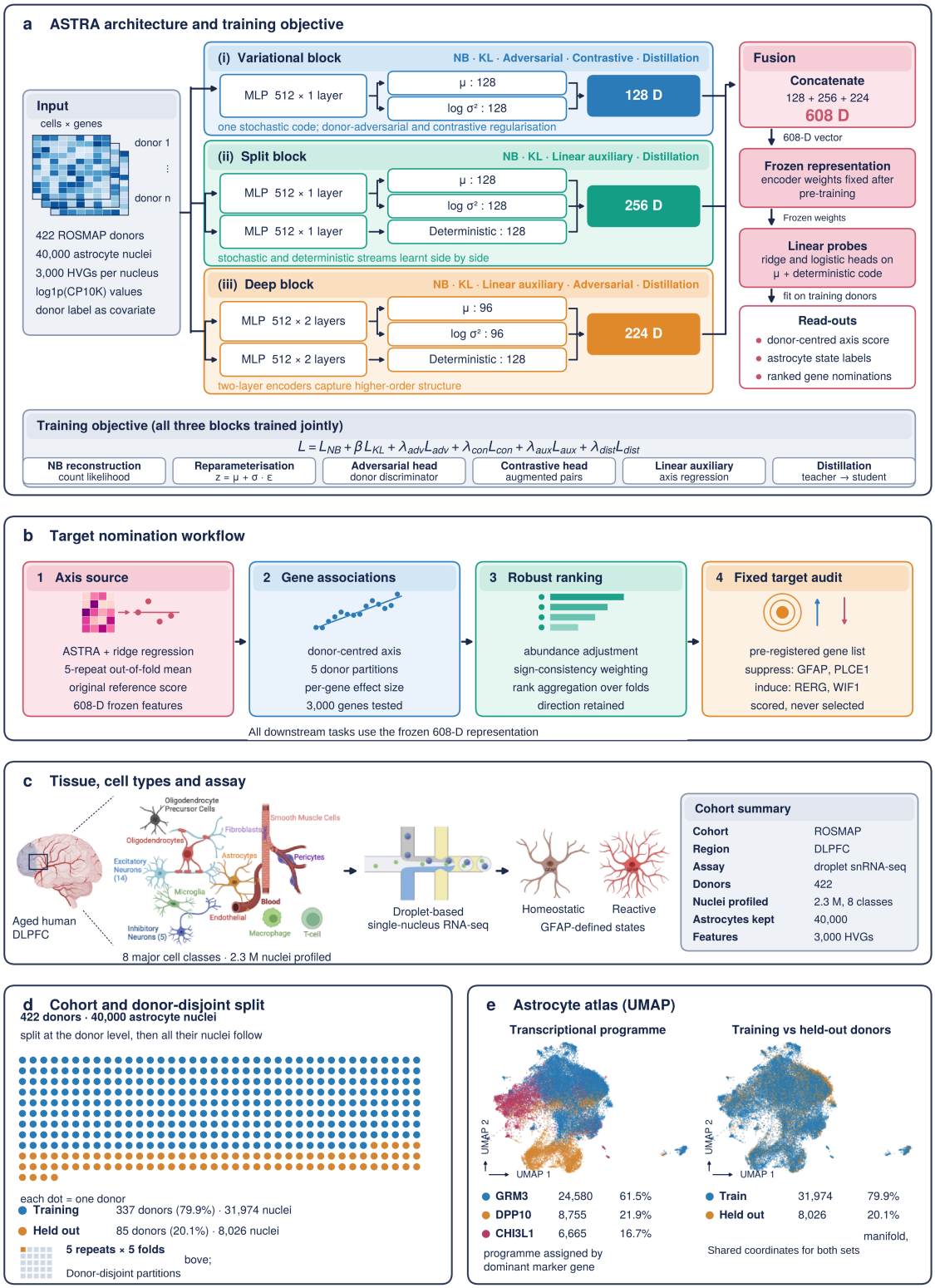


Figure 1: **Architecture, nomination workflow and cohort. a,** Three encoder blocks supply 128+256+224 features. We concatenate posterior means and deterministic codes and describe each block's reconstruction decoder and decoder-only donor embedding in Methods. **b,** We apply the same gene-ranking procedure to ASTRA out-of-fold predictions and reference scores. Overall ranks determine the targets selected for detailed analysis. **c,** Conceptual overview of astrocyte functions and other cell types. **d,** Cohort and partition sizes for repeat 1, fold 1. **e,** Expression map of all 40,000 nuclei, with identical coordinates in the subtype and donor-split views. We fit PCA/UMAP on training donors and project held-out donors through the fitted transforms.

ing donor outcomes by averaging the cell embeddings.

ASTRA's strong reference-score prediction is particularly relevant because we use that score for target discovery. The same representation also distinguishes astrocyte subtypes and predicts donor outcomes. The full endpoint comparisons identify the complementary strengths of ASTRA, scVI, PCA and pseudobulk (Figs. 2, 3).

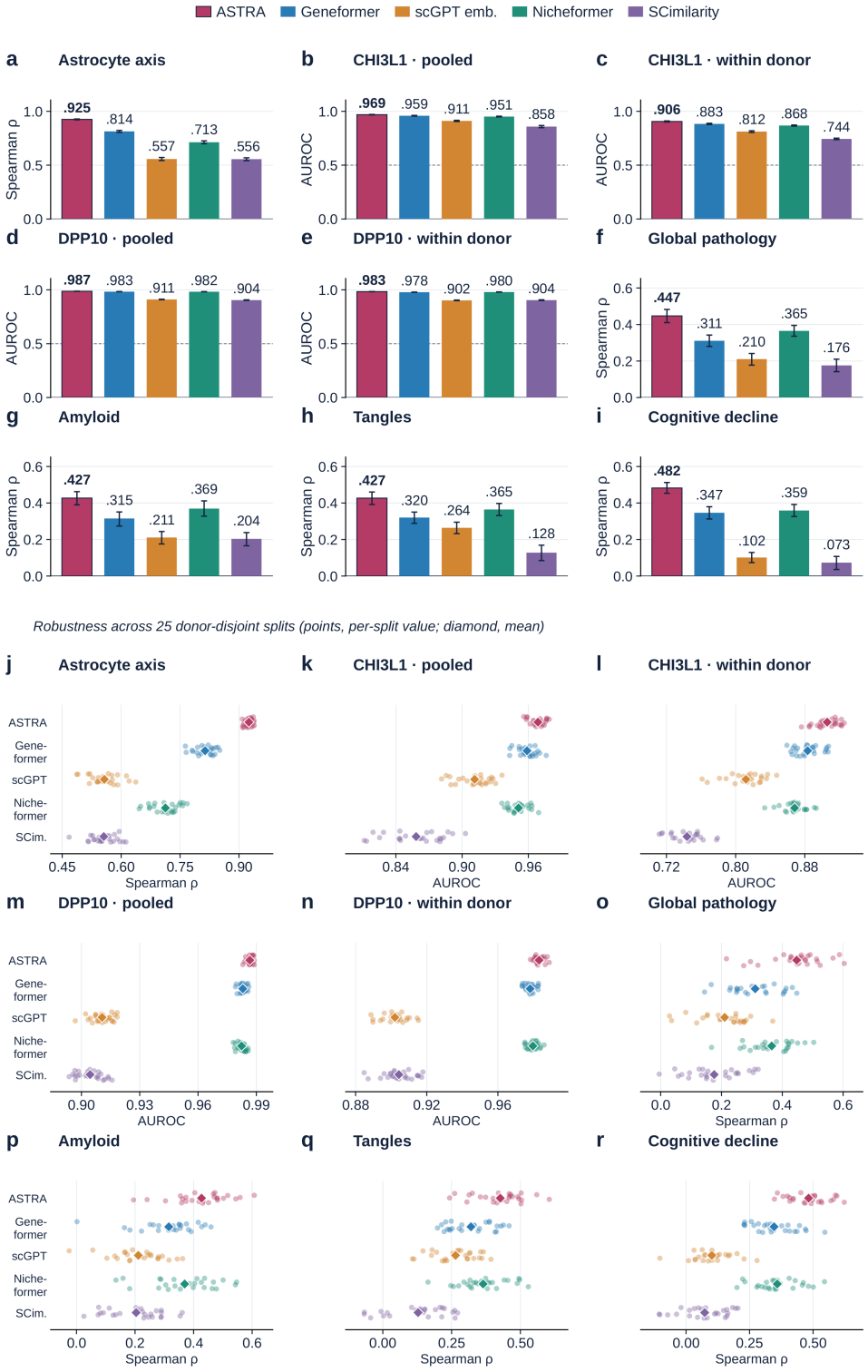


Figure 2: **Performance using pretrained models with fixed encoders. a–i,** Mean performance for ASTRA, Geneformer, scGPT, Nicheformer and SCimilarity on one reference-score task, four subtype evaluations and four donor-outcome tasks. Error bars show 95% bootstrap intervals over the 25 overlapping donor-disjoint splits. Dashed classification baselines indicate AUROC 0.5. **j–r,** Corresponding split-level distributions for the same nine endpoints; each point represents one evaluation split and diamonds indicate means. Colors identify the same representations in both panel groups. Higher correlations and AUROCs indicate better prediction. The four donor outcomes are global pathology burden, amyloid load, tangle density and cognitive decline.

3.3 Architecture analyses reveal complementary representations

After reducing ASTRA to 30 dimensions, we retain an axis correlation of 0.8850 and a pathology correlation of 0.4174 (Fig. 4b). This comparison with scVI suggests that the learned features contribute to performance even at matched representation sizes. Effects vary across block removals and changes to the training objective. We com-

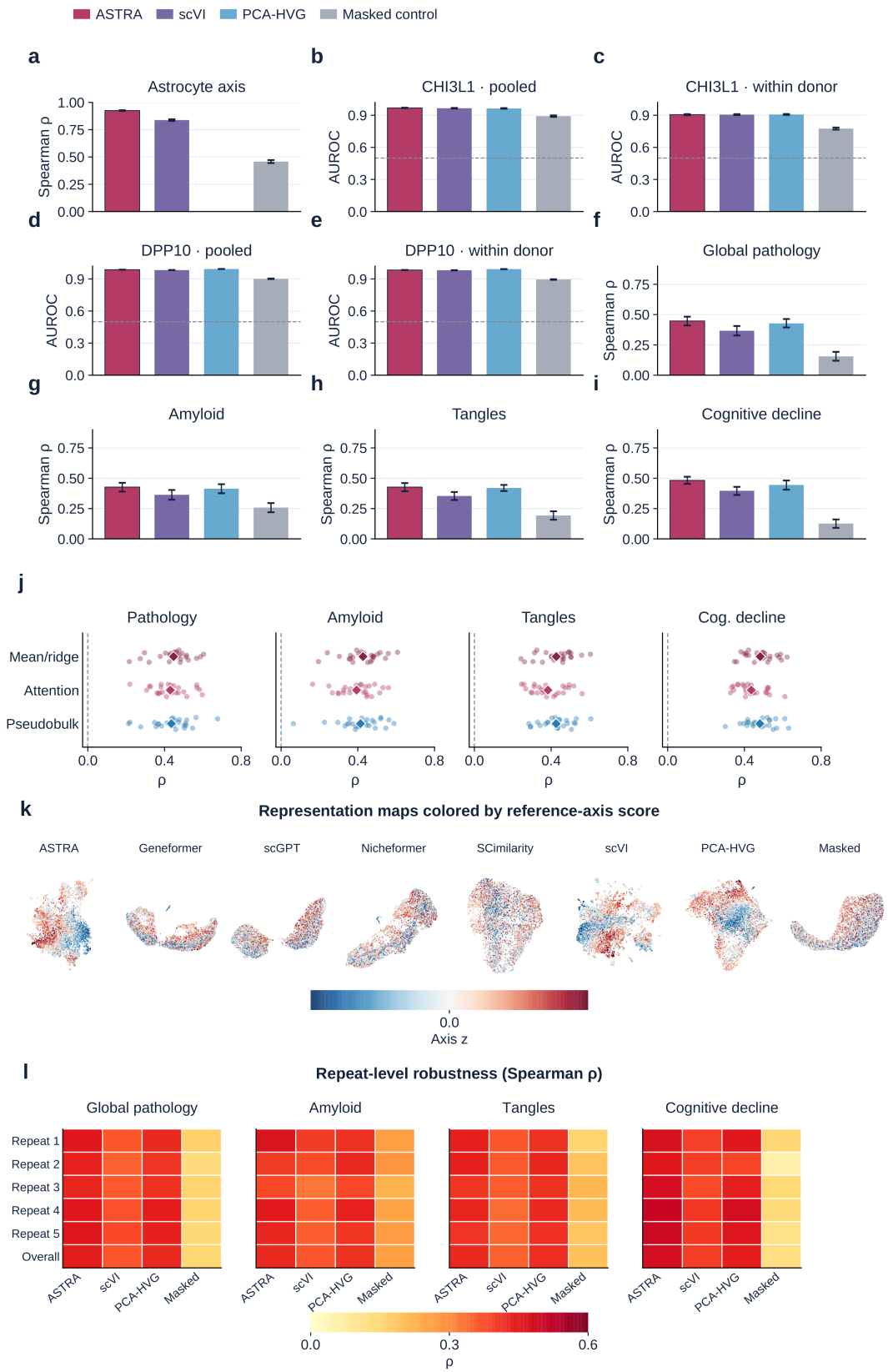


Figure 3: **Performance against task-matched baselines. a–i,** Mean performance and 95% bootstrap intervals for ASTRA, scVI, PCA applied to highly variable genes (PCA-HVG), and the masked-expression control. We evaluate PCA on the eight endpoints independent of the expression-PC reference axis. **j,** Correlations for the four donor outcomes across all 25 splits for mean-pooled ASTRA with ridge regression, attention-pooled ASTRA and 50-PC donor pseudobulk; diamonds indicate means. **k,** Descriptive two-dimensional views of the eight representations for a common sample of 8 000 nuclei, colored by standardized reference-axis score. Each representation has its own coordinates. **l,** Repeat-level and overall mean donor-outcome correlations for ASTRA, scVI, PCA-HVG and the masked-expression control. Repeated splits share donors, so we interpret them as overlapping evaluation estimates.

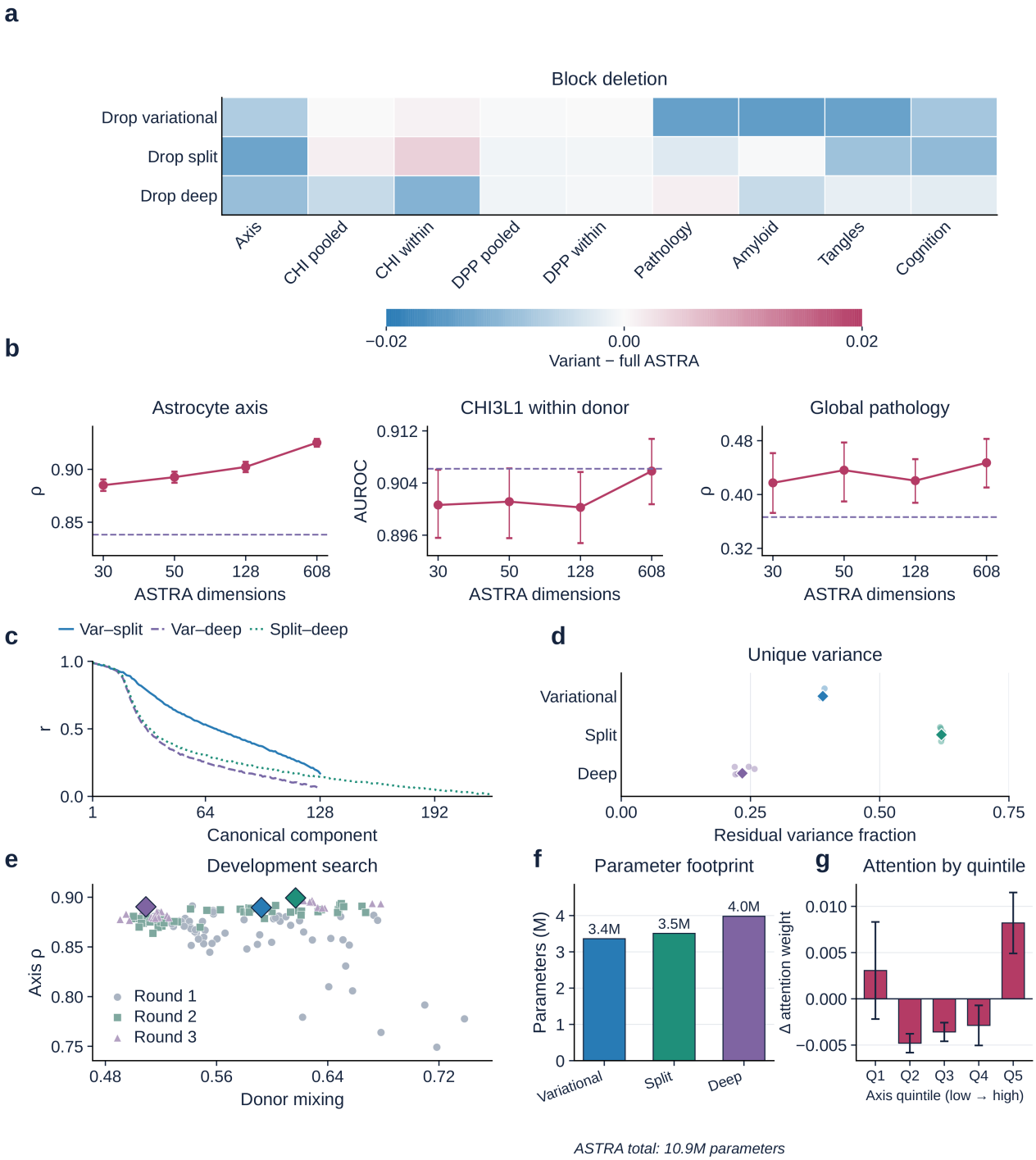


Figure 4: **Architecture development and complementary representation structure. a,** Change in performance after removing each encoder block, measured as variant minus full ASTRA across all nine endpoints. Negative values indicate lower performance after removal. **b,** Performance after projecting ASTRA to 30, 50 or 128 dimensions, compared with its full representation; error bars show 95% bootstrap intervals and dashed lines indicate scVI performance. **c,** Held-out canonical-correlation spectra for the three block pairs. **d,** Residual variance after regressing each block on the other two; larger fractions indicate less redundancy under this regression, and diamonds indicate means. **e,** Development configurations across three search rounds, with the selected blocks highlighted. **f,** Parameter counts of the three selected blocks and their combined total. **g,** Attention-weight differences across reference-axis quintiles. Panels summarize development analyses; architecture selection and benchmark evaluation share observations.

pare training variants with their parent configurations in 27 contrasts. Confidence intervals include zero in 18 comparisons, and the median absolute performance difference is 0.0022.

Using canonical correlations on held-out cells, we identify both shared and distinct information across encoder blocks. Mean variational–split correlation is 0.565, while split–deep correlation is 0.264 (Fig. 4c). Regressing each block on the other two leaves variance fractions 0.390, 0.619 and 0.234 for variational, split and deep blocks (Fig. 4d).

Across our 144-configuration search, we use different task combinations in successive development rounds, including donor mixing in the first round. We also use benchmark results to select the combined representation (Fig. 4e), so architecture selection and evaluation share the same benchmark, as detailed in Methods.

We use control analyses to test how donor composition affects performance and why training and test donors should remain separate. Shuffling subtype labels within a donor preserves that donor's subtype counts while breaking the link between an individual cell's expression and its label. Under this control, ASTRA's pooled *CHI3L1* AUROC of 0.8351 remains informative across donors, whereas within-donor AUROC falls to 0.5007. Using random cell partitions instead of donor partitions raises reference-score correlation from 0.9252 to 0.9297 for ASTRA and from 0.8382 to 0.8472 for scVI. Through these controls, we quantify the influence of donor composition and partition choice.

We examine the architecture through block deletion, canonical correlation and residual variance (Fig. 4). The variational block contributes strongly to axis and donor-outcome performance. The other two blocks explain least of the split block's variation: its residual-variance fraction is 0.619. The corresponding fractions are 0.390 for the variational block and 0.234 for the deep block. Thus, the blocks capture partly distinct information, and their removal affects individual tasks differently. Together, these development analyses support our choice to combine the three blocks into ASTRA's shared representation.

3.4 ASTRA identifies reversal candidates with strong regional concordance

We next identify genes whose expression changes with the disease-aligned score and decide whether to nominate reduced or increased expression. We average five out-of-fold predictions per cell and apply the same gene-association procedure to ASTRA-derived and reference scores across 15 051 genes. Within each donor group, we subtract each donor's mean score and standardize gene expression across cells. We average the gene–score associations and adjust them for gene abundance. We then give greater weight to associations that agree in direction across donor groups. The strength of the adjusted association determines the gene's rank. The original association direction determines whether we nominate suppression or induction. Here, suppression means reducing expression of a positively associated gene, while induction means increasing expression of a negatively associated gene. Each nomination therefore identifies a gene and whether to reduce or increase its expression. The association has the same sign across all five donor partitions for 5 552 genes with ASTRA and 4 692 with the reference (Fig. 5a,b). We select the leading pairs for detailed analysis: *GFAP*/*PLCE1* for suppression and *RERG*/*WIF1* for induction, with overall ranks of 1, 2, 10 and 12.

We evaluated the target rankings across six broad astrocyte populations, termed supertypes in the reference atlas, and ten brain regions, yielding 60 population–region comparisons. We use 14 420 common genes and regional Wald statistics from SEA-AD [22, 23]. Each comparison measures agreement between a gene ranking and external expression changes in one population and region. ASTRA exceeds the reference in 57 of 60 comparisons, with 51 bootstrap intervals excluding zero and median difference 0.0118 (Fig. 6a,b). It exceeds the case–control ranking in 60 of 60 comparisons. The labels early, intermediate and late refer to anatomical region groups in the staging scheme used for this analysis. The supertype-level contrast between early and late anatomical region groups is 0.0385 [0.0188, 0.0566], $P = 0.01562$, but the distinct region-level tier permutation is nonsignificant ($P = 0.5332$). These comparisons test how well the rankings transfer across brain regions and astrocyte populations.

SEA-AD provides an external test across independently collected donors, regional samples and processing procedures [22, 23]. The rank agreement supports similar expression relationships across the two datasets. Comparing six populations across ten regions shows where this agreement is stronger or weaker (Fig. 6).

We also assess support for the genes selected at a fixed cutoff. This complements the correlation across the full shared gene set. Median model-minus-reference precision differences are -0.010 for suppression ($P = 0.945$) and -0.010 for induction ($P = 0.967$). We compare each ranking's precision with a background matched for gene expression level and detection frequency (Fig. 6d). Gene-level comparisons also identify differences between the discovery and regional data (Fig. 5d). For *RERG*, the direction-signed external statistic is -0.199, indicating disagreement with the nomination direction. After expanding our searches to include gene aliases, we classify the nominations into groups of 415 limited-literature and 85 literature-supported nominations. These labels describe search coverage. We compare the external replication scores of these groups using a rank-biserial effect size. This measures how strongly their ranked values differ. The pooled effect is +0.0017 [-0.1294, +0.1391], $P = 0.981$ (Fig. 5c). These comparisons show how external expression support varies with the amount of retrieved literature.

3.5 Regulatory and chemical-response analyses connect nominations to mechanisms

To determine which biological processes organize the nominated genes, we test enrichment of established pathway and regulatory gene sets. Enrichment asks whether genes linked to the same process or regulator are concentrated toward one end of our ranking. A transcription factor regulates the expression of other genes; its enrichment is assessed through those associated genes (Methods, Section 2.12). Grouping overlapping enriched gene sets yields 21 modules for ASTRA and 12 for the reference. These include inflammation and cytokine responses, signaling, protein translation and metabolism (Fig. 7a). The 12 factors we display include STAT3, EGR1, NFIC, RELB, ATF6 and SMAD3 on the suppression side (Fig. 7b). These enrichments connect the ranked genes to established transcription-factor and pathway signatures. Gene-expression patterns and ligand–receptor co-expression provide additional context.

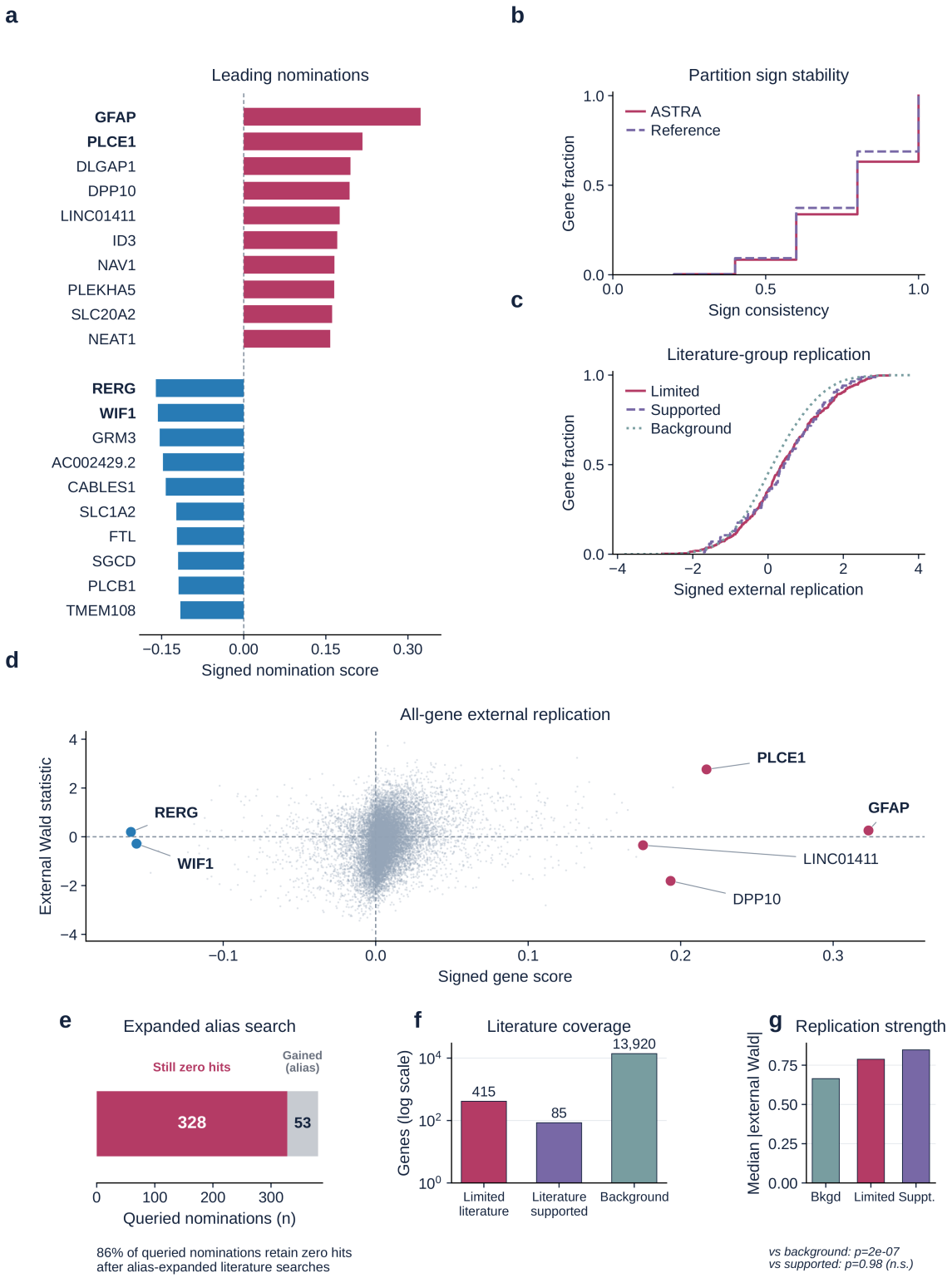


Figure 5: **Candidate reversal targets and external expression evidence.** **a,** Ten leading nominations per direction, ordered by the overall rank; the four targets examined in detail are highlighted. Scores describe how gene expression relates to the state score. Positive scores nominate suppression; negative scores nominate induction. **b,** Sign-consistency distributions across five donor partitions for 15,051 genes per axis. **c,** Direction-signed external Wald distributions for limited-literature, literature-supported and background genes; positive values indicate agreement with the nominated direction. **d,** Raw external Wald statistics against ASTRA nomination scores for all 14,420 common genes. **e,** Expanded alias searches for previously zero-hit nominations: 328 retain zero retrieved hits and 53 gain a hit. **f,** Sizes of the expanded literature-coverage groups: 415 limited-literature, 85 supported and 13,920 background genes. **g,** Median absolute external Wald statistics for these groups, with the associated comparisons. We define literature coverage through the gene-symbol and alias searches described in Methods.

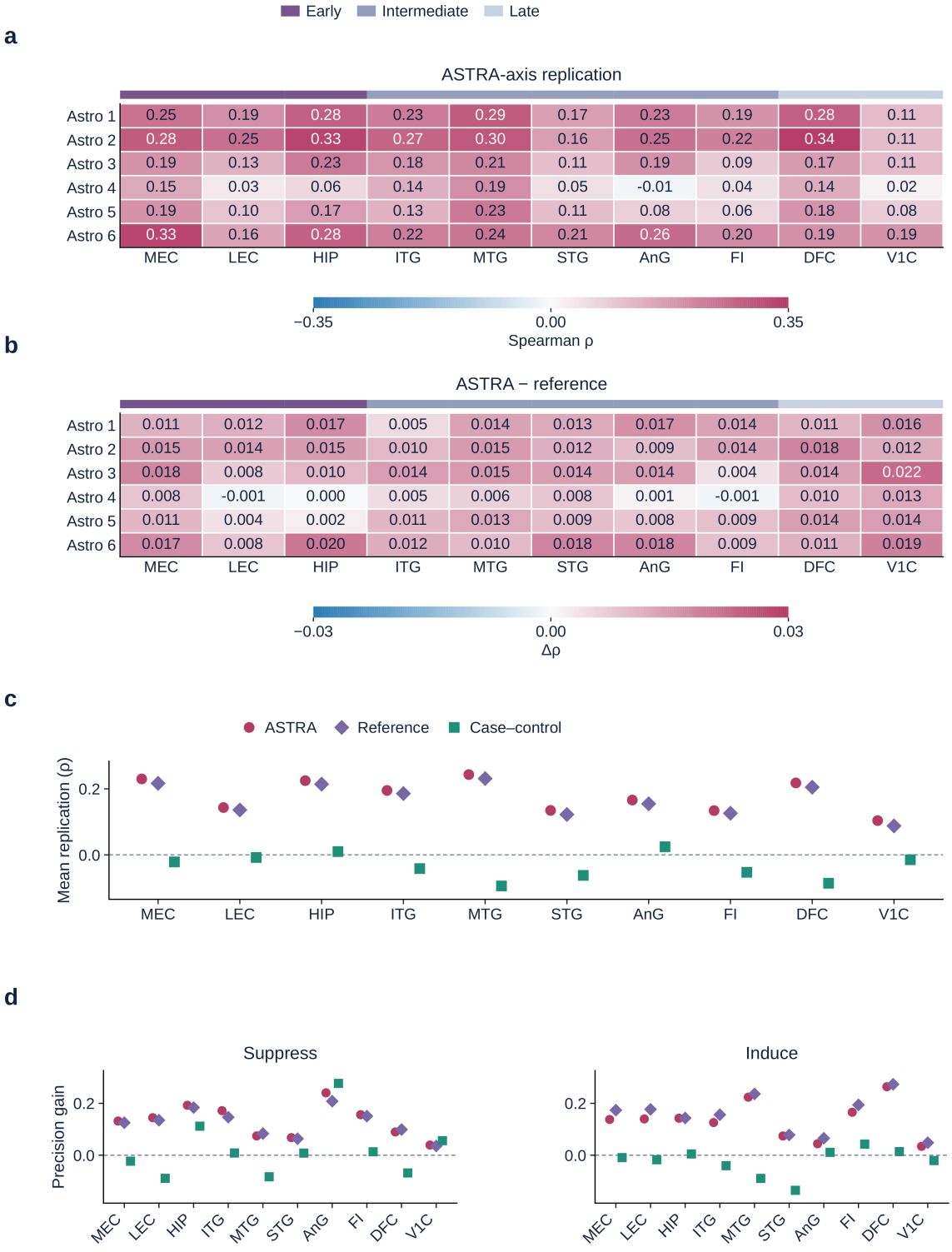


Figure 6: **Regional agreement and matched null comparisons. a,** ASTRA score–Wald correlation across all six supertypes and ten regions. **b,** ASTRA-minus-reference correlation differences, across the same comparisons. Positive differences favor ASTRA. Color strips identify the early, intermediate and late region groups. **c,** Mean correlation across the six astrocyte populations in each region, for ASTRA, reference and case–control rankings. We report the staging analysis for individual populations separately. **d,** Precision above a background matched for gene expression level and detection frequency, shown separately for suppression and induction. Each ranking uses its own matched background. The supertype early-minus-late test and region-level tier permutation ($P = 0.5332$) assess distinct forms of spatial variation. We describe aggregation and uncertainty estimates in Methods.

We then compare the ranked program with published perturbation signatures, lists of genes that increase or decrease after a chemical or genetic intervention. A strong reversal match identifies an intervention whose measured expression changes oppose the nominated disease-associated pattern. Our GEO/CREEDS-derived screen covers 3 037 experiments, 270 compounds and 991 gene perturbations [28]. We show the complete compound ranking and 12

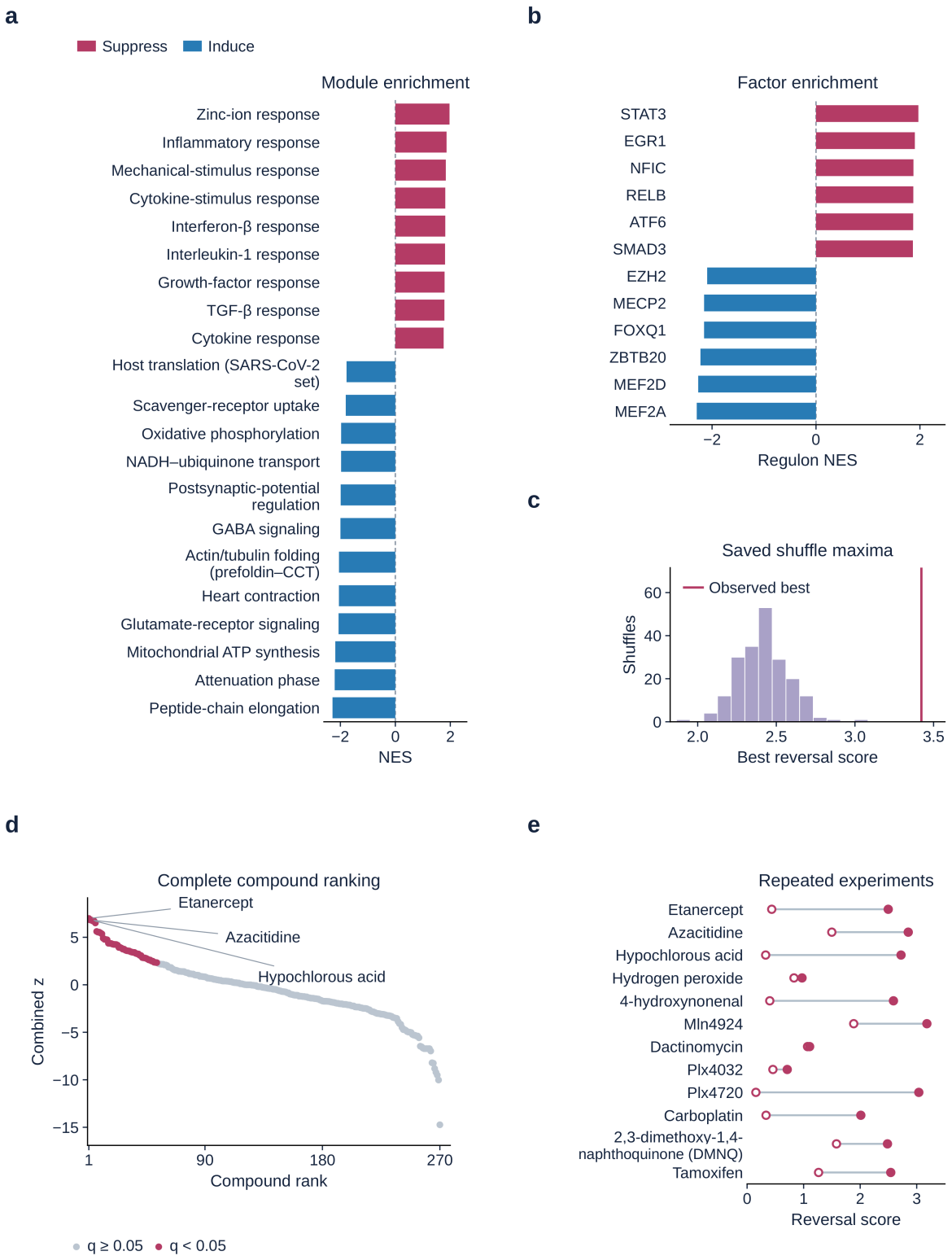


Figure 7: **Regulatory enrichment and perturbation signatures.** **a,** The 21 groups of overlapping enriched terms (modules), with signed normalized enrichment scores (NES) and counts of genes driving each signal. **b,** Strongest significant term for each of six selected factors per direction. Labels identify the source gene sets used for enrichment. **c,** The 200 gene-label shuffle maxima and observed best reversal; the observed maximum exceeds all 200 shuffle maxima. **d,** The complete 270-compound ranking, with experiment-level q flags indicating significance after multiple-testing adjustment. **e,** The top 12 compounds with at least two experiments by combined z. Open/filled points mark median/maximum reversal. The GEO/CREEDS-derived signatures span species and cellular contexts [28]; we consider the context of each experiment when interpreting the results.

multi-experiment summaries in Fig. 7d,e. The highest observed reversal score, 3.4226, exceeds the highest scores obtained in each of 200 gene-label shuffles (Fig. 7c). Combined z summarizes evidence across experiments, while median and maximum reversal describe typical and strongest observed responses. The source experiments span species

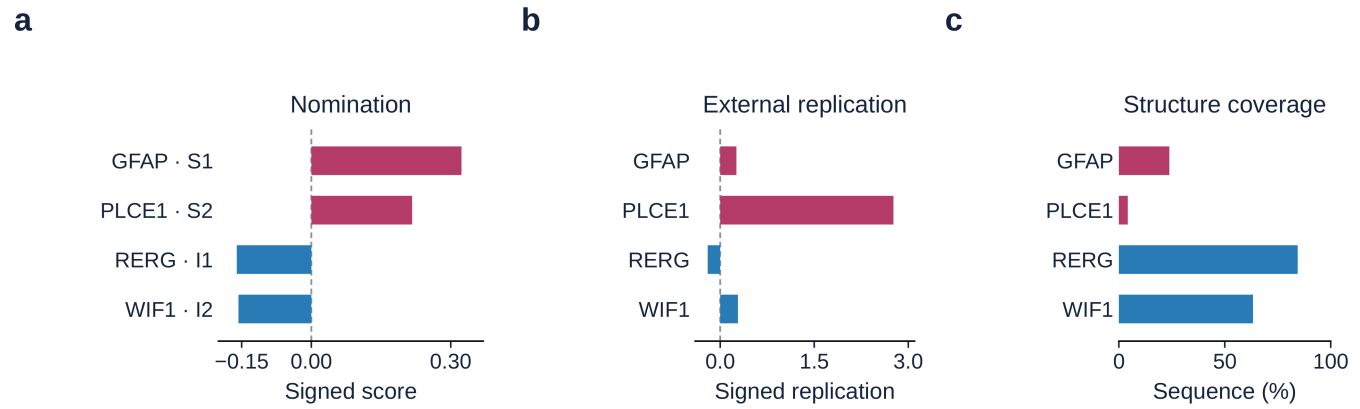


Figure 8: **Comparative evidence for four leading targets. a,** Signed nomination scores; S1/S2 and I1/I2 are within-direction ranks. **b,** Direction-signed external Wald statistics for the four targets. **c,** Largest fraction of the full-length protein resolved in the selected experimental structures: 6A9P, 2C5L, 2ATV and 2YGQ. Coverage counts resolved residues mapped to the full-length reference sequence; WIF1's larger model is low resolution [35]. Detailed study controls and sources appear in Figures 9 and 10 and the corresponding target sections.

and cellular contexts. To connect the screen to our four-target analysis, we examine estradiol, decitabine, phenytoin and *Gfap*-knockdown signatures at the individual-experiment level (Fig. 9c).

By combining pathway enrichment with chemical-response analysis, we connect candidate reversal programs to specific regulatory factors and perturbations. The enrichment results describe the biological organization of the ranked genes, while signature reversal identifies interventions associated with opposing expression changes. We use these results to select experiments that test how the nominated targets affect astrocyte state.

3.6 Integrated evidence identifies intervention routes for four leading targets

For each target, we examine practical ways to reduce or increase its expression or alter its activity. Figure 8 compares the four selected targets. Figure 9 brings together chemical-response modeling, structures and expression-signature comparisons. Figure 10 summarizes published experiments for each target. We also examine a broader set of 50 targets using disease associations, genetic constraint, tissue expression and cancer-cell dependency. UniProt annotations identify GFAP and RERG as intracellular proteins, PLCE1 as a cytoplasmic protein with peripheral membrane association, and WIF1 as a secreted protein [33]. Cellular location helps determine how a target can be approached experimentally. Lowering RNA or increasing gene expression changes how much protein a cell produces. Enzyme modulation changes its catalytic activity, while supplementation supplies an extracellular protein. We use these distinct routes to organize the evidence for the four targets.

In our four-target analysis, we bring together 57 intervention–outcome observations from published studies and eight protein structures: five experimentally determined domains or complexes and three full-length AlphaFold models [35, 36]. We map each structure to the full-length protein to identify resolved regions and compare the experimental constructs. Bound GDP, GTP and WIF-domain phospholipid provide information about molecular interactions alongside the intervention evidence.

The chemical analyses address two complementary questions. The GEO/CREEDS comparison asks which published perturbations oppose the nominated gene-expression pattern. The separate reactive-astrocyte screen measures a cellular response to compounds through GBP2 staining. We model that measured response to assess how much information chemical structure provides about the screen outcome (Fig. 9a). Morgan fingerprints summarize local chemical substructures, and a random forest combines predictions from multiple decision trees. We encode compound structures with these fingerprints and train a random forest on 1,180 development compounds to model the normalized GBP2-positive fraction reduction. The calibration and test sets contain 393 and 393 compounds, with chemical series kept separate between sets. Held-out mean absolute error is 0.226, with a split-conformal interval radius of 0.58 at nominal 90% coverage. We find that the random forest has error comparable to median, mean and nearest-neighbor baselines and therefore use it descriptively. We base our biological interpretation on the published compound and target measurements summarized in Figure 10a–c.

3.7 GFAP: RNA targeting and chemical modulation support expression suppression

GFAP encodes glial fibrillary acidic protein, a component of the astrocyte intermediate-filament network. To examine routes for its suppression, we review published experiments that lower GFAP RNA or protein. Clomipramine lowered mouse GFAP promoter activity by 26% and protein by 57% after 21 days of 40 mg/kg intraperitoneal dosing every other day (16 treated, 15 vehicle) [40]. The separate reactive-astrocyte phenotypic screen of Clayton et al. [39] reports normalized GBP2-positive fraction reductions of 98.9% for JSH-23 and 94.3% for RGFP966, with cell-number ratios shown alongside the GBP2 measurements

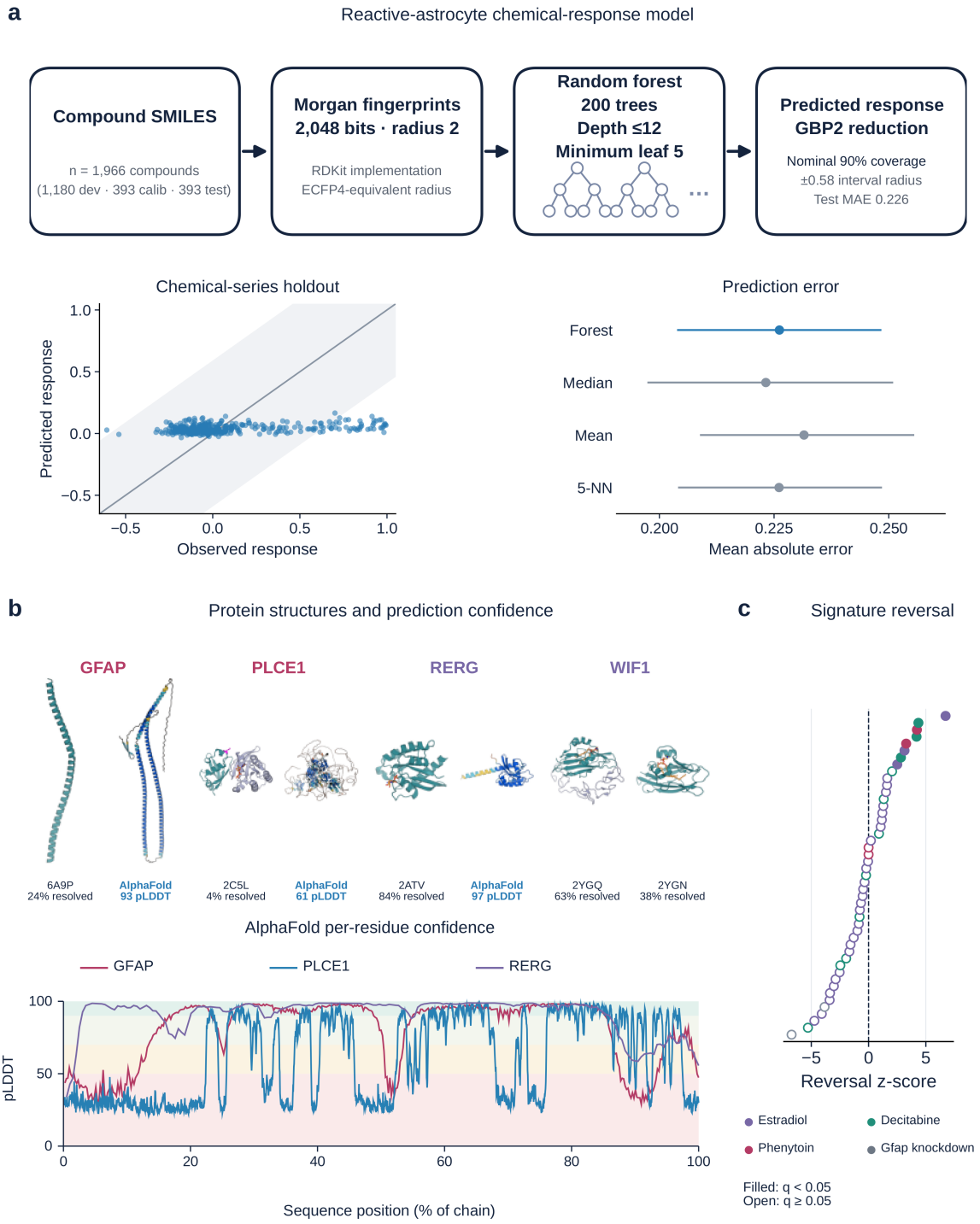


Figure 9: **Chemical-response modeling, protein structures and signature reversal. a,** Reactive-astrocyte chemical-response model. We encode compound structures from SMILES notation as Morgan fingerprints (2,048 bits, radius 2). A 200-tree random forest (depth $\leq$12, minimum leaf 5) uses 1,180 development compounds. Calibration and testing use 393 and 393 compounds, with chemical series kept separate across sets. Plots compare observed and predicted GBP2 responses and show mean absolute error with 95% intervals. Baselines use the median, mean and Tanimoto-weighted 5-nearest-neighbor predictions. Screen measurements derive from Clayton et al. [39]. **b,** Structural coverage for the four targets. Each pair shows the experimental fragment with greatest sequence coverage beside a full-length AlphaFold model (GFAP, PLCE1, RERG) or a second experimental structure (WIF1). Annotations give the percentage of resolved residues and median pLDDT; traces show confidence along the predicted sequence [35, 36]. **c,** Individual expression-signature comparisons for estradiol, decitabine, phenytoin and *Gfap* knockdown against the ASTRA axis. Filled points indicate $q < 0.05$ and open points indicate $q \geq 0.05$ [28].

(Fig. 10a). The published clomipramine experiments delayed aggregate accumulation; daily dosing at the same level caused weight loss in a pilot. Later authors reported an unsuccessful clomipramine approach in mutant-GFAP mice [45]. The expression-lowering effects therefore depend on the experimental model and dosing regimen.

Antisense oligonucleotides (ASOs), short nucleic-acid sequences designed to bind a target RNA, offer a di-



Measured compound evidence

Target	Compound	GBP2 ↓ (%)	Cell ratio	Target evidence
GFAP	JSH-23 (probe)	98.9	0.98	RNA −52%
	Clomipramine (Rx)	11.9	0.98	Protein −57%; reporter mouse
	Phenytoin (Rx)	6.7	1.00	RNA/protein ↓; U251-MG
	RGFP966 (probe)	94.3	0.79	RNA: q ≥ 0.05
PLCE1	RGFP966 (probe)	94.3	0.79	RNA −34%
	JSH-23 (probe)	98.9	0.98	RNA: q ≥ 0.05
	Zinc pyrithione (activity)	—	—	ε IC50 1.2/7.5 µM
	Sanguinarine (activity)	—	—	ε/γ1 IC50 1.9/1.6 µM

Figure 10: **Target-specific intervention evidence for the four nominated genes. a,** GFAP and PLCE1 chemical-genetic evidence. Transcript panels show $\log_2$ fold changes for reactive versus physiological astrocytes and for compound-treated versus vehicle-treated reactive astrocytes, with FDR-adjusted significance. A $\log_2$ fold change of $+1$ denotes a doubling and -1 a halving relative to the stated comparison. The table combines normalized GBP2 reductions and cell-number ratios from the astrocyte screen with target-specific evidence from the indicated primary studies [39, 40, 41, 42]. **b,** RERG: estradiol induction and ESR2/RERG-knockdown $\log_2$ fold changes with reported P values in primary endometriotic stromal cells [43]. **c,** WIF1: reported SPR K_d against Wnt3a across domain-truncation constructs, with reported uncertainty; lower K_d indicates tighter binding [44]. We discuss published thermal-stability, zinc-delivery, promoter-methylation and cell-cycle results in the corresponding target sections.

rect way to lower the transcript. Lead ASOs achieved $>$ 90% *Gfap* mRNA reduction in wild-type mice, with improvements in independently measured pathology and function in Alexander-disease mouse/rat studies [45, 46]. These studies provide evidence that lowering GFAP can affect tissue pathology and function in Alexander-disease models, where pathogenic GFAP variants drive disease. They motivate a target-lowering strategy whose effects

in AD-associated astrocyte states can be tested separately. Studies in human U251-MG astrocytoma cells also report GFAP lowering with clomipramine or phenytoin. Clomipramine or carbamazepine also changed the organization of mutant GFAP filaments. We describe these results qualitatively because the published doses or units conflict [41].

The experimentally determined GFAP 1B-domain structure covers 103 of the full protein's 432 residues. We compare it with an AlphaFold model whose median pLDDT indicates confident prediction (Fig. 9b) [35, 36]. The thermal unfolding midpoint of purified human GFAP was 51.8 °C alone, 58.2 °C with ceftriaxone and 55.3 °C with phenytoin sodium [47]. The thermal midpoint describes the temperature of the protein-unfolding transition. These measurements characterize protein stability and complement the transcript and cellular readouts.

From these published findings, we identify several experimentally grounded routes for GFAP intervention. Antisense oligonucleotides provide direct transcript lowering, small molecules provide published cellular expression effects, and biophysical measurements characterize protein stability. On this basis, we prioritize selective GFAP lowering for future experiments with controls that restore GFAP expression and independently assess astrocyte function.

3.8 PLCE1: enzyme inhibition and astrocyte genetics identify a signaling target

PLCE1 encodes phospholipase C epsilon (PLCε), an enzyme involved in intracellular signaling. Its catalytic activity provides a measurable endpoint for chemical modulation. Zinc pyrithione (ZPT) inhibits purified human PLCε with $IC_{50} = 1.2\,\mu M$ using a fluorogenic substrate, versus $> 50\,\mu M$ for the tested PLCβ4/δ1/γ1 isoforms; the biotin-PIP_2 assay gives $7.5\,\mu M$ (see also Fig. 10a). Sanguinarine and chelerythrine also inhibit PLCε but similarly inhibit PLCγ1, limiting selectivity claims [42]. Here, IC_{50} is the compound concentration that produces half-maximal inhibition in the stated assay. Comparing it across enzyme isoforms helps assess selectivity, while differences between substrates or assay formats can change the measured value.

Zinc sulfate itself inhibits the enzyme, while pyrithione promotes cellular zinc delivery. In Caco2 cells, $5\,\mu M$ zinc sulfate plus equimolar pyrithione and $5\,\mu M$ ZPT produced 8.78- and 9.13-fold labile-zinc signals, respectively. The cellular assay measures labile zinc, including weakly bound zinc [42]. Experiments in primary mouse astrocytes separately tested the role of PLCε catalytic activity. After stimulation with thrombin or lysophosphatidic acid (LPA), the response involves protein kinase D (PKD), nuclear factor κB (NF-κB) and cyclooxygenase-2 (COX-2). Restoring the wild-type enzyme after gene deletion restores this response; restoring a catalytically inactive version does not [48]. Deletion removes the gene, and complementation restores it with either an active or inactive version. This comparison tests whether the inflammatory response depends on the enzyme's catalytic function. The biochemical and genetic experiments provide separate evidence for chemical modulation and astrocyte pathway function.

The human 2C5L structure we select contains the Y2176L RA2 fragment bound to HRAS, with GTP contacting HRAS. The full-length prediction has 52.4% of residues below pLDDT 70 (Fig. 9b) [35, 36]. The rat catalytic-core structure 6PMP adds an experimental view of the enzyme core [49]. Future experiments can test whether a compound acts on the enzyme and whether its effects on astrocytes depend on PLCE1.

For PLCE1, we connect chemical modulation of the enzyme with published genetic evidence for its role in astrocyte signaling. Purified-enzyme inhibition, cellular zinc delivery and astrocyte deletion/complementation experiments define three complementary components of this hypothesis. A combined chemical-genetic experiment can test their connection in a single astrocyte system.

3.9 RERG: estrogen signaling provides an expression-induction mechanism

For RERG, the nominated intervention is increased expression. We examine published evidence that estrogen signaling can increase RERG expression, alongside the independent expression data used to assess the nomination. In primary human endometriotic stromal cells, 100 nM estradiol increased RERG mRNA 1.55-fold at two hours and protein 1.54-fold at six hours (Fig. 10b) [43]. Knockdown reduces expression experimentally. Reducing ESR2, which encodes estrogen receptor β (ERβ), lowered RERG protein to 0.48 of control; RERG knockdown reduced RERG protein to 0.40 and the proliferation-associated protein PCNA to 0.71 (Fig. 10b). Experiments with the estrogen-receptor agonists DPN and PPT provide additional evidence of receptor-mediated regulation.

These experiments identify an estrogen-linked route to RERG induction and a proliferative role in endometriotic stromal cells [43]. The GDP-bound 2ATV construct covers 168 of 199 residues, providing experimental structural context (Fig. 9b) [35]. We rank RERG at overall position 10, while its direction-signed external statistic of -0.199 disagrees with our nominated direction. Testing selective RERG induction with receptor- and gene-specific controls would directly assess its effect on astrocyte state.

3.10 WIF1: epigenetic induction and protein supplementation provide intervention routes

WIF1 encodes Wnt inhibitory factor 1, a secreted protein that binds Wnt ligands. Its extracellular function motivates two intervention routes: increasing cellular WIF1 production or supplying the protein directly. Decitabine increased WIF1 mRNA in tested bladder-cell models at $1\,\mu M$ for 72 hours; trichostatin A alone did not consistently induce WIF1, and combination protein increases were modest [50]. In separate methylation experiments using the 5637 bladder-cancer cell line, treatment reduced the methylated promoter fraction from 92.5% untreated to 69% at $30\,\mu M$. These measurements distinguish promoter methylation from transcript induction; decitabine and trichostatin A act through broader epigenetic mechanisms.

Recombinant human WIF-domain supplementation provides a distinct functional route. At $0.29\,\mu M$ for 48 hours, it inhibited T24-cell growth by 40% and increased the frac-

tion of cells in the G1 phase of the cell cycle from 67% to 80%, with supporting Wnt/β-catenin and SKP2 readouts [51]. These experiments measure protein supplementation in bladder-cell models. Surface plasmon resonance (SPR) measures molecular binding. The reported dissociation constant, K_d, describes binding affinity, with lower values indicating tighter binding under comparable conditions. The measurements show that affinity depends on which protein domains are present: WIF-domain, C-terminally truncated WIF+EGF-I–V and full-length WIF1 bind Wnt3a at reported K_d values of 3400, 21.6 and 62.8 nM, respectively (Fig. 10c) [44].

The lipid-containing WIF1 structures provide experimental views of domain organization. The refined 2YGQ coordinates resolve a subset of the expressed construct, with EGF IV/V outside the resolved model (Fig. 9b) [44]. These findings support comparing two approaches: restoring WIF1 expression and adding defined forms of the protein. The experiments could measure secretion, Wnt responses and astrocyte function.

Through the WIF1 analysis, we illustrate how structural information can refine an intervention hypothesis. The SPR measurements assign distinct Wnt3a affinities to the isolated WIF domain, the C-terminally truncated WIF+EGF-I–V construct and the full-length protein. The affinity differences support comparing defined protein forms when developing a WIF1-restoration strategy.

4 Discussion

We developed ASTRA to model astrocyte expression states and identify genes that could help reverse disease-associated changes. Its deep learning component converts gene-expression profiles into a shared 608-feature representation. The subsequent prediction and gene-ranking analyses connect this learned representation to candidate therapeutic targets. ASTRA led the aggregate benchmark ranking and achieved higher mean performance than the four pretrained models used as fixed feature extractors across all nine endpoints. Its reference-score correlation was 0.9252. We then used predictions for held-out donors to identify genes for reduced or increased expression. The subtype and donor-outcome tasks test which information the representation retains. Reference-score prediction supplies the cell-level estimate that we use to rank genes. Together, these analyses connect model performance to the selection of candidate intervention targets.

4.1 From representation learning to target discovery

The three encoder blocks capture partly different expression features. The variational block contributes strongly to reference-score and donor-outcome prediction, while the split and deep blocks add other information. Combining them allows us to analyze individual cells and predict donor outcomes from averages of their cell embeddings. Comparisons with scVI, PCA and pseudobulk show how performance differs across tasks and help explain the value of each approach [15, 16].

Foundation models make broad atlas-derived features available for biological studies. Our comparison asks how well those features support astrocyte prediction when the pretrained encoders remain fixed. ASTRA complements this approach by learning a representation from astrocytes in the study cohort and connecting its state predictions to directional gene rankings. This links the choice of expression representation to the next step in therapeutic discovery: selecting genes whose reduced or increased expression merits testing.

Our target-discovery method identifies genes whose expression rises or falls with the predicted state score. We subtract each donor's mean score to focus on differences among that donor's cells. We also adjust for gene expression level and give greater weight to associations that agree in direction across donor groups. Applying the same procedure to ASTRA-derived and reference scores allows a direct comparison of the resulting rankings. ASTRA improves agreement with the independent atlas in 57 of 60 population–region comparisons. This supports the use of its rankings beyond the discovery cohort.

Each additional analysis addresses a different part of target discovery. Regional comparisons test whether the rankings agree with expression changes in other donors and brain regions. Pathway enrichment relates the genes to known biological processes. Published perturbation signatures identify interventions whose expression changes oppose the nominated pattern. Together, these results help select targets and explain why particular interventions merit testing.

4.2 Intervention hypotheses for the leading targets

Published experiments suggest a different intervention route for each leading target. For GFAP, RNA-targeting approaches and chemical treatments can lower expression. For PLCE1, biochemical studies identify enzyme inhibitors, while genetic experiments link its activity to inflammatory signaling in astrocytes. For RERG, estrogen signaling provides a published route to increased expression. For WIF1, the evidence supports studying both increased expression and direct protein supplementation. Measurements of Wnt3a binding also show why the choice of WIF1 protein form matters.

The organism, cell model and measured response determine how each study informs the proposed intervention. We therefore propose selective RNA lowering for GFAP, tests of enzyme activity and gene dependence for PLCE1, controlled increases in RERG expression, and comparisons of secreted WIF1 protein forms. Experiments should first confirm target engagement: whether the intervention changes the intended target or its activity. They should then assess cell viability across doses, include controls that restore the targeted gene or protein, and measure astrocyte function independently. These steps would help determine whether an expression response reflects the intended mechanism and whether it improves the functions that motivate treatment.

4.3 Implications for therapeutic discovery

ASTRA addresses an early decision in drug development: which target to act on and whether to reduce or increase its expression. We connect human astrocyte expression data to a ranked set of targets and proposed interventions. The regional comparisons show that the rankings capture signals in an independent cohort. The target-level analyses then identify ways to test these predictions through small molecules, RNA lowering, expression induction or protein supplementation. This gives experimental screening a defined objective and an expression pattern against which to assess the response.

The broader therapeutic goal is to restore the support that astrocytes provide to neurons. Candidate interventions could be tested for effects on astrocyte metabolism, neurotransmitter handling and inflammatory signaling, alongside their gene-expression responses. Subsequent experiments could assess effects on neurons and brain tissue. ASTRA helps prioritize the genes and interventions entering this process. By making target selection more systematic and specific to astrocyte biology, we aim to support treatments that preserve the cellular functions needed for cognition in AD.

4.4 Future work

We plan to experimentally validate ASTRA's target nominations using Perturb-seq, which combines pooled genetic perturbations with single-cell RNA sequencing to measure how individual perturbations alter cellular expression programs [52]. We will prioritize suppression of *GFAP* and *PLCE1*, alongside induction of *RERG* and *WIF1*, in human astrocyte models. These experiments will test whether changing each target reverses the disease-associated expression changes predicted by ASTRA. We will assess the transcriptomic responses together with cell viability and astrocyte function to identify targets that support beneficial changes in cellular state.

Guided by these validation results, we plan to design, synthesize and test candidate small molecules that suppress the expression of supported target genes or inhibit their encoded proteins where appropriate. We will seek compounds that reproduce beneficial molecular and functional responses to genetic perturbation. Measurements of target engagement, selectivity and dose-dependent effects will guide compound optimization. This experimental program will connect ASTRA's computational target discovery to the development of astrocyte-directed therapeutics for AD.

4.5 Limitations

Our state-of-the-art claim applies to the evaluated benchmark. We used benchmark observations to select the architecture, repeated splits share donors, and models differ in their access to expression data during training. The target rankings come from observational expression associations. Prospective perturbation studies are needed to establish effects on astrocyte function and AD progression. Supporting experiments span different biological systems, including non-neural cells, and RERG shows disagreement with the external expression data. The chemical-response model has error similar to simple baselines, so we use it descriptively. These findings guide the next steps: evaluation in independent cohorts, experiments in human astrocytes and functional testing of the nominated targets.

We combine accurate astrocyte-state prediction with a method for selecting genes to suppress or induce. The resulting rankings, regional comparisons and target-specific evidence define practical starting points for AD intervention studies. With suitable cohorts and reference expression patterns, the same strategy could support target discovery in other neurodegenerative diseases. Our goal is to connect the analysis of disease-associated cell states to treatments that restore cellular function.

Data availability

ROSMAP single-nucleus expression data and donor annotations are available through the resources described in the original cohort publications, subject to their data-use and access requirements [17, 3]. The independent regional expression data are available through SEA-AD and its associated publications [22, 23]. Published perturbation signatures, intervention measurements and protein structures are available from the cited primary studies and the GEO/CREEDS, RCSB Protein Data Bank and AlphaFold Protein Structure Database resources [28, 39, 35, 36]. We use ChEMBL-derived data under CC BY-SA 3.0 [38].

Acknowledgments

I thank my mentor, Prof. Manolis Kellis, and Dr. Riley Mangan for their guidance and support, and for providing access to the laboratory's data and computing resources. These resources enabled the analyses presented in this paper. I also thank Dr. Riley Mangan for reviewing this work and providing valuable feedback.

References

[1] World Health Organization. Dementia. Fact sheet (2026). https://www.who.int/news-room/fact-sheets/detail/dementia

[2] Mathys, H., Dávila-Velderrain, J., Peng, Z. *et al.* Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332–337 (2019). doi:10.1038/s41586-019-1195-2

[3] Mathys, H., Peng, Z., Boix, C. A. *et al.* Single-cell atlas reveals correlates of high cognitive function, dementia, and resilience to Alzheimer's disease pathology. *Cell* **186**, 4365–4385 (2023). doi:10.1016/j.cell.2023.08.039

[4] Mathys, H., Boix, C. A., Akay, L. A. *et al.* Single-cell multiregion dissection of Alzheimer's disease. *Nature* **632**, 858–868 (2024). doi:10.1038/s41586-024-07606-7

[5] Green, G. S., Fujita, M., Yang, H.-S. *et al.* Cellular communities reveal trajectories of brain ageing and Alzheimer's disease. *Nature* **633**, 634–645 (2024). doi:10.1038/s41586-024-07871-6

[6] Escartin, C., Galea, E., Lakatos, A. *et al.* Reactive astrocyte nomenclature, definitions, and future

directions. *Nat. Neurosci.* **24**, 312–325 (2021). doi:10.1038/s41593-020-00783-4

[7] Patani, R., Hardingham, G. E. & Liddelow, S. A. Functional roles of reactive astrocytes in neuroinflammation and neurodegeneration. *Nat. Rev. Neurol.* **19**, 395–409 (2023). doi:10.1038/s41582-023-00822-1

[8] Liddelow, S. A., Guttenplan, K. A., Clarke, L. E. *et al.* Neurotoxic reactive astrocytes are induced by activated microglia. *Nature* **541**, 481–487 (2017). doi:10.1038/nature21029

[9] Habib, N., McCabe, C., Medina, S. *et al.* Disease-associated astrocytes in Alzheimer's disease and aging. *Nat. Neurosci.* **23**, 701–706 (2020). doi:10.1038/s41593-020-0624-8

[10] Theodoris, C. V., Xiao, L., Chopra, A. *et al.* Transfer learning enables predictions in network biology. *Nature* **618**, 616–624 (2023). doi:10.1038/s41586-023-06139-9

[11] Cui, H., Wang, C., Maan, H. *et al.* scGPT: toward building a foundation model for single-cell multi-omics using generative AI. *Nat. Methods* **21**, 1470–1480 (2024). doi:10.1038/s41592-024-02201-0

[12] Heimberg, G., Kuo, T., DePianto, D. J. *et al.* A cell atlas foundation model for scalable search of similar human cells. *Nature* **638**, 1085–1094 (2025). doi:10.1038/s41586-024-08411-y

[13] Tejada-Lapuerta, A., Schaar, A. C., Gutgesell, R. *et al.* Nicheformer: a foundation model for single-cell and spatial omics. *Nat. Methods* **22**, 2525–2538 (2025). doi:10.1038/s41592-025-02814-z

[14] Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nat. Methods* **15**, 1053–1058 (2018). doi:10.1038/s41592-018-0229-2

[15] Kedzierska, K. Z., Crawford, L., Amini, A. P. & Lu, A. X. Zero-shot evaluation reveals limitations of single-cell foundation models. *Genome Biol.* **26**, 101 (2025). doi:10.1186/s13059-025-03574-x

[16] Boiarsky, R., Singh, N. M., Buendia, A. *et al.* Deeper evaluation of a single-cell foundation model. *Nat. Mach. Intell.* **6**, 1443–1446 (2024). doi:10.1038/s42256-024-00949-w

[17] Bennett, D. A., Buchman, A. S., Boyle, P. A. *et al.* Religious Orders Study and Rush Memory and Aging Project. *J. Alzheimers Dis.* **64**, S161–S189 (2018). doi:10.3233/jad-179939

[18] Rush Alzheimer's Disease Center. Global AD pathology burden (gpath). Clinical and pathological variable documentation. `https://www.radc.rush.edu/docs/var/detail.htm?variable=gpath`

[19] Ganin, Y., Ustinova, E., Ajakan, H. *et al.* Domain-adversarial training of neural networks. In *Domain Adaptation in Computer Vision Applications* 189–209 (Springer, 2017). doi:10.1007/978-3-319-58347-1_10

[20] Becht, E., McInnes, L., Healy, J. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **37**, 38–44 (2019). doi:10.1038/nbt.4314

[21] Ilse, M., Tomczak, J. M. & Welling, M. Attention-based deep multiple instance learning. In *Proc. 35th International Conference on Machine Learning* (*PMLR*) **80**, 2127–2136 (2018).

[22] Gabitto, M. I., Travaglini, K. J., Rachleff, V. M. *et al.* Integrated multimodal cell atlas of Alzheimer's disease. *Nat. Neurosci.* **27**, 2366–2383 (2024). doi:10.1038/s41593-024-01774-5

[23] Hawrylycz, M., Kaplan, E. S., Travaglini, K. J. *et al.* SEA-AD is a multimodal cellular atlas and resource for Alzheimer's disease. *Nat. Aging* **4**, 1331–1334 (2024). doi:10.1038/s43587-024-00719-8

[24] Subramanian, A., Tamayo, P., Mootha, V. K. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005). doi:10.1073/pnas.0506580102

[25] Fang, Z., Liu, X. & Peltz, G. GSEApy: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics* **39**, btac757 (2023). doi:10.1093/bioinformatics/btac757

[26] Chen, E. Y., Tan, C. M., Kou, Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013). doi:10.1186/1471-2105-14-128

[27] Kuleshov, M. V., Jones, M. R., Rouillard, A. D. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016). doi:10.1093/nar/gkw377

[28] Wang, Z., Monteiro, C. D., Jagodnik, K. M. *et al.* Extraction and analysis of signatures from the Gene Expression Omnibus by the crowd. *Nat. Commun.* **7**, 12846 (2016). doi:10.1038/ncomms12846

[29] Ochoa, D., Hercules, A., Carmona, M. *et al.* The next-generation Open Targets Platform: reimagined, redesigned, rebuilt. *Nucleic Acids Res.* **51**, D1353–D1359 (2023). doi:10.1093/nar/gkac1046

[30] Karczewski, K. J., Francioli, L. C., Tiao, G. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020). doi:10.1038/s41586-020-2308-7

[31] Karlsson, M., Zhang, C., Méar, L. *et al.* A single-cell type transcriptomics map of human tissues. *Sci. Adv.* **7**, eabh2169 (2021). doi:10.1126/sciadv.abh2169

[32] Arafeh, R., Shibue, T., Dempster, J. M., Hahn, W. C. & Vazquez, F. The present and future of the Cancer Dependency Map. *Nat. Rev. Cancer* **25**, 59–73 (2025). doi:10.1038/s41568-024-00763-x

[33] UniProt. Reviewed human protein sequences and annotations: GFAP (P14136), PLCE1 (Q9P212), RERG (Q96A58) and WIF1 (Q9Y5W5). `https://www.uniprot.org`

[34] PubChem. Chemical identity and structure records for clomipramine (CID 2801), zinc pyrithione (CID 26041), estradiol (CID 5757) and decitabine (CID 451668). `https://pubchem.ncbi.nlm.nih.gov`

[35] RCSB Protein Data Bank. Protein structures 6A9P, 2C5L, 2ATV, 2YGN and 2YGQ and associated ligand records. https://www.rcsb.org

[36] AlphaFold Protein Structure Database. Human GFAP, PLCE1 and RERG models AF-P14136-F1, AF-Q9P212-F1 and AF-Q96A58-F1. https://alphafold.ebi.ac.uk

[37] EMBL-EBI ChEMBL. Release 37: target and bioactivity database. https://www.ebi.ac.uk/chembl/

[38] EMBL-EBI ChEMBL. Data terms: Creative Commons Attribution–ShareAlike 3.0 Unported. https://chembl.gitbook.io/chembl-interface-documentation/about

[39] Clayton, B. L. L., Kristell, J. D., Allan, K. C. *et al.* A phenotypic screening platform for identifying chemical modulators of astrocyte reactivity. *Nat. Neurosci.* **27**, 656–665 (2024). doi:10.1038/s41593-024-01580-z

[40] Cho, W., Brenner, M., Peters, N. *et al.* Drug screening to identify suppressors of GFAP expression. *Hum Mol Genet* **19**, 3169–3178 (2010). doi:10.1093/hmg/ddq227

[41] Bachetti, T., Zanni, E. D., Adamo, A. *et al.* Beneficial Effect of Phenytoin and Carbamazepine on *GFAP* Gene Expression and Mutant GFAP Folding in a Cellular Model of Alexander's Disease. *Front Pharmacol* **12**, 723218 (2021). doi:10.3389/fphar.2021.723218

[42] Yoshikawa, Y., Naito, M., Emi, A. *et al.* Increase of intracellular Zn^{2+} concentration directly inhibits phospholipase Cε and suppresses inflammation and tumour formation in mice. *Sci Rep* **15**, 41853 (2025). doi:10.1038/s41598-025-25886-5

[43] Monsivais, D., Dyson, M. T., Yin, P. *et al.* ERβ- and prostaglandin E2-regulated pathways integrate cell proliferation via Ras-like and estrogen-regulated growth inhibitor in endometriosis. *Mol Endocrinol* **28**, 1304–1315 (2014). doi:10.1210/me.2013-1421

[44] Malinauskas, T., Aricescu, A. R., Lu, W. *et al.* Modular mechanism of Wnt signaling inhibition by Wnt inhibitory factor 1. *Nat Struct Mol Biol* **18**, 886–893 (2011). doi:10.1038/nsmb.2081

[45] Hagemann, T. L., Powers, B., Mazur, C. *et al.* Antisense suppression of glial fibrillary acidic protein as a treatment for Alexander disease. *Ann Neurol* **83**, 27–39 (2018). doi:10.1002/ana.25118

[46] Hagemann, T. L., Powers, B., Lin, N.-H. *et al.* Antisense therapy in a rat model of Alexander disease reverses GFAP pathology, white matter deficits, and motor impairment. *Sci Transl Med* **13**, eabg4711 (2021). doi:10.1126/scitranslmed.abg4711

[47] Ruzza, P., Vitale, R. M., Hussain, R. *et al.* Interactions of GFAP with ceftriaxone and phenytoin: SRCD and molecular docking and dynamic simulation. *Biochim Biophys Acta* **1860**, 2239–2248 (2016). doi:10.1016/j.bbagen.2016.04.021

[48] Dusaban, S. S., Purcell, N. H., Rockenstein, E. *et al.* Phospholipase C epsilon links G protein-coupled receptor activation to inflammatory astrocytic responses. *Proc Natl Acad Sci U S A* **110**, 3609–3614 (2013). doi:10.1073/pnas.1217355110

[49] Rugema, N. Y., Garland-Kuntz, E. E., Sieng, M. *et al.* Structure of phospholipase Cε reveals an integrated RA1 domain and previously unidentified regulatory elements. *Commun Biol* **3**, 445 (2020). doi:10.1038/s42003-020-01178-8

[50] Shen, C.-H., Li, P.-Y., Wang, S.-C. *et al.* Epigenetic regulation of human WIF1 and DNA methylation situation of WIF1 and GSTM5 in urothelial carcinoma. *Heliyon* **9**, e16004 (2023). doi:10.1016/j.heliyon.2023.e16004

[51] Tang, Y., Simoneau, A. R., Liao, W.-X. *et al.* WIF1, a Wnt pathway inhibitor, regulates SKP2 and c-myc expression leading to G1 arrest and growth inhibition of human invasive urinary bladder cancer cells. *Mol Cancer Ther* **8**, 458–468 (2009). doi:10.1158/1535-7163.MCT-08-0885

[52] Dixit, A., Parnas, O., Li, B. *et al.* Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* **167**, 1853–1866.e17 (2016). doi:10.1016/j.cell.2016.11.038