Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Semi-Independent Transcriptomic, Morphological, Connectivity Dimensions of the Mouse Brain Revealed by CoT-Mining 258 Million Multimodal Associations

Hanchuan Peng

h@braintell.org

Fudan University https://orcid.org/0000-0002-3478-3942

Lijun Wang

Institute for Brain and Intelligence, Southeast University

Fuhui Long

Adadia

Biological Sciences - Article

Keywords:

Posted Date: March 31st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8864665/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Semi-Independent Transcriptomic, Morphological, and Connectivity Dimensions of the Mouse Brain Revealed by CoT-Mining 258 Million Multimodal Associations

Lijun Wang[1,2,+], Fuhui Long[3], Hanchuan Peng[1,4,+,*]

[1] New Cornerstone Science Laboratory, Institute for Brain and Intelligence, Southeast University, Nanjing, China
[2] New Cornerstone Science Laboratory, Institute for Brain and Intelligence, Fudan University, Shanghai, China
[3] Adadia LLC, Seattle, USA
[4] Shanghai Academy of Natural Sciences (SANS), Fudan University, Shanghai, China
[+] Equal contribution
[*] Corresponding author: Hanchuan Peng (h@braintell.org)

Abstract

We address a fundamental question in brain cell typing: to what extent do transcriptomic, morphological, and connectivity features of neurons correspond to one another, and are they redundant or largely independent? To this end, we constructed NeuroXiv2, a large unified multimodal mouse-brain neuron database, integrating neuronal morphology, connectivity, and transcriptomic profiles across 1,385 hierarchical brain regions, 182,483 reconstructed neurons, and 5.24 million cells with expression data for 1,122 genes. We next built the NeuroXiv2 knowledge graph, encoding 258.4 million cross-modal associations among nodes spanning brain regions, neurons, microenvironmental subregions, and transcriptomic cell types. We then developed AI-Powered Open Mining with Chain-of-Thought (AIPOM-CoT) reasoning, an agent that integrates multimodal information and outperforms single-modality analyses in cell-type separation. Using this approach, we achieved a brain-wide, single-neuron resolution quantification demonstrating that neuronal morphology, brain connectivity, and transcriptomic profiles are neither completely redundant or fully independent. Indeed they form three semi-independent dimensions. Further, we showed that combining any two substantially improves predictive power, however none is fully deterministic. We also applied AIPOM-CoT to identify homogeneous Car3-positive neuronal subclasses jointly defined by anatomy, connectivity, and molecular signatures. Overall, the conceptual advances and methodological innovations presented here demonstrate the value of the NeuroXiv2 database, the power of AI-driven analysis, and the broader need for scalable, integrative cross-modal discovery in neuroscience.

Introduction

Over the past decade, neuroscience datasets have expanded rapidly across multiple modalities at single-cell resolution, particularly in the mouse brain(Ament et al., 2022; Aransay et al., 2015; Bria et al., 2016; Gong et al., 2016; Hawrylycz et al., 2023). This growth has fundamentally transformed our ability to investigate neural architecture along several high-throughput dimensions. First, in brain connectivity, mesoscale neurite-projection datasets generated using diverse microscopy techniques now provide standardized, high-resolution measurements of long-range wiring(Oh et al., 2014; Ragan et al., 2012). Second, in neuronal morphology, large collections of single-neuron reconstructions now span extensive brain territories and reveal a rich combination of conserved and divergent structural features(Peng et al., 2021). Third, in molecular profiling, whole-brain spatial and single-cell transcriptomics enable precise localization of diverse transcriptomic cell types within their anatomical context(Yao et al., 2023; Yao et al., 2021; Zhang et al., 2023). Previous studies have begun to examine correspondences among these features—for example,

between axonal projection patterns and molecular diversity(Peng et al., 2021), between neuronal connectivity and morphology(L. Liu et al., 2025), and between local morphology and long-range connectivity(Y. Liu et al., 2025). However, a systematic answer to a fundamental question in brain cell typing remains lacking: to what extent do transcriptomic, morphological, and connectivity features correspond to one another, and are they redundant or largely independent?

A key challenge is how to integrate these modalities to infer meaningful cellular, morphological, and connectivity relationships at single-neuron resolution on a brain-wide scale, particularly when the underlying datasets do not originate from the same individual cells. A common strategy is to register these datasets to the Allen Mouse Brain Common Coordinate Framework version 3 (CCFv3) (Wang et al., 2020), a population-average reference atlas comprising approximately 300 manually curated regions. This anatomical framework allows multimodal datasets to be compared in a common space: although single-cell correspondence is typically not achievable, spatially adjacent "superpixels" may still be jointly analyzed. Because brain tissues imaged under different sample-preparation and acquisition protocols can exhibit markedly different appearances, additional atlases have been constructed in modality- or protocol-specific spaces(Feng et al., 2025; Qu et al., 2022), often retaining the regional definitions of CCFv3. Meanwhile, emerging technologies, such as analyses of dendritic microenvironments, have uncovered brain subdivisions far finer than those represented in CCFv3(Y. Liu et al., 2025). As the level of spatial granularity revealed by the finer scale subregions exceeds conventional atlas parcellations, brain anatomy itself would also complicate the integrative analysis of multimodal datasets.

A further difficulty is that, practically datasets of these modalities are typically housed in separate repositories, use heterogeneous formats, and are often defined across different brain regions, coordinate resolutions, time points or experimental conditions. Substantial harmonization is therefore required before joint analysis is even possible. As a result, cross-modal integration remains highly labor-intensive and requires specialized expertise, spanning multiple databases as well as extensive custom code, creating a major bottleneck for systematic comparative studies. Moreover, because most existing work still centers on single modalities (e.g. (Jiang et al., 2025) or implements only limited multimodal integration, there are few robust precedents for quantitatively and objectively assessing what constitutes a comprehensive integrative analysis.

The uncertainty of the detectable neuronal patterns themselves represents another barrier to truly integrative analysis. Recent work has shown that a tremendous number of previously undocumented neuronal patterns can be discovered by systematically mining 3.7 petavoxels of data across multiple spatial scales, ranging from whole-brain, centimeter-level organization down to sub-micrometer synaptic structures(Liu et al., 2024). In particular, many diverse, previously undescribed neuronal patterns can be revealed, with their variability quantitatively characterized. This is both encouraging and cautionary: encouraging because multimodal data integration may offer similar opportunities, and cautionary because many patterns are likely to be missed when an improperly assembled dataset is analyzed. However, because the conserved patterns are not known *a priori*, technical approaches must be designed carefully to provide both panoramic, brain-wide coverage and sufficient resolution to examine these patterns in detail when needed.

These considerations motivated us to construct a large, representative multimodal neuronal database, NeuroXiv-KG, encompassing 258 million relationships among diverse neuronal entities in the whole mouse brain. We further developed a Chain-of-Thought (CoT) AI agent to infer complex associations embedded within this dataset. We show that conserved and divergent patterns across morphology, connectivity, and transcriptional profiles become apparent only when modalities are integrated, and are difficult to discern from any single modality alone. These three dimensions are much less redundant or independent than one might anticipate, providing the first quantitative evidence for why and how multimodal features should be jointly leveraged. We demonstrate the utility of this framework in Car3-positive neurons, identifying subclasses, delineating anatomical niches, mapping projection networks, and profiling molecular

fingerprints of repeatedly targeted cell populations. NeuroXiv-KG provides the structured multimodal substrate, while AIPOM-CoT provides the reasoning layer that compiles natural language questions into auditable analytical workflows over this substrate.

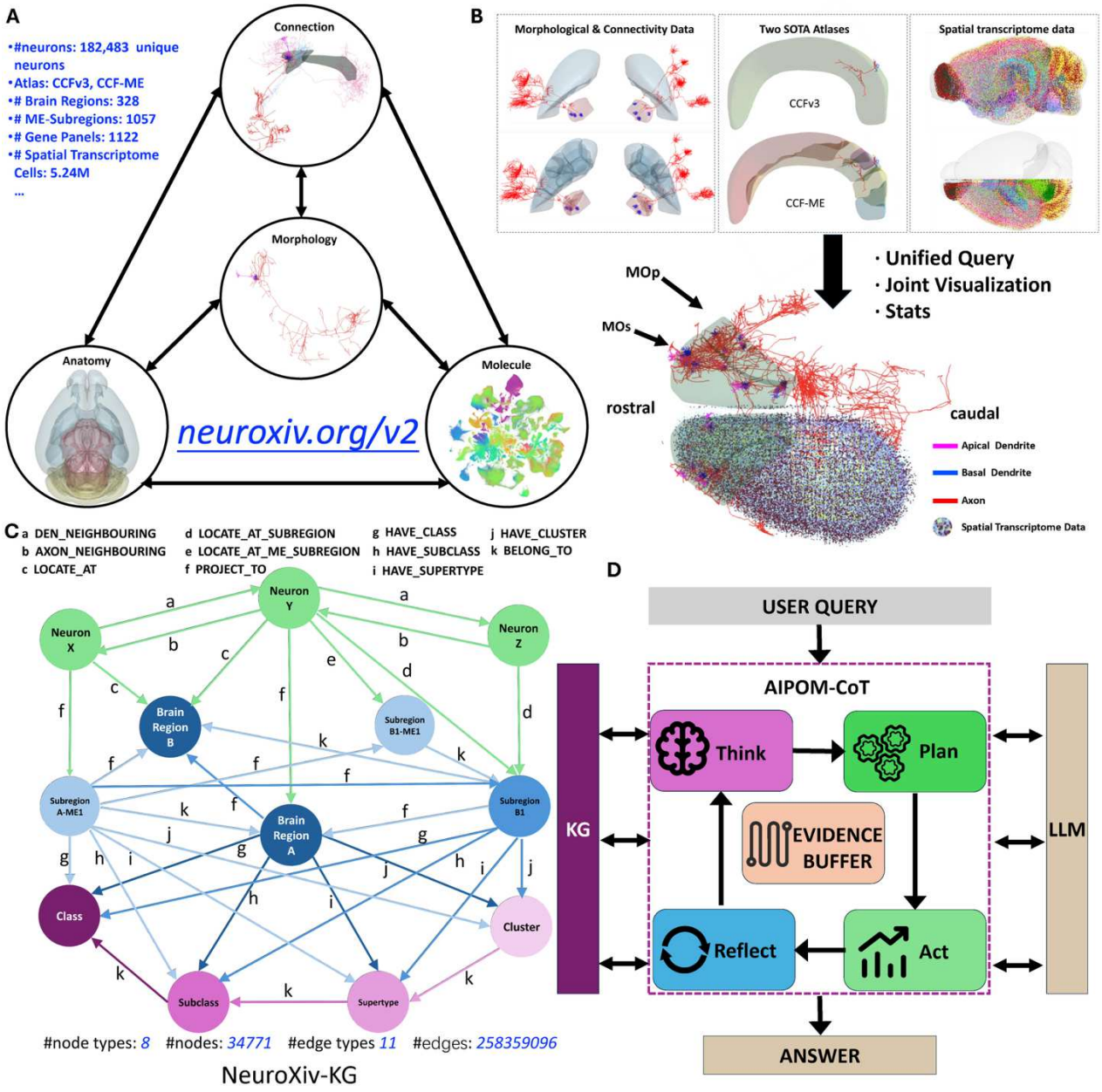


Figure 1. *Integrating multimodal neuroscience data into a unified knowledge graph with AI-powered automated mining.* ***A.*** *Four complementary data modalities integrated into a single knowledge graph: anatomy (328 CCFv3 regions and 1,057 CCF-ME subregions), neuronal morphology (182,483 reconstructed neurons), circuit connectivity (brain-wide projection patterns), and molecular profiles (expression of 1,122 genes across 5.24 million cells).* ***B.*** *Unified spatial registration and joint visualization of neuronal morphology, projection connectivity, anatomical atlases (CCFv3 and CCF-ME), and MERFISH-based spatial transcriptomics within a common coordinate framework.* ***C.*** *NeuroXiv-KG schema comprising eight node types and eleven edge types encoding spatial localization, axonal connectivity, hierarchical taxonomy, and atlas organization, resulting in 34,771 nodes interconnected by 258.4 million edges.* ***D.*** *AIPOM-CoT reasoning architecture, in which* ***Think, Plan, Act, and Reflect*** *modules interact iteratively*

with NeuroXiv-KG. A central Evidence Buffer tracks data quality, statistical significance, and modality coverage, enabling autonomous, provenance-aware question answering.

Results

NeuroXiv2 and NeuroXiv-KG: databasing 258 million multimodal associations of brain-wide m-types, c-types, t-types and s-types of neurons

We considered four major categories of neurodata: molecular composition, cellular morphology, circuit connectivity, and anatomical organization. In prior studies, these modalities are often formalized as complementary neuronal classification dimensions, corresponding respectively to *t*-types (transcriptomic types), *m*-types (morphological types), *c*-types (connectivity types), and *s*-types (soma-location types)(Gouwens et al., 2019; L. Liu et al., 2025; Peng et al., 2021; Zeng, 2022; Zeng & Sanes, 2017). Note that, as the literature often uses "projection" and "connectivity" interchangeably, we follow this convention below especially in figures where no ambiguity arises.

To integrate these modalities, we constructed a *NeuroXiv2* database through a three-stage pipeline (**Figure 1**). First, we assembled a large-scale, multimodal database (**Figure 1A and Supplementary Figures 1-2**) comprising 182,483 single-neuron morphological reconstructions aggregated from multiple public sources(Gao et al., 2022; Gao et al., 2023; Jiang et al., 2025; Jiao et al., 2025; Liu et al., 2024; Peng et al., 2021; Qiu et al., 2024; Winnubst et al., 2019), brain-wide axonal projection connectivity matrices spanning major brain regions(L. Liu et al., 2025; Xiong et al., 2025), and transcriptomic profiles of 1,122 genes across 5.24 million cells(Zhang et al., 2023). Together, these datasets capture neuronal structure, long-range connectivity, and molecular identity comprehensively. Second, all morphological, connectivity, and molecular datasets were spatially registered to the CCFv3 reference atlas using *mBrainAligner(Li et al., 2022; Qu et al., 2022)*, when necessary, ensuring objective cross-modal comparison within a unified coordinate framework (**Figure 1B and Supplementary Figures 2-3**). Third, to enhance spatial resolution and provide better granularity of analysis, we incorporated 1,057 newly defined subregions from the recent atlas Common Coordinate Framework Microenvironment (CCF-ME) (Y. Liu et al., 2025), in addition to the 328 previously established, larger brain regions in CCFv3. In total, 1,385 hierarchically organized brain regions were used as anatomical references for integrative analyses of neuronal morphology, connectivity, and molecular signatures (**Figure 1B and Supplementary Figures 1-3**).

Next, we constructed *NeuroXiv-KG*, a knowledge graph (KG) that encodes relationships among heterogeneous data objects as a structured semantic network (**Figure 1C and Supplementary Figures 4-5**). Because all datasets were spatially registered, we defined a unified schema comprising (a) eight node types, including neurons, brain regions, subregions, and transcriptomic cell types organized across four hierarchical levels (*Class, Subclass, Supertype,* and *Cluster*), and (b) eleven edge types capturing spatial localization, axonal projection connectivity, hierarchical taxonomy, and atlas organization (**Figure 1C**, **Methods, and Supplementary Figures 4-5**). The resulting knowledge graph contains 34,771 nodes connected by 258.4 million edges, enabling cross-modal queries that traverse molecular, morphological, connectivity, and anatomical domains (**Supplementary Figure 5**). *NeuroXiv-KG* thus transforms previously siloed, single-modality datasets into a unified, machine-readable multimodal resource that allows large-scale integrative analysis and automated reasoning.

AIPOM-CoT: an AI agent for deep-mining of multimodal associations

The unified data representation in *NeuroXiv-KG* allows domain-specific, natural-language–driven discovery. We developed *AIPOM-CoT* (AI-Powered Open Mining with Chain-of-Thought), a reasoning

agent that translates biological questions into multi-step, goal-directed analytical workflows (**Figure 1D and Supplementary Figures 6-8**).

AIPOM-CoT consists of four tightly coupled modules (**Figure 1D and Supplementary Figure 6**). (a) "**Think"** generates high-level reasoning strategies by decomposing user queries into executable sub-tasks and selecting relevant graph schemas. (b) "**Plan**" structures these sub-tasks into ordered analytical steps with explicit biological intent. dependencies. (c) "**Act**" executes knowledge graph retrievals and associated statistical analyses, including effect-size estimation, confidence interval calculation, and false-discovery-rate correction. (d) "**Reflect**" evaluates accumulated evidence, identifies missing information or uncertainty, and determines whether additional analysis is required. A central *Evidence Buffer* tracks data quality, statistical significance, and modality coverage, allowing the agent to autonomously terminate when sufficient evidence has been accumulated.

This architecture enables AIPOM-CoT to address diverse biological questions without pre-defined workflows (**Figure 2 and Supplementary Figure 6**). For example, when asked *"What can you tell me about Claustrum neurons?"*, the agent autonomously identifies relevant molecular, morphological, connectivity, and anatomical modalities; formulates appropriate graph queries; integrates retrieved evidence; and produces an interpretable summary with full provenance (**Supplementary Figure 8**), including data snapshots, tool invocation sequences, parameters, and intermediate inputs and outputs. By explicitly logging all reasoning steps and computational operations, AIPOM-CoT enables users to inspect, validate, modify, and extend analyses, transforming atlas interaction from manual, single-step querying into planned, reproducible, and multimodal biological investigation.

Rather than relying on keyword matching or fixed templates, AIPOM-CoT first performs semantic intent analysis, reasoning about the underlying scientific objective of each query (**Figure 2A and Supplementary Figure 7**). User questions are classified by analytical intent (for example, profiling, comparison, screening, or explanation), expected depth, relevant data modalities, and target entity types. This structured interpretation routes queries to appropriate planning strategies, enabling flexible handling of diverse question types. Instead of retrieving predefined workflows, AIPOM-CoT dynamically synthesizes analytical steps by reasoning over the knowledge graph schema (**Figure 2B and Supplementary Figure 7**). Each step is defined by its biological purpose, required relationships, and data dependencies, and candidate strategies are evaluated for relevance and expected information gain before execution. The core reasoning engine operates in iterative **Think–Plan–Act–Reflect** cycles, with **Reflect** enabling autonomous decision-making (**Figure 2C and Supplementary Figure 7**). After each execution step, the agent evaluates whether results match expectations, identifies key findings and uncertainties, and determines appropriate next actions. Based on this assessment, the agent may continue, revise the plan, explore emerging patterns in greater depth, pivot to alternative analytical directions, or terminate once sufficient evidence has been obtained. These decisions arise from contextual reasoning over accumulated evidence (**Supplementary Figure 8**) rather than fixed rules, allowing flexible analytical trajectories even for similarly phrased questions.

All evidence is accumulated in a structured buffer that tracks data quality, statistical support, modality coverage, and analysis provenance (**Supplementary Figure 9**). Overall confidence is computed by integrating data completeness, coverage across molecular, morphological, and connectivity modalities, and statistical significance. When confidence exceeds a predefined threshold and uncertainties are resolved, the agent may autonomously terminate the analysis, even if additional steps were originally planned. Conversely, stalled confidence prompts replanning or deeper investigation. Analysis proceeds until one of three conditions is met: the agent determines that the question has been adequately answered, predefined computational limits are reached, or evidence quality plateaus, indicating diminishing returns. In practice, most analyses terminate autonomously based on sufficient evidence rather than resource constraints, and the termination reason is logged to ensure transparency and auditability.

The **Act** module executes analysis using a set of reusable neuroscience tools (**Figure 2D**), including multimodal fingerprint construction across molecular, morphological, and connectivity features; similarity and divergence measurements; and statistical validation. All operations record parameters and data snapshots to ensure reproducibility. For example, when asked *"Which regions show the highest cross-modal mismatch?"*, the agent identifies this as a systematic screening task and applies a comparative strategy across brain regions. It constructs multimodal fingerprints, quantifies divergence between modalities, and evaluates statistical support before ranking regions by mismatch strength. The analysis terminates automatically once sufficient evidence is obtained, returning an interpretable summary together with figures, complete provenance, and machine-readable records that enable full reproduction of the results.

It can be seen that our architecture enables adaptive, transparent, and reproducible multimodal analysis, handling novel questions through schema-aware reasoning and self-correcting workflows. In another word, NeuroXiv-KG serves as the structured multimodal substrate, whereas AIPOM-CoT functions as the reasoning layer that translates natural language queries into explicit, auditable analytical workflows operating on this substrate.

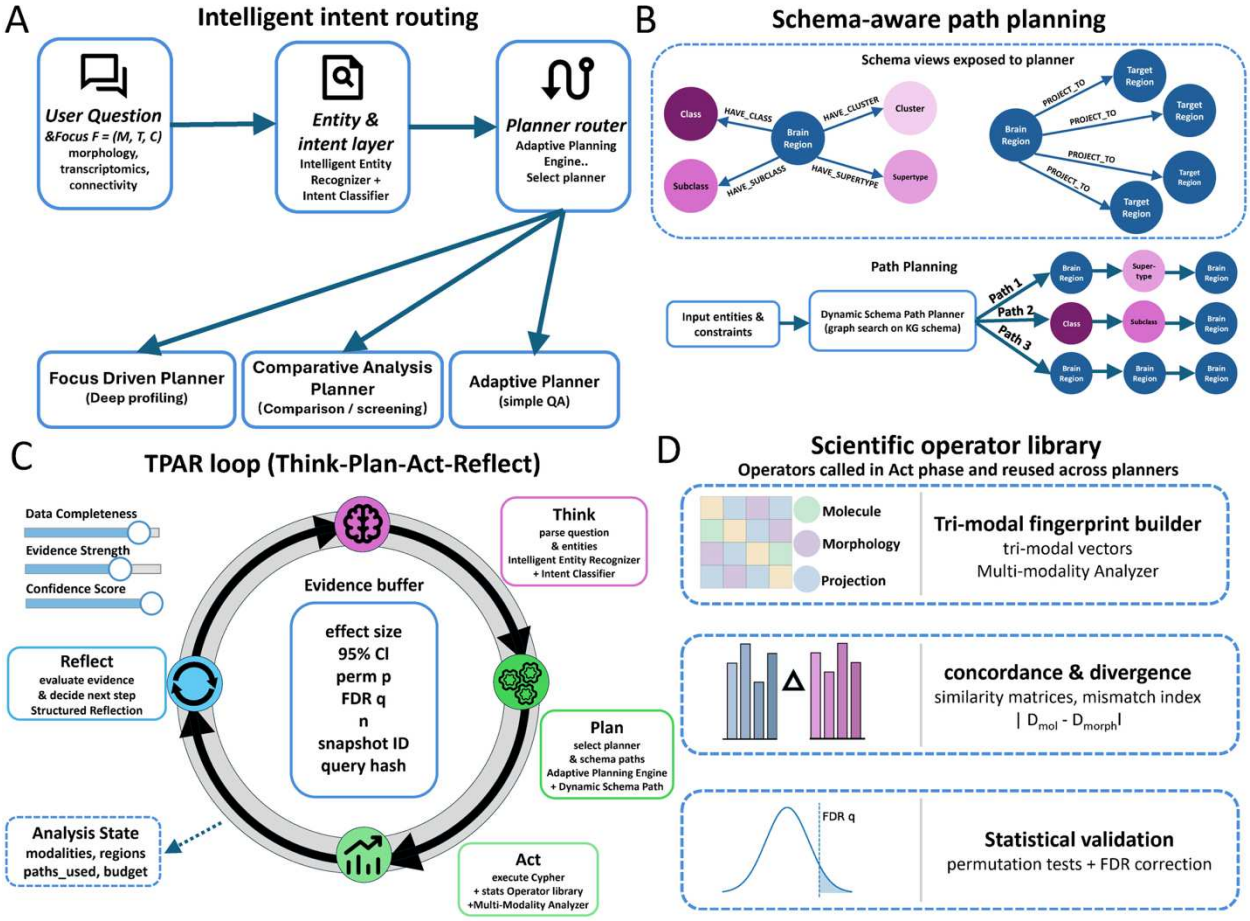


Figure 2. *AIPOM-CoT architecture for schema-adaptive knowledge graph mining.* ***A.*** *LLM-driven semantic analysis classifies user intent and routes queries to appropriate planners (Focus-Driven, Comparative, or Adaptive).* ***B.*** *Planners dynamically generate analytical steps by reasoning over the knowledge graph schema.* ***C. Think-Plan-Act-Reflect*** *cycle enables autonomous decision-making: continue, replan, deepen, pivot, or terminate based on accumulated evidence.* ***D. Act*** *module executes reusable neuroscience tools with full execution logging for reproducibility.*

AIPOM-CoT excels in adaptive planning and evidence integration

We evaluated the generalization ability of AIPOM-CoT across a diverse set of neuroscience queries using a systematic benchmark spanning simple factual retrieval to complex multimodal analyses. AIPOM-CoT was compared against ReAct(Yao et al., 2022) (a state-of-the-art agentic framework), GPT-4o (OpenAI, 2024) (direct prompting), retrieval-augmented generation(Guu et al., 2020; Lewis et al., 2020), and template-based approaches (Badenes-Olmedo & Corcho, 2024; Francis et al., 2018) under a comprehensive evaluation framework that assessed not only answer accuracy, but also planning coherence, reasoning quality, and self-correction — capabilities essential for autonomous scientific discovery. AIPOM-CoT achieved an overall performance score of 0.78 and an agent capability index of 0.81, substantially outperforming all baselines (**Figure 3A**). Performance gains were most pronounced for complex queries requiring multi-step workflows (+34.5% over ReAct; +18.2% over GPT-4o), demonstrating that schema-aware tool binding, evidence-driven reflection, and adaptive planning yield measurable advantages precisely in regimes where intermediate decisions critically shape outcomes.

Across four core dimensions, AIPOM-CoT showed systematic advantages (**Figure 3B**). Planning quality (0.79 vs. 0.62) reflected schema adaptive synthesis of analytical strategies derived from graph structure rather than template retrieval. When asked to compare cortical profiles, AIPOM-CoT generated complete six step workflows that included PCA reduction and statistical validation, whereas ReAct omitted critical steps. Reasoning capability (0.81 vs. 0.56) stemmed from systematic multimodal sampling. AIPOM-CoT retrieved molecular, morphological, and projection data with 94% completeness, compared with 73% for ReAct and 58% for RAG. Self-correction (0.81 vs. 0.69) resulted from LLM based reflection that performed metacognitive analysis after each step, detecting insufficient evidence in 35% of queries (8 of 23 cases) and successfully recovering in 88% (7 of 8) through strategic revisions such as pooling taxonomic levels, expanding spatial radii, or switching metrics, compared with 40% (2 of 5) for ReAct. Query understanding (0.83 vs. 0.69) reflected structured semantic analysis that explicitly reasoned about question category, depth, and required modalities.

Planning quality was decomposed into coherence (0.71 vs. 0.67), efficiency (0.89 vs. 0.70), and adaptability (0.74 vs. 0.48) (**Figure 3C, top left**). Coherence showed comparable performance, indicating that both systems maintain step to step consistency. Efficiency highlighted AIPOM-CoT's superior depth selection. Simple queries terminated in two to three steps after completeness was confirmed, whereas complex analyses executed six to eight coordinated steps. In contrast, ReAct exhibited either redundant validation or premature termination. Adaptability, the most discriminative dimension, quantified strategic adjustments when initial approaches produced unexpected patterns. In 15 of 23 tasks, AIPOM-CoT detected insufficient coverage and autonomously broadened searches by expanding to parent regions when fewer than 10 neurons were returned, adding morphological validation when molecular variance exceeded 0.4, or invoking reciprocal checks when cross modal mismatch exceeded 0.6. By contrast, ReAct maintained committed strategies in 18 of 23 tasks despite contradictory evidence, reflecting fundamental architectural differences. AIPOM-CoT employs a reflection loop that evaluates multiple options at each step, whereas ReAct relies on fixed templates that limit mid analysis revision.

Reasoning performance was further decomposed into logical consistency and evidence integration (**Figure 3C, top right**). AIPOM-CoT achieved higher logical consistency than ReAct (0.74 vs. 0.54) and near perfect evidence integration (0.98 vs. 0.42). Logical consistency assessed whether successive claims remained internally aligned without unsupported leaps. AIPOM-CoT was more stable because each step was executed as an explicit knowledge graph operation with typed outputs, ensuring that downstream statements were constrained by retrieved and computed results. Evidence integration measured whether conclusions were supported by complete and traceable evidence rather than isolated fragments. AIPOM-

CoT consistently assembled supporting evidence by traversing complementary relations and aggregating results into claim linked records, thereby reducing modality gaps and improving end to end traceability.

Our approach also improved the self-correction performance of AIPOM-CoT (**Figure 3C, bottom left**). Compared with ReAct, AIPOM-CoT more effectively detected failures (0.81 vs. 0.73), applied corrective actions (0.82 vs. 0.68), and iteratively refined outputs (0.78 vs. 0.71). Critically, AIPOM-CoT employs quantitative insufficiency criteria, including empty query returns, weak statistical support, and high uncertainty, to trigger replanning, whereas ReAct largely relies on coarse binary signals. As a result, AIPOM-CoT produced the most complete outputs (**Figure 3C, bottom right**), achieving a score of 0.98 compared with ReAct (0.37), GPT-4o (0.92), and RAG (0.52). In practice, AIPOM-CoT consistently reports confidence intervals, false discovery rate corrected *p* values, and execution traces, which are frequently missing from baseline methods but are essential for distinguishing robust biological signals from sampling noise.

Our results show AIPOM-CoT's capabilities in autonomous task decomposition, systematic multimodal integration, quantitative evidence evaluation, and strategic flexibility generalize robustly across diverse query types, including factual lookup, comparative analysis, hypothesis generation, and exploratory discovery. Performance advantages are consistent across all evaluated dimensions, with particularly strong gains in adaptive planning (54%), evidence integration (perfect score of 0.98), and answer completeness (169%). These improvements, together with full provenance and traceability (**Supplementary Figure 8**) demonstrated in the subsequent results, suggest that AIPOM-CoT functions as a general discovery platform rather than a system optimized for narrow or task specific benchmarks.

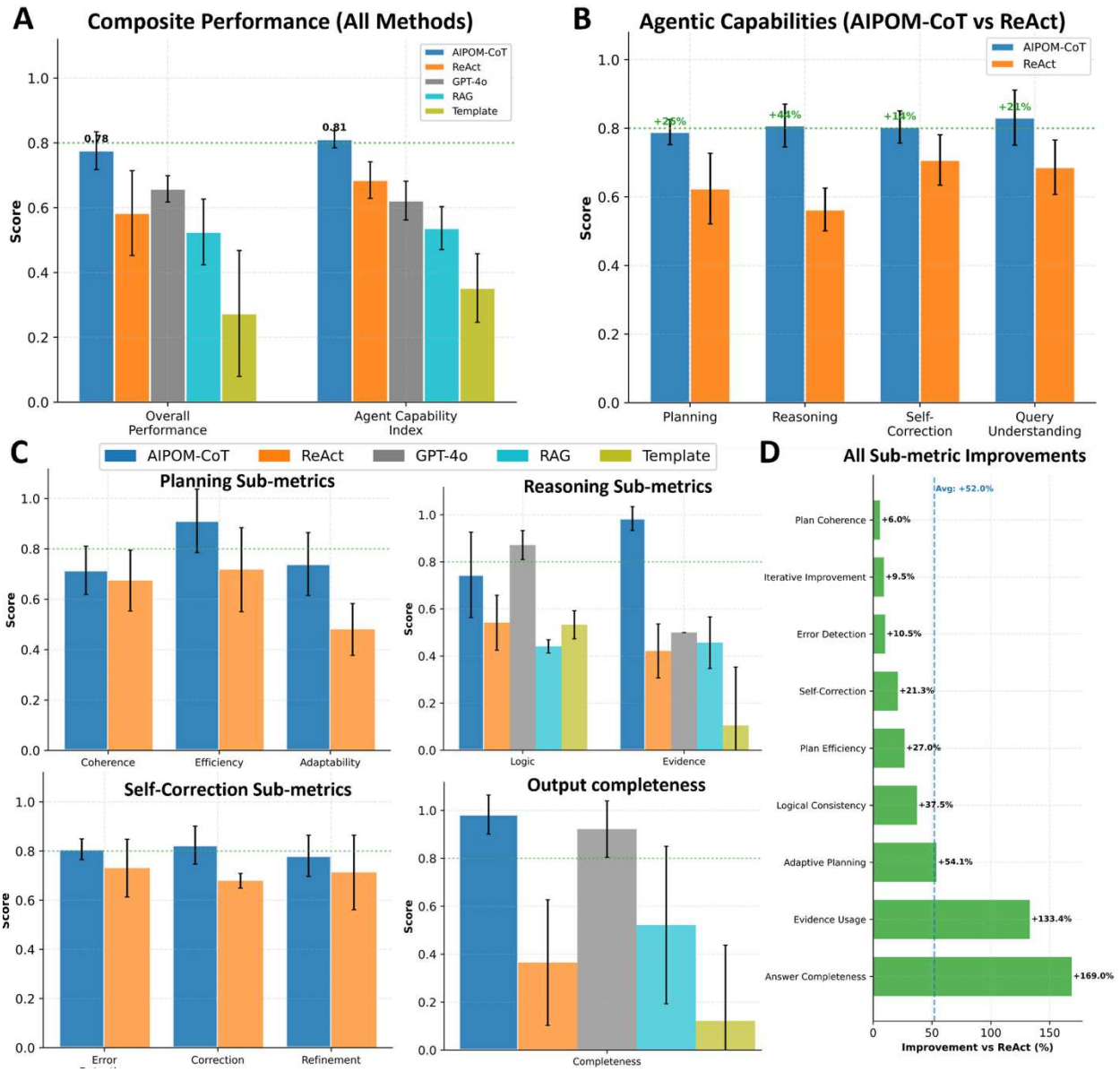


Figure 3. *Performance evaluation of AIPOM-CoT's agentic capabilities.* ***A.*** *Composite performance comparison across five methods (AIPOM-CoT, ReAct (Yao et, al; 2023), GPT-4o (OpenAI, 2024), RAG, Template) on overall performance and agent capability index.* ***B.*** *Core agentic capabilities comparison between AIPOM-CoT and ReAct across four dimensions: planning, reasoning, self-correction, and query understanding.* ***C.*** *Detailed sub-metric decomposition for each capability dimension. Planning evaluated through coherence, efficiency, and adaptability. Reasoning assessed via logical consistency and evidence integration. Self-correction measured by error detection, correction implementation, and iterative refinement. Output quality quantified through answer completeness.* ***D.*** *Relative performance improvements of AIPOM-CoT over ReAct across all sub-metrics.*

NeuroXiv-KG and AIPOM-CoT quantify brain region-wise cross-modality consistency and divergence

Having established automated cross-modal integration for targeted queries, we next examined whether the system could support systematic, whole-brain analyses of cross-modal relationships. We instructed AIPOM-CoT to construct tri-modal fingerprints encompassing molecular, morphological, and projection

features for all 32 major brain regions represented in the knowledge graph, and to compute pairwise similarity matrices for each modality (**Figure 4A**).

The resulting similarity patterns revealed a pronounced contrast across modalities. Morphological and projection fingerprints exhibited clear block structures, with groups of regions showing high mutual similarity, such as cortical areas in morphology and hypothalamic subregions in projection patterns. This modular organization indicates that relatedness at the levels of neuronal shape and connectivity often extends beyond strict anatomical boundaries. In contrast, the molecular fingerprint matrix was characterized by sparse inter-region similarity, with elevated values largely restricted to the diagonal. This asymmetry, namely modular organization in morphology and projection versus strong regional specificity in molecular composition, suggests that molecular profiles are predominantly region specific, whereas morphological and projection properties can converge across anatomically distant brain regions.

To quantify the prevalence of cross-modal concordance versus divergence across brain regions, we computed cross-modal mismatch indices for all region pairs (**Figure 4B**). The molecule–morphology divergence matrix captured cases in which molecular composition diverged from morphological organization, whereas the molecule–projection divergence matrix highlighted region pairs whose molecular profiles were misaligned with their projection patterns. On average, about 40% of region pairs exhibited systematic cross-modal divergence, including 57% for molecule–morphology and 24% for molecule–projection comparisons. These results indicate that multimodal concordance, in which similarity in one modality predicts similarity in another, is not the dominant organizational principle. Instead, brain organization appears to be structured along partially independent axes, with morphological and projection similarities frequently arising independently of molecular composition.

To illustrate the biological significance of these divergence patterns can be seen in two representative cases automatically identified by AIPOM-CoT. First, we analyzed the lateral hypothalamic area (LHA) and the tuberal nucleus (TU), which exhibit high morphological similarity but pronounced molecular divergence (**Figure 4C**) for the currently available data. Their morphological feature profiles, including axonal length, branching complexity, and dendritic arbor characteristics, are nearly indistinguishable for the available neurons in the NeuroXiv2 database. In contrast, their neuronal molecular compositions differ substantially. LHA contains a highly heterogeneous population comprising Skor1+, Foxb1+, and Pitx2+ glutamatergic neurons, as well as specialized neuropeptidergic subtypes such as orexin and melanin-concentrating hormone neurons that support its role as a multifunctional hypothalamic hub regulating arousal, feeding, reward, and motivated behavior. TU, however, is dominated by a more homogeneous GABAergic population specialized for hedonic feeding regulation and receives selective neurotensin inputs from the lateral septum.

Second, we compared the secondary motor cortex (MOs) with LHA, which exhibit similarity in projection patterns despite complete divergence in molecular identity (**Figure 4D**). Although their molecular compositions are entirely distinct, with MOs comprising laminar organized cortical glutamatergic neurons including IT type, ET type, and CT type populations, and LHA consisting of hypothalamic glutamatergic neurons, both regions project to a largely overlapping set of downstream targets. These shared targets include the striatum (STR, CP, and ACB), thalamic nuclei (MD, VM, and ZI), and multiple brainstem structures (PAG, MRN, and MB). While the relative projection strengths differ, with MOs preferentially engaging cortico-striatal motor loops and LHA emphasizing hypothalamic brainstem arousal pathways, the extensive target overlap indicates a potentially shared control architecture. In this framework, distinct brain systems coordinate behavior through common downstream effectors, a concept consistent with the convergence of input pathways onto shared relay neurons in the motor thalamus(Koster & Sherman, 2024) and the finding that multiple hypothalamic populations drive similar behaviors through convergent output to the brainstem(Lovett-Barron et al., 2020).

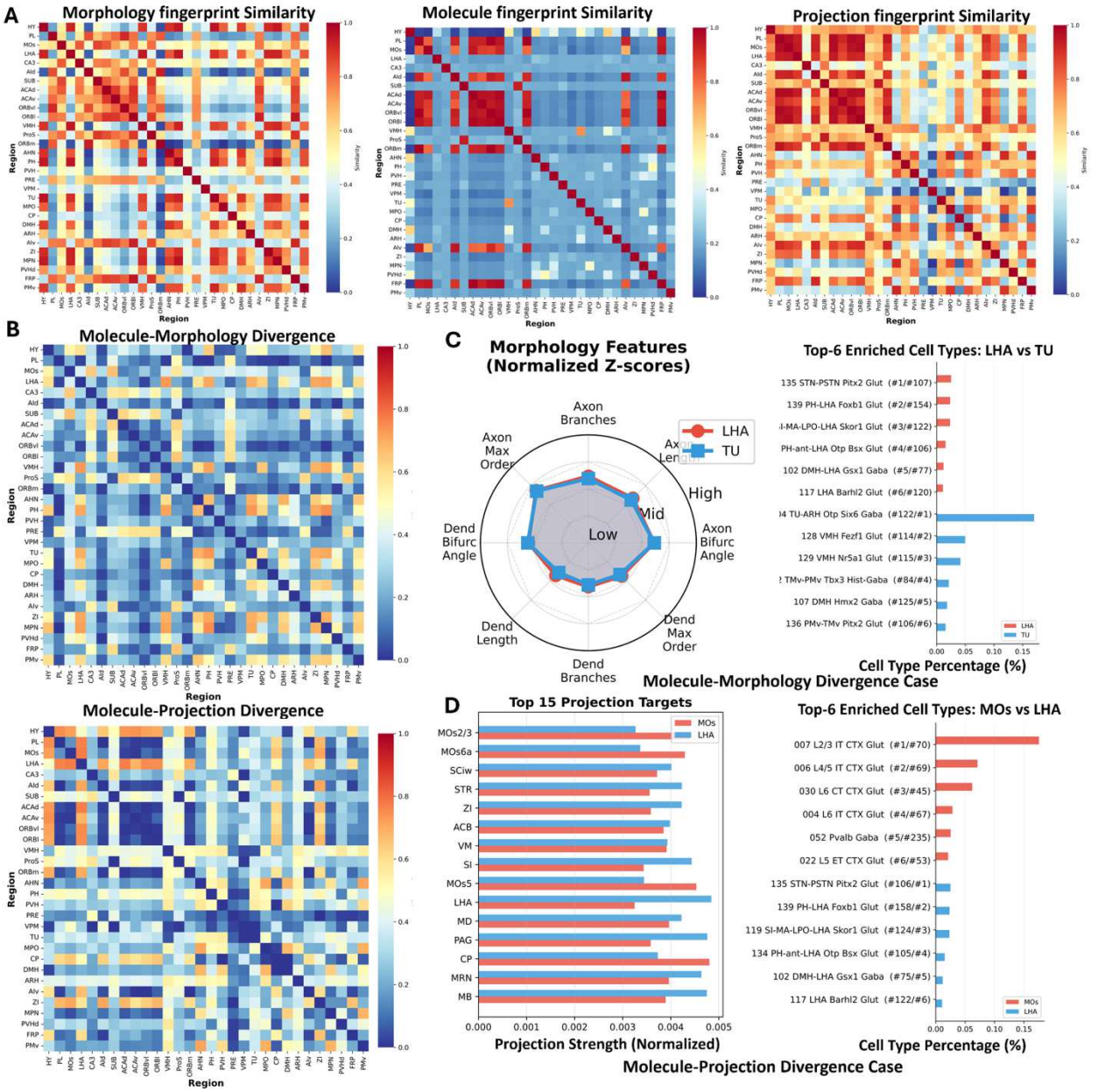


Figure 4. *Whole-brain tri-modal fingerprinting reveals cross-modal concordance and divergence.* ***A***. *Similarity matrices for morphological, molecular, and projection fingerprints across 30 brain regions. Morphology and projection exhibit modular block structure, whereas molecular profiles show strong region specificity.* ***B***. *Cross-modal mismatch matrices quantify systematic divergence between modalities, with ~40% of region pairs exhibiting cross-modal dissociation.* ***C***. *LHA versus TU example showing strong morphological similarity but pronounced molecular divergence. We compared the eight main morphological features of the two brain regions (left), as well as the composition and ranking of the top 6 cell types in each (right).* ***D***. *MOs versus LHA example showing convergent projection targets despite complete molecular divergence. We compared the top 15 projection targets and their corresponding projection strengths for two brain regions (left), as well as the composition and ranking of the top 6 cell types within each region (right).*

Semi-independency of morphology, connectivity and transcriptome demensions highlights the importance of their integration

The cross-modal divergence observed in **Figure 4** indicates that molecular, morphological, and connectivity patterns frequently encode non-redundant information. This observation raises a central question: When modalities diverge in a nontrivial manner, how much complementary information does each modality contribute, and what measurable gains result from their integration? To address it, we quantified the benefits of multimodal data fusion at two spatial scales: cross-region predictive accuracy and within-region spatial homogeneity.

We first evaluated whether features from one modality could predict the regional identity defined by another modality, and whether combining modalities improved predictive performance (**Figure 5A**). Using machine-learning classifiers (**Methods**) trained on morphological features, molecular microenvironment profiles, and projection patterns, we performed three cross-modal prediction tasks: predicting projection targets from morphology and molecular data, predicting morphological groupings from molecular and projection data, and predicting molecular composition from morphological and projection data. Single-modality predictors achieved only moderate accuracy, ranging from 0.48 to 0.63 across tasks, indicating that each modality captures only partial information about the others.

In clear contrast, multimodal predictors that integrated two complementary modalities using late-fusion strategies consistently outperformed single-modality models (**Figure 5C**), achieving accuracies of 0.62 to 0.74, corresponding to improvements of 15% to 42% depending on the task. For projection prediction, morphology alone reached an accuracy of 0.55 and molecular profiles alone reached 0.52, whereas their combination achieved 0.74, representing a 34% improvement over the best single modality. For morphology prediction, projection patterns alone achieved 0.63 and molecular data alone achieved 0.56, while their fusion reached 0.73, yielding a 16% improvement. For molecular prediction, projection data alone reached 0.52 and morphology alone reached 0.48, but their integration achieved 0.62, corresponding to a 19% improvement. Together, these consistent gains across all three tasks demonstrate that the modalities contribute complementary information. Each modality captures distinct aspects of neuronal organization that cannot be fully inferred from the others, and their integration substantially enhances predictive power.

Beyond cross-modal prediction, we want to find out whether multi-modality data can make the boundaries between brain regions more distinct and the internal structures more stable (**Figure 5B**). We evaluated whether multimodal integration improves region-level separability by computing confusion matrices for region classification (**Methods**). Single-modality features exhibited substantial cross-region overlap, yielding global confusion scores of 3.4295 for morphology (**Figure 5D, top right**), 1.6189 for molecular profiles (**Figure 5D, bottom left**), and 2.2827 for projection patterns (**Figure 5D, bottom right**). In contrast, the multimodal representation markedly reduced cross-region confusion, achieving a global confusion score of 0.8867 (**Figure 5D, top left**). Quantitatively, multimodal integration reduced the global confusion score by 74.1% relative to morphology alone, by 45.2% relative to molecular features alone, and by 61.2% relative to projection features alone. These results demonstrate that integrating complementary modalities yields substantially more discriminative and region-specific representations.

In parallel, we examined whether multimodal integration improves spatial homogeneity within anatomically defined brain regions. We compared within-region feature dispersion for single-modality representations with that of the integrated multimodal representation (**Methods**). The distributions of spatial homogeneity change across regions consistently shifted toward improvement following multimodal integration (**Figure 5E**), indicating a general reduction in within-region dispersion. Specifically, dispersion was reduced in 34.6% of regions when starting from molecular representations (**Figure 5E, top**), in 58.8% of regions when starting from morphology-based representations (**Figure 5E, middle**), and in 59.6% of regions when starting from connectivity-based representations (**Figure 5E, bottom**). All these distributions demonstrate that incorporating complementary modalities frequently tightens within-region similarity structure, yielding representations that are more spatially homogeneous within anatomical boundaries.

Overall, our analyses demonstrate that multimodal data fusion confers systematic benefits for cross-modal prediction and brain region separability. Across three cross-modal prediction tasks, multimodal integration improved accuracy by 15–42% (**Figure 5C**), indicating that morphology, molecular profiles, and projection patterns each contribute distinct and complementary information about neuronal and regional identity rather than encoding redundant signals. At the cross-region level, multimodal representations substantially reduced ambiguity in region classification, as reflected by lower confusion scores and sharper region separability compared with single-modality features (**Figure 5D**). In addition, multimodal integration reduced within-region dispersion by 34.6–59.6%, depending on the baseline modality (**Figure 5E**), thereby enhancing spatial homogeneity within anatomical regions. Collectively, these results support a theory that brain regions are defined by the joint attributes of molecular, structural, and connectivity properties, rather than by any single feature space alone.

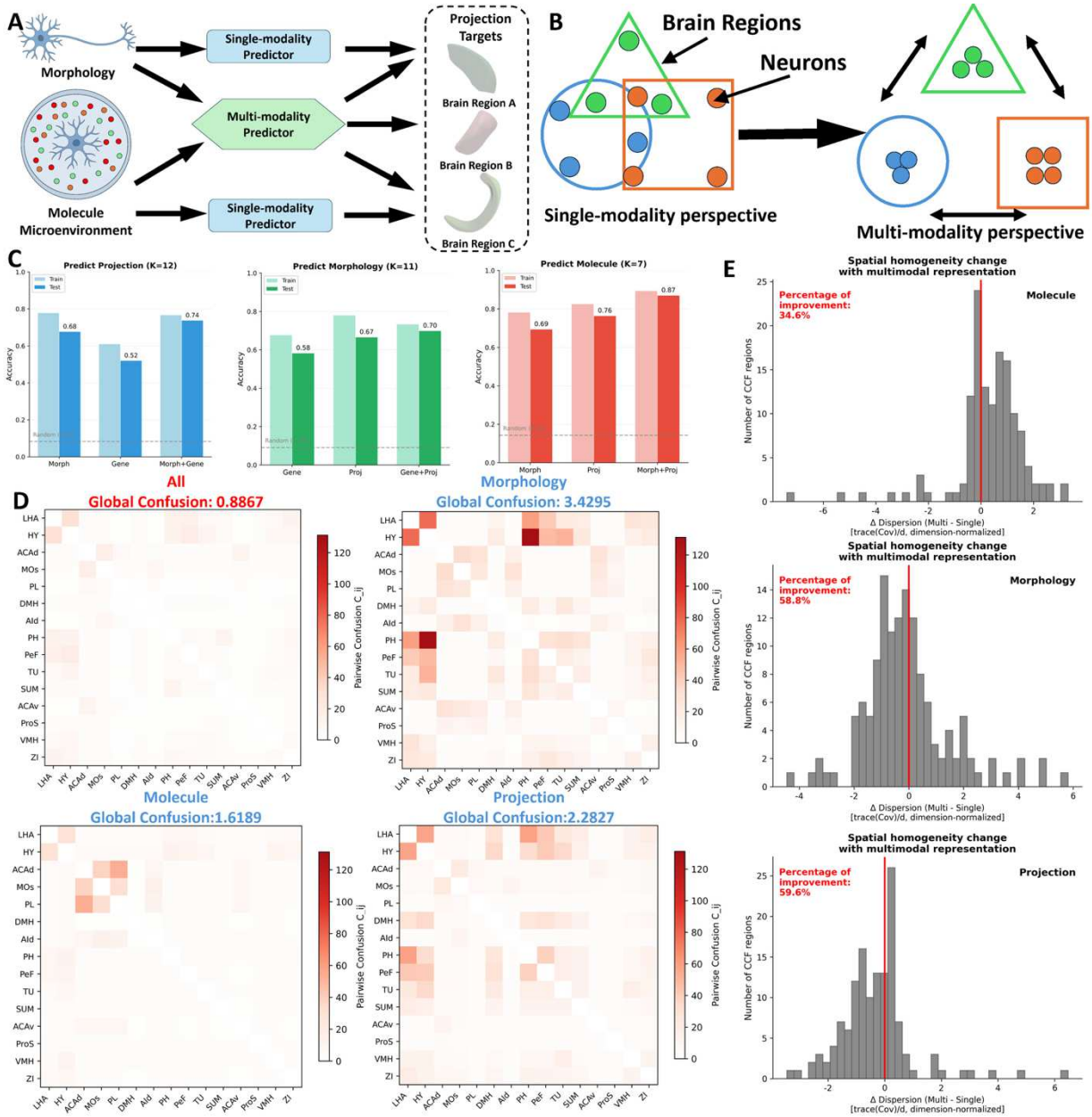


***Figure 5**. Multimodal data fusion enhances prediction and regional organization. **A.** Cross-modal prediction of regional identity from complementary modalities. **B.** Multimodal integration reduces cross-region confusion and within-region dispersion. **C.** Late fusion improves prediction accuracy by 15–42%, indicating complementary information across modalities. **D.** Cross-region confusion matrices show sharper regional boundaries with multimodal integration. **E.** Multimodal representations reduce within-region dispersion by 34.6–59.6% across most regions.*

Automated neuron-retrieval and multi-modality data integration: a case study for Car3

To demonstrate the capacity of AIPOM-CoT for automated cross-modal analysis, we posed an open-ended biological question to it: "*Can you tell me something about Car3+ neurons?*" (**Figure 6A and Supplementary Figure 6-8**). Without manual data curation or preselected datasets, the agent autonomously decomposed this query into a multi-step analytical workflow. Specifically, it (1) identified transcriptomic subclasses enriched for Car3 expression, (2) determined the brain regions showing the strongest enrichment of these subclasses, (3) retrieved morphological reconstructions of neurons from the enriched regions, (4) mapped the projection targets of the reconstructed neurons, and (5) characterized the molecular composition of the target regions.

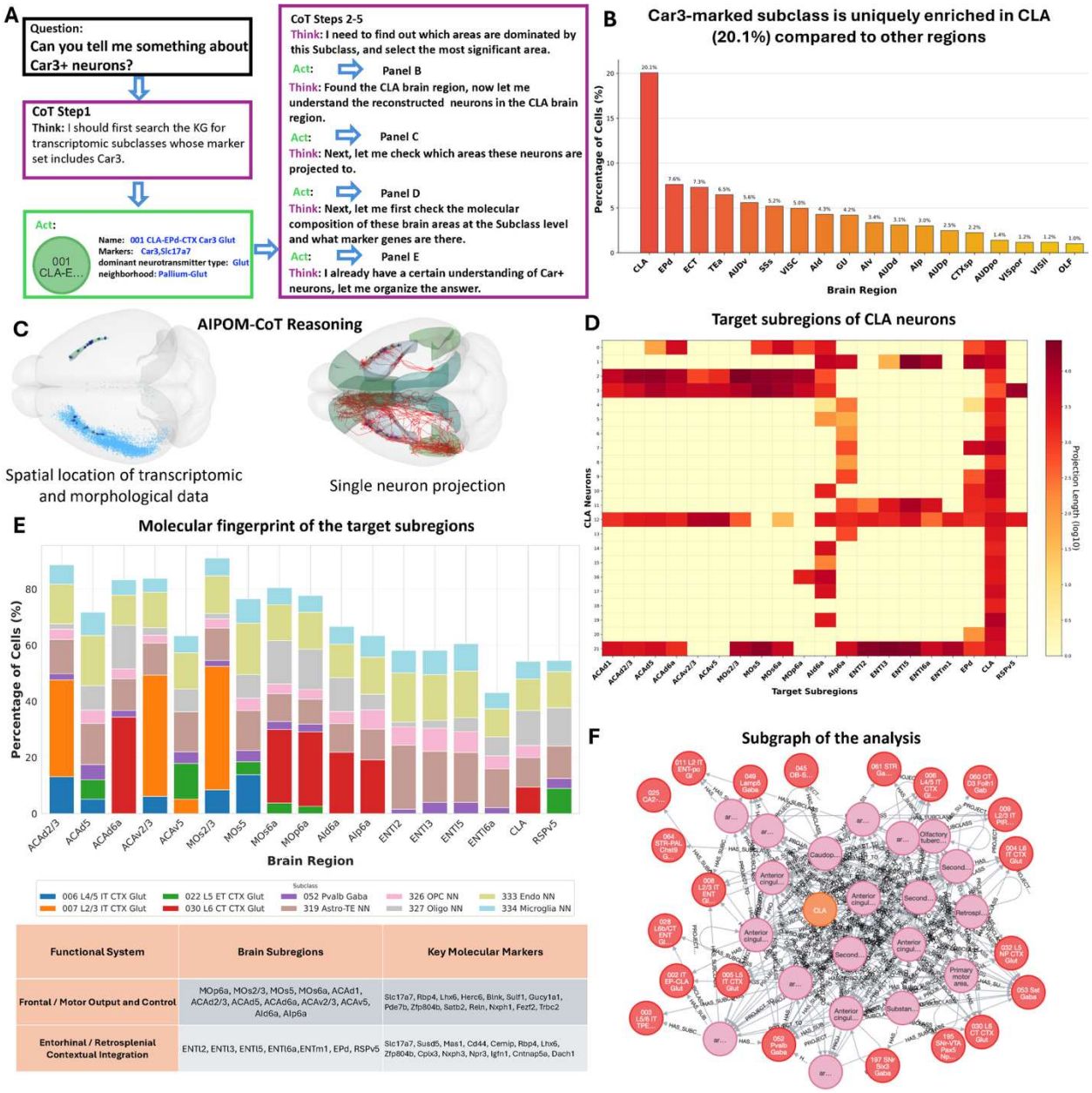
AIPOM-CoT first queried the knowledge graph to identify transcriptomic subclasses expressing *Car3*, identifying the "003 CLA-EPd-CTX Car3 Glut" subclass as the dominant *Car3*-expressing population. The agent then automatically ranked brain regions by their enrichment for this subclass, revealing that the claustrum (CLA) contains approximately 20.1% of cells belonging to this subclass, substantially higher than any other region examined (**Figure 6B and Supplementary Figure 8**). Importantly, this enrichment was identified through automated retrieval and quantitative analysis across the entire knowledge graph, rather than through hypothesis-driven region selection or manual literature review.

Having identified CLA as the region with maximal enrichment of the *Car3* subclass, AIPOM-CoT next retrieved morphological reconstructions of CLA neurons from the knowledge graph (**Figure 6C, left and Supplementary Figure 6**). The system then automatically extracted projection information from these reconstructions, identifying the target subregions innervated by CLA neurons (**Figure 6C, right**; **Figure 6D and Supplementary Figures 7-8**). This analysis revealed a heterogeneous projection pattern, with CLA neurons targeting multiple cortical areas with varying strengths, including frontal, motor, and sensory regions, as well as selected subcortical structures.

To complete the multimodal profile, AIPOM-CoT automatically generated molecular fingerprints for each CLA projection target by querying the cell type composition of the target regions (**Figure 6E and Supplementary Figures 7-8**). This analysis revealed substantial molecular diversity among regions receiving CLA input, with some targets dominated by specific cortical layer markers and others enriched for particular GABAergic subtypes. The system organized these results into an interpretable summary, including a knowledge graph visualization that explicitly links *Car3* neurons, their projection targets, and the molecular characteristics of those targets (**Figure 6F and Supplementary Figure 6**).

Although the primary contribution of this demonstration is the automated retrieval-to-integration capability itself, the analysis also produced an integrated multimodal profile of *Car3*+ claustral neurons. The strong enrichment of *Car3*+ cells in the claustrum, together with their widespread cortical and selective subcortical projections, is consistent with proposed roles of the claustrum in large-scale cortical coordination and cross-modal integration. The molecular heterogeneity of projection targets further suggests that *Car3*+ neurons may differentially influence distinct cortical processing streams. However, testing these hypotheses requires targeted experimental validation beyond the scope of automated knowledge graph analysis.

All steps from the initial question to the final integrated summary were executed automatically and completed in approximately 60 seconds, in contrast to the hours or days typically required for manual cross-modal integration. The analysis includes full provenance tracking (**Supplementary Figure 8**), recording the exact knowledge graph queries, data sources, and computational parameters at each stage, thereby enabling independent validation and iterative refinement. Importantly, the central advance illustrated here extends beyond the specific biological observations to the system's ability to assemble comprehensive multimodal neuronal profiles on demand from arbitrary natural language queries, without manual data curation or predefined workflows. Together, these capabilities establish a new paradigm for automated, multi-step cross-modal analysis spanning molecular, morphological, and connectivity data, with end-to-end traceability to support reproducibility.



Functional System	Brain Subregions	Key Molecular Markers
Frontal / Motor Output and Control	MOp6a, MOs2/3, MOs5, MOs6a, ACAd1, ACAd2/3, ACAd5, ACAd6a, ACAv2/3, ACAv5, Ald6a, Alp6a	Slc17a7, Rbp4, Lhx6, Herc6, Bink, Sulf1, Gucy1a1, Pde7b, Zfp804b, Satb2, Reln, Nxph1, Fezf2, Trbc2
Entorhinal / Retrosplenial Contextual Integration	ENTl2, ENTl3, ENTl5, ENTl6a,ENTm1, EPd, RSPv5	Slc17a7, Susd5, Mas1, Cd44, Cemip, Rbp4, Lhx6, Zfp804b, Cplx3, Nxph3, Npr3, Igfn1, Cntnap5e, Dach1

Figure 6*. Automated multi-modality profiling of Car3+ neurons from natural language query.* ***A.*** *AIPOM-CoT decomposes "Tell me about Car3+ neurons" into executable workflow: marker identification → regional enrichment ranking → morphology retrieval → projection mapping → target profiling.* ***B.*** *Automated statistical analysis*

discovers claustrum (CLA) as maximally enriched for Car3+ subclasses (~20.1%), far exceeding other regions. ***C.*** *Spatial co-registration: transcriptomic Car3 expression (left) aligned with CLA neuron morphology (right) in common atlas coordinates.* ***D.*** *Projection targets automatically aggregated from CLA neurons, revealing consistent innervation of cortical (MOs/ACAd L2/3/L5, AI) and entorhinal (ENTl) regions.* ***E.*** *Automated molecular fingerprints of targets partition into frontal-motor system (IT-excitatory dominated) and entorhinal-retrosplenial system (mixed IT signatures).* ***F.*** *Complete execution provenance enables reproducible replay.*

Discussion

Conceptually, this work shows that transcriptomic, morphological, and connectivity features define semi-independent organizational axes of the brain, necessitating explicit multimodal integration rather than single-modality analysis. To that end, we developed NeuroXiv-KG and AIPOM-CoT to address the fragmentation of neuroscience data across molecular, morphological, and connectivity domains. NeuroXiv-KG provides a unified substrate in the form of a typed, cross-modal property graph comprising 258 million relationships, enabling direct traversal from molecular markers to anatomical regions, neuronal morphology, and projection targets. Building on this foundation, AIPOM-CoT translates natural language questions into multi-step analytical workflows that combine structured graph traversal with statistical computation, including effect sizes, confidence intervals, and permutation tests with false discovery rate correction, while maintaining full provenance (**Supplementary Figure 8**) through recorded queries, tool-calling sequences, random seeds, and data snapshots. Together, these components transform atlas interaction from manual browsing into planned, auditable multimodal analysis.

One broadly important finding is that the cross-modal divergence observed here has conceptual implications for brain organization. Molecular taxonomies alone do not fully predict neuronal morphology or connectivity, while connectivity-centric descriptions that ignore molecular diversity risk oversimplifying functional organization. A comprehensive understanding of brain structure therefore requires explicit multimodal integration that treats molecular, structural, and connectional data as complementary rather than redundant. Knowledge graphs provide the representational substrate for such integration, and automated reasoning agents supply the analytical machinery to explore these relationships systematically.

In a few specifically shown applications, the *Car3*+ neuron case study illustrates this capability in practice. From an open-ended prompt, AIPOM-CoT autonomously identified *Car3*-expressing transcriptomic subclasses, quantified regional enrichment in the claustrum, retrieved morphological reconstructions, mapped projection targets, and generated molecular fingerprints of target regions. This end-to-end analysis was completed in approximately one minute, compared with the conventionally required days of work in manual cross-modal integration. More broadly, whole-brain tri-modal fingerprinting revealed that approximately 40% of region pairs exhibit systematic cross-modal divergence. Morphological and projection features showed modular organization, whereas molecular composition was largely region-specific. Representative cases demonstrated distinct mechanisms underlying this divergence: LHA and TU shared morphological organization despite molecular divergence, consistent with shared computational demands and distinct developmental origins, while MOs and LHA converged on projection targets despite molecular divergence, supporting functional integration between cortical motor control and hypothalamic motivational systems. Together, these results indicate that molecular, morphological, and connectivity features define semi-independent organizational axes of brain structure.

Within the broader neuroscience ecosystem, our platform fills a distinct gap. Existing resources provide key components, such as curated atlases, data infrastructure, imaging pipelines, or general-purpose agentic planning, but none integrate these elements into a single system that supports automated, cross-modal analysis with statistical auditing and reproducibility. NeuroXiv-KG and AIPOM-CoT address this gap by combining purpose-built graph representations, natural language queries compiled into explicit tool-calling chains, built-in statistical discovery, and full provenance tracking (**Supplementary Figure 8**).

Benchmarking demonstrates that this integration yields consistent improvements in accuracy, interpretability, and audit completeness over graph-only pipelines, retrieval-augmented language models, and manual workflows, particularly for complex, multi-step analyses. Together, these capabilities establish a unified platform for automated and auditable cross-modal neuroscience, enabling adaptive analytical workflows that scale with question complexity while preserving reproducibility and supporting systematic discovery from large multimodal brain datasets. It is important to note that this work enables analyses that were previously impractical, including on-demand, whole-brain cross-modal queries with statistical validation and full provenance (**Supplementary Figure 8**), without manual data curation.

Several limitations remain. Region-level aggregation can obscure single-cell heterogeneity, the current planning module has limited stochasticity, comprehensive permutation testing is computationally intensive, and finally the current coverage of data is restricted to the adult mouse brain. Future extensions will incorporate cell-type-resolved connectivity where available, introduce calibrated uncertainty estimation, formalize exemplar selection strategies, expand benchmarking with expert-graded task suites, extend the tool function library to include causal inference primitives, and explore integration with experimental platforms and even data of other species to support hypothesis generation.

Data availability

NeuroXiv2 provides downloadable resources covering atlas-registered neuronal morphology data, molecular data, and derived findings (e.g., associated metadata, visualizations, and text-mining outputs). These materials can be accessed through the NeuroXiv2 portal (https://neuroxiv.org/v2), with a mirrored distribution endpoint hosted on AWS for standardized releases (https://download.neuroxiv.org). All files are distributed under the Creative Commons Attribution–NonCommercial 4.0 International license (CC BY-NC 4.0; https://creativecommons.org/licenses/by-nc/4.0/legalcode) and are intended for non-commercial scholarly use only. Users are required to maintain provenance (including proper citation/attribution) and must not republish the content in ways that obscure origin or enable academic misconduct.

Code availability

The NeuroXiv2 codebase is publicly hosted on GitHub (https://github.com/SEU-ALLEN-codebase/NeuroXiv2). The repository is structured into three primary components, namely NeuroXiv2-KG, AIPOM-CoT, and Data-Analysis. NeuroXiv2-KG contains the code used for knowledge-graph construction and supporting utilities. AIPOM-CoT provides the implementation of the agent. Data-Analysis includes the scripts that reproduce the main analyses and results reported in this study. The entire codebase is released under the MIT license.

Acknowledgements

This work was mainly supported by several grants awarded to H.P., who is a New Cornerstone Investigator and a Shanghai Academy of Natural Sciences (SANS) Senior Investigator. We thank Shengdian Jiang, Zhixi Yun, Longxiao Yuan, and other members in Peng Lab for contribution of early development and discussion of this work.

Author contributions

H.P. and F.L. conceptualized this study. H.P. managed this work and instructed the detailed development of methods and experiments. L.W. developed the database and the AI platform, performed data analyses.

F.L. contributed ideas of AI-based open mining for large public databases. L.W. and H.P. wrote the manuscript with input from all coauthors.

Declaration of Interests

The authors declare no competing interests.

References

1. Ament, S. A., Adkins, R. S., Carter, R., Chrysostomou, E., Colantuoni, C., Crabtree, J., Creasy, H. H., Degatano, K., Felix, V., Gandt, P., Garden, Gwenn A., Giglio, M., Herb, B. R., Khajouei, F., Kiernan, E., McCracken, C., McDaniel, K., Nadendla, S., Nickel, L.,…White, Owen R. (2022). The Neuroscience Multi-Omic Archive: a BRAIN Initiative resource for single-cell transcriptomic and epigenomic data from the mammalian brain. *Nucleic Acids Research*, *51*(D1), D1075–D1085. https://doi.org/10.1093/nar/gkac962
2. Aransay, A., Rodríguez-López, C., García-Amado, M., Clascá, F., & Prensa, L. (2015). Long-range projection neurons of the mouse ventral tegmental area: a single-cell axon tracing analysis. *Frontiers in neuroanatomy*, *9*, 59.
3. Badenes-Olmedo, C., & Corcho, O. (2024). MuHeQA: Zero-shot question answering over multiple and heterogeneous knowledge bases. *Semantic Web*, *15*(5), 1547–1561.
4. Bria, A., Iannello, G., Onofri, L., & Peng, H. (2016). TeraFly: real-time three-dimensional visualization and annotation of terabytes of multidimensional volumetric images. *Nature Methods*, *13*(3), 192–194.
5. Feng, Z., Li, X., Luo, Y., Liu, X., Long, B., Jiang, T., Jia, X., Chen, X., Luo, J., Chai, X., Wang, Z., Ren, M., Lu, X., Yao, G., Zhao, M., Li, Y., Liu, Z., Ni, H., Dou, C.,…Luo, Q. (2025). A mouse brain stereotaxic topographic atlas with isotropic 1-μm resolution. *Nature*, *645*(8080), 448–456. https://doi.org/10.1038/s41586-025-09211-8
6. Francis, N., Green, A., Guagliardo, P., Libkin, L., Lindaaker, T., Marsault, V., Plantikow, S., Rydberg, M., Selmer, P., & Taylor, A. (2018). Cypher: An evolving query language for property graphs. Proceedings of the 2018 international conference on management of data,
7. Gao, L., Liu, S., Gou, L., Hu, Y., Liu, Y., Deng, L., Ma, D., Wang, H., Yang, Q., Chen, Z., Liu, D., Qiu, S., Wang, X., Wang, D., Wang, X., Ren, B., Liu, Q., Chen, T., Shi, X.,…Yan, J. (2022). Single-neuron projectome of mouse prefrontal cortex. *Nature Neuroscience*, *25*(4), 515–529. https://doi.org/10.1038/s41593-022-01041-5
8. Gao, L., Liu, S., Wang, Y., Wu, Q., Gou, L., & Yan, J. (2023). Single-neuron analysis of dendrites and axons reveals the network organization in mouse prefrontal cortex. *Nature Neuroscience*, *26*(6), 1111–1126. https://doi.org/10.1038/s41593-023-01339-y
9. Gong, H., Xu, D., Yuan, J., Li, X., Guo, C., Peng, J., Li, Y., Schwarz, L. A., Li, A., & Hu, B. (2016). High-throughput dual-colour precision imaging for brain-wide connectome with cytoarchitectonic landmarks at the cellular level. *Nature Communications*, *7*(1), 12142.
10. Gouwens, N. W., Sorensen, S. A., Berg, J., Lee, C., Jarsky, T., Ting, J., Sunkin, S. M., Feng, D., Anastassiou, C. A., Barkan, E., Bickley, K., Blesie, N., Braun, T., Brouner, K., Budzillo, A., Caldejon, S., Casper, T., Castelli, D., Chong, P.,…Koch, C. (2019). Classification of electrophysiological and morphological neuron types in the mouse visual cortex. *Nature Neuroscience*, *22*(7), 1182–1195. https://doi.org/10.1038/s41593-019-0417-0
11. Guu, K., Lee, K., Tung, Z., Pasupat, P., & Chang, M. (2020). Retrieval augmented language model pre-training. International conference on machine learning,
12. Hawrylycz, M., Martone, M. E., Ascoli, G. A., Bjaalie, J. G., Dong, H. W., Ghosh, S. S., Gillis, J., Hertzano, R., Haynor, D. R., Hof, P. R., Kim, Y., Lein, E., Liu, Y., Miller, J. A., Mitra, P. P., Mukamel, E., Ng, L., Osumi-Sutherland, D., Peng, H.,…Zingg, B. (2023). A guide to the BRAIN

Initiative Cell Census Network data ecosystem. *PLoS Biol*, *21*(6), e3002133. https://doi.org/10.1371/journal.pbio.3002133

13. Jiang, S., Wang, L., Yun, Z., Chen, H., Liu, L., Yao, J., & Peng, H. (2025). NeuroXiv: AI-powered open databasing and dynamic mining of brain-wide neuron morphometry. *Nature Methods*, *22*(6), 1195–1198. https://doi.org/10.1038/s41592-025-02687-2
14. Jiao, Z., Gao, T., Wang, X., Wang, A., Ma, Y., Feng, L., Gao, L., Gou, L., Zhang, W., Biglari, N., Boxer, E. E., Steuernagel, L., Ding, X., Yu, Z., Li, M., Gao, M., Hao, M., Zhou, H., Cao, X.,…Xu, X. (2025). Projectome-based characterization of hypothalamic peptidergic neurons in male mice. *Nature Neuroscience*, *28*(5), 1073–1088. https://doi.org/10.1038/s41593-025-01919-0
15. Koster, K. P., & Sherman, S. M. (2024). Convergence of inputs from the basal ganglia with layer 5 of motor cortex and cerebellum in mouse motor thalamus. *Elife*, *13*, e97489. https://doi.org/10.7554/eLife.97489
16. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., & Rocktäschel, T. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in neural information processing systems*, *33*, 9459–9474.
17. Li, Y., Wu, J., Lu, D., Xu, C., Zheng, Y., Peng, H., & Qu, L. (2022). mBrainAligner-Web: a web server for cross-modal coherent registration of whole mouse brains. *Bioinformatics*, *38*(19), 4654–4655. https://doi.org/10.1093/bioinformatics/btac549
18. Liu, L., Yun, Z., Manubens-Gil, L., Chen, H., Xiong, F., Dong, H., Zeng, H., Hawrylycz, M., Ascoli, G. A., & Peng, H. (2025). Connectivity of single neurons classifies cell subtypes in mouse brains. *Nature Methods*, *22*(4), 861–873. https://doi.org/10.1038/s41592-025-02621-6
19. Liu, Y., Jiang, S., Li, Y., Zhao, S., Yun, Z., Zhao, Z.-H., Zhang, L., Wang, G., Chen, X., Manubens-Gil, L., Hang, Y., Gong, Q., Li, Y., Qian, P., Qu, L., Garcia-Forn, M., Wang, W., De Rubeis, S., Wu, Z.,…Peng, H. (2024). Neuronal diversity and stereotypy at multiple scales through whole brain morphometry. *Nature Communications*, *15*(1), 10269. https://doi.org/10.1038/s41467-024-54745-6
20. Liu, Y., Zhao, S., Yun, Z., Xiong, F., & Peng, H. (2025). A mouse brain atlas based on dendritic microenvironments. *Nature Neuroscience*. https://doi.org/10.1038/s41593-025-02119-6
21. Lovett-Barron, M., Chen, R., Bradbury, S., Andalman, A. S., Wagle, M., Guo, S., & Deisseroth, K. (2020). Multiple convergent hypothalamus–brainstem circuits drive defensive behavior. *Nature Neuroscience*, *23*(8), 959–967. https://doi.org/10.1038/s41593-020-0655-1
22. Oh, S. W., Harris, J. A., Ng, L., Winslow, B., Cain, N., Mihalas, S., Wang, Q., Lau, C., Kuan, L., Henry, A. M., Mortrud, M. T., Ouellette, B., Nguyen, T. N., Sorensen, S. A., Slaughterbeck, C. R., Wakeman, W., Li, Y., Feng, D., Ho, A.,…Zeng, H. (2014). A mesoscale connectome of the mouse brain. *Nature*, *508*(7495), 207–214. https://doi.org/10.1038/nature13186
23. OpenAI. (2024). *Hello GPT-4o*. OpenAI. Retrieved 2026-02-09 from https://openai.com/index/hello-gpt-4o/
24. Peng, H., Xie, P., Liu, L., Kuang, X., Wang, Y., Qu, L., Gong, H., Jiang, S., Li, A., Ruan, Z., Ding, L., Yao, Z., Chen, C., Chen, M., Daigle, T. L., Dalley, R., Ding, Z., Duan, Y., Feiner, A.,…Zeng, H. (2021). Morphological diversity of single neurons in molecularly defined cell types. *Nature*, *598*(7879), 174–181. https://doi.org/10.1038/s41586-021-03941-1
25. Qiu, S., Hu, Y., Huang, Y., Gao, T., Wang, X., Wang, D., Ren, B., Shi, X., Chen, Y., Wang, X., Wang, D., Han, L., Liang, Y., Liu, D., Liu, Q., Deng, L., Chen, Z., Zhan, L., Chen, T.,…Xu, C. (2024). Whole-brain spatial organization of hippocampal single-neuron projectomes. *Science*, *383*(6682), eadj9198. https://doi.org/doi:10.1126/science.adj9198
26. Qu, L., Li, Y., Xie, P., Liu, L., Wang, Y., Wu, J., Liu, Y., Wang, T., Li, L., Guo, K., Wan, W., Ouyang, L., Xiong, F., Kolstad, A. C., Wu, Z., Xu, F., Zheng, Y., Gong, H., Luo, Q.,…Peng, H. (2022). Cross-modal coherent registration of whole mouse brains. *Nat Methods*, *19*(1), 111–118. https://doi.org/10.1038/s41592-021-01334-w
27. Ragan, T., Kadiri, L. R., Venkataraju, K. U., Bahlmann, K., Sutin, J., Taranda, J., Arganda-Carreras, I., Kim, Y., Seung, H. S., & Osten, P. (2012). Serial two-photon tomography for automated ex vivo mouse brain imaging. *Nature Methods*, *9*(3), 255–258. https://doi.org/10.1038/nmeth.1854

28. Wang, Q., Ding, S.-L., Li, Y., Royall, J., Feng, D., Lesnar, P., Graddis, N., Naeemi, M., Facer, B., Ho, A., Dolbeare, T., Blanchard, B., Dee, N., Wakeman, W., Hirokawa, K. E., Szafer, A., Sunkin, S. M., Oh, S. W., Bernard, A.,…Ng, L. (2020). The Allen Mouse Brain Common Coordinate Framework: A 3D Reference Atlas. *Cell*, *181*(4), 936–953.e920. https://doi.org/https://doi.org/10.1016/j.cell.2020.04.007
29. Winnubst, J., Bas, E., Ferreira, T. A., Wu, Z., Economo, M. N., Edson, P., Arthur, B. J., Bruns, C., Rokicki, K., Schauder, D., Olbris, D. J., Murphy, S. D., Ackerman, D. G., Arshadi, C., Baldwin, P., Blake, R., Elsayed, A., Hasan, M., Ramirez, D.,…Chandrashekar, J. (2019). Reconstruction of 1,000 Projection Neurons Reveals New Cell Types and Organization of Long-Range Connectivity in the Mouse Brain. *Cell*, *179*(1), 268–281.e213. https://doi.org/https://doi.org/10.1016/j.cell.2019.07.042
30. Xiong, F., Liu, L., & Peng, H. (2025). Reconstruction of a connectome of single neurons in mouse brains by cross-validating multi-scale multi-modality data. *Nature Methods*, *22*(12), 2670–2683. https://doi.org/10.1038/s41592-025-02784-2
31. Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2022). ReAct: Synergizing Reasoning and Acting in Language Models. *arXiv e-prints*, arXiv:2210.03629. https://doi.org/10.48550/arXiv.2210.03629
32. Yao, Z., van Velthoven, C. T. J., Kunst, M., Zhang, M., McMillen, D., Lee, C., Jung, W., Goldy, J., Abdelhak, A., Aitken, M., Baker, K., Baker, P., Barkan, E., Bertagnolli, D., Bhandiwad, A., Bielstein, C., Bishwakarma, P., Campos, J., Carey, D.,…Zeng, H. (2023). A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain. *Nature*, *624*(7991), 317–332. https://doi.org/10.1038/s41586-023-06812-z
33. Yao, Z., van Velthoven, C. T. J., Nguyen, T. N., Goldy, J., Sedeno-Cortes, A. E., Baftizadeh, F., Bertagnolli, D., Casper, T., Chiang, M., Crichton, K., Ding, S.-L., Fong, O., Garren, E., Glandon, A., Gouwens, N. W., Gray, J., Graybuck, L. T., Hawrylycz, M. J., Hirschstein, D.,…Zeng, H. (2021). A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell*, *184*(12), 3222–3241.e3226. https://doi.org/https://doi.org/10.1016/j.cell.2021.04.021
34. Zeng, H. (2022). What is a cell type and how to define it? *Cell*, *185*(15), 2739–2755. https://doi.org/https://doi.org/10.1016/j.cell.2022.06.031
35. Zeng, H., & Sanes, J. R. (2017). Neuronal cell-type classification: challenges, opportunities and the path forward. *Nature Reviews Neuroscience*, *18*(9), 530–546. https://doi.org/10.1038/nrn.2017.85
36. Zhang, M., Pan, X., Jung, W., Halpern, A. R., Eichhorn, S. W., Lei, Z., Cohen, L., Smith, K. A., Tasic, B., Yao, Z., Zeng, H., & Zhuang, X. (2023). Molecularly defined and spatially resolved cell atlas of the whole mouse brain. *Nature*, *624*(7991), 343–354. https://doi.org/10.1038/s41586-023-06808-9

Methods

Datasets

NeuroXiv 2.0 substantially extends the original platform by integrating spatial transcriptomic data with existing morphological reconstructions, enabling systematic cross-modal analysis of cellular architecture and molecular profiles. Building upon the morphological datasets previously described in NeuroXiv (Jiang et al., 2025), we have incorporated MERFISH whole mouse brain spatial transcriptomics datasets and expanded the morphological collection with hypothalamic peptidergic neurons.

Morphology Datasets

In addition to previously incorporated datasets (Jiang et al., 2025; Liu et al., 2024; Peng et al., 2021), we additionally integrated a hypothalamus dataset that captures single-cell projection patterns of peptidergic neurons (Jiao et al., 2025). These reconstructions were obtained from the Brain Science Data Center of the

Chinese Academy of Sciences (Wang, X. and Jiao, Z. Single-cell projection atlas of mouse hypothalamic peptidergic neurons. https://cstr.cn/33145.11.BSDC.1752068512.1942934383845957634, 2025). The dataset provides comprehensive axonal architectures that are particularly valuable for characterizing hypothalamic circuit organization within a unified atlas reference.

Spatial transcriptomics Datasets

NeuroXiv2 incorporates MERFISH-based spatial transcriptomics as the molecular layer for spatially resolved cell-type annotation and cross-modal integration with atlas-registered morphology. The MERFISH resource used in this study is publicly accessible through multiple official distribution channels, including the Brain Image Library, which provides raw and processed data together with codebook and probe information, the Allen Brain Cell Atlas interactive browser at https://knowledge.brain-map.org/data/5C0201JSVE04WY6DMVC/explore with accompanying dataset documentation at

- https://alleninstitute.github.io/abc_atlas_access/descriptions/Zhuang-ABCA-1.html,
- https://alleninstitute.github.io/abc_atlas_access/descriptions/Zhuang-ABCA-2.html,
- https://alleninstitute.github.io/abc_atlas_access/descriptions/Zhuang-ABCA-3.html,
- https://alleninstitute.github.io/abc_atlas_access/descriptions/Zhuang-ABCA-4.html,

and the CELLxGENE collection at

- https://cellxgene.cziscience.com/collections/0cca8620-8dee-45d0-aef5-23f032a5cf09.

These sources collectively provide the molecular measurements and supporting metadata that enable standardized integration into NeuroXiv2.

Connectivity datasets

NeuroXiv2 provides connectivity datasets that are derived from the integrated morphology datasets after standardization and atlas registration. The Connectivity data between neurons, between neurons and brain regions, and between brain regions is obtained by quantifying the distances between neurons and the projection strengths from neurons to brain regions under a common coordinate framework. These connectivity datasets are generated systematically from the morphology resources described above and are distributed as derived datasets to support search, comparison, and downstream analyses.

Multi-modality data aggregation and integration

NeuroXiv2 follows the data aggregation workflow established in NeuroXiv(Jiang et al., 2025) to convert morphology resources from SEU-ALLEN and third-party sources into a consolidated format. Briefly, reconstructions are harmonized into the SWC format to ensure consistent encoding of soma and neurite compartments, and the SEU-ALLEN datasets are already available as standardized SWC files that have been registered to the Allen Mouse Brain Common Coordinate Framework version 3(Wang et al., 2020). For external resources, ION datasets are converted into the SWC format, aligning the structure domain types, while MouseLight reconstructions are transformed from their released JSON representation into SWC files. Each reconstruction is then screened with structural quality criteria to ensure data usability. Accepted neurons are mapped to the CCFv3 atlas, using mBrainAligner(Li et al., 2022; Qu et al., 2022), and the atlas-oriented reconstructions are resampled to ensure that the distance between parent and child nodes was set to 1 μm. After atlas mapping, neurons are curated with standardized identifiers, and basic information such as the soma region are computed by assigning soma location and quantifying axonal and dendritic arborization strength using neurite length. Morphology features and connectivity relationships are further computed to support retrieval, comparison and analysis.

Multi-modality data integration in NeuroXiv2 is achieved by anchoring morphology data, spatial transcriptomics data, and connectivity data to shared anatomical coordinates. Morphological reconstructions are aligned to CCFv3 through registration with manual verification, and spatial transcriptomic measurements are placed into the same reference space using transformation matrices released by the Allen Institute(Yao et al., 2023). The connectivity data is obtained by quantifying the spatial distances between neurons registered in the atlas and the projection strengths from the neurons to brain regions. In addition to the CCFv3 atlas, NeuroXiv2 supports interpretation within the morphology informed microenvironment atlas CCF-ME(Y. Liu et al., 2025) so that brain region level and micro-environment level analyses can be conducted consistently across modalities.

Knowledge-graph construction

Using all the multi-modality data in NeuroXiv2, we built a knowledge-graph namely NeuroXiv-KG and implemented it in Neo4j Community Edition version 5.22.0. NeuroXiv-KG is deployed in a docker container using docker version 28.3.2. The graph schema was designed to support unified queries across anatomy, morphology, spatial transcriptomics, and connectivity.

The schema of NeuroXiv-KG was defined after all multi-modality data were anchored to the shared anatomical coordinates. We represented anatomical structures using CCFv3 brain region nodes (Region) and their standard subdivisions (Subregion), and we additionally included microenvironment parcellations (ME-Subregion) from the CCF-ME atlas to increase spatial specificity for multi-modality data aligned analyses. We represented individual reconstructed neurons as Neuron nodes and linked them to their atlas assigned locations through spatial localization relations at the brain region, subregion, and microenvironment levels. For the molecular data, we represented MERFISH mapped cell types as transcriptomic tier nodes organized into four hierarchical levels (Class, Subclass, Supertype and Cluster), and we encoded the hierarchy so that higher level labels can be traversed to lower-level clusters within the same graph. The node set therefore enables integrated navigation from anatomical units to neurons and from anatomical units to transcriptomic classes, as well as bridging between these two modalities through shared spatial anchoring.

We encoded multi-modality associations as typed relations that capture spatial containment, hierarchical organization, connectivity and local morphological adjacency. Spatial localization relations associate each neuron with its assigned region, subregion, and me-subregion. Projection relations represent axonal connectivity and store normalized weights that quantify projection strength in atlas space. Taxonomic relations connect transcriptomic tiers across the hierarchy and associate cells with their corresponding tier labels, enabling queries that aggregate molecular composition within anatomical units. To support local circuit context, we added relations that represent axonal neighboring and dendritic neighboring pairs derived from spatial proximity and local arbor overlap criteria computed in atlas space during data processing.

We store key measurements, feature values, and other computed properties in the form of graph attributes to enable efficient cross modal queries without repeated computation. For morphology, we computed morphological features of per neuron that include axonal and dendritic length, branch counts, maximum branch order and so on and aggregated these features to anatomical units to produce region resolved profiles. For connectivity, we harmonized projection information across sources and attached weighted projection relations to anatomical source target pairs to enable comparisons across datasets in a consistent reference space. For transcriptomics, we computed anatomical compositions as proportions across the Class, Subclass, Supertype, and Cluster levels by aggregating MERFISH mapped cells within each anatomical unit. These summaries were stored as node or relation properties that can be accessed directly during KG traversal.

AIPOM-CoT

AIPOM-CoT is an agentic analysis system that interfaces with NeuroXiv KG to answer free text requests using an explicit iterative reasoning and execution procedure. For each user prompt, AIPOM-CoT runs a structured cycle that combines intent understanding, executable planning, graph-based action, and evidence driven self-evaluation. This loop is designed to make multi-modality analyses reproducible and to ensure that the final response is supported by traceable evidence from the knowledge graph.

Think–Plan–Act–Reflect Loop

For a free-text prompt, AIPOM-CoT executes a four-stage iterative cycle:

1. In the **Think** stage, AIPOM-CoT interprets the user prompt and converts it into a structured analysis objective. It identifies which modalities are required for the request, including neuronal morphology, circuit connectivity and molecular profiles. It also determines the spatial scope, such as a single region, a region-to-region comparison, or a whole brain survey, and it estimates the level of analytical depth that will be needed. The output of Think is a statement of data requirements and an expected workflow complexity that guides subsequent planning and execution.
2. In the **Plan** stage, AIPOM-CoT inspects the current schema of NeuroXiv-KG and constructs an executable task graph. The agent decomposes the objective into steps that each step serves a biological purpose and has explicit dependencies on previous steps. For example, a typical plan begins with defining the neuron population of interest, then derives molecular enrichment or transcriptomic composition, retrieves morphological features, and finally summarizes projection patterns across atlas regions when connectivity is relevant. For each step, the plan specifies the tool function class to be used, the expected input and output data structures, and any constraints such as minimum sample size or required modality coverage.
3. In the **Act** stage, AIPOM-CoT translates the executable plan into parameterized Cypher queries over the knowledge graph (**Supplementary Figure 5**). It then uses these queries to drive a tool call chain that performs the required statistics, visualization steps, and so on, and it returns structured outputs together with the metadata needed for downstream evaluation and replay.
4. In the **Reflect** stage, AIPOM-CoT assesses whether the accumulated evidence is sufficient to support a conclusion and whether additional analysis is warranted. It updates a structured evidence buffer (**Supplementary Figure 8**) that records data completeness, modality coverage, and statistical support, together with quantitative summaries that include effect sizes, confidence intervals, and so on used for each comparison. AIPOM-CoT then computes an overall confidence score by integrating evidence quality, cross modality coverage, and statistical support, and it determines whether to continue, replan, deepen, pivot, or terminate. Termination is triggered when predefined confidence, coverage, and stability criteria are met, and the termination decision and its reason are recorded in the provenance trace for replay and audit.

The tool function library

AIPOM-CoT executes analyses through a reusable tool function library that provides composable primitives for retrieval, computation, statistical testing, and visualization. Each tool function exposes a standardized interface that specifies required inputs, produced outputs, and run descriptors that capture the parameters used for the call. The tool functions are designed to be chained by the Act module so that multi step workflows can be executed consistently across different query intents and datasets.

The retrieval layer provides tool functions for graph search and schema aware traversal. These tools support filtering and ranking by node properties, transcriptomic composition, and projection weights, as well as constrained path queries that traverse predefined relation types to connect anatomical units, neurons, and

transcriptomic tiers. The aggregation layer provides region resolved summarization tools that compute modality specific summaries across neurons or cells within an anatomical unit.

For statistical analysis, the library includes enrichment tests that assess overrepresentation in a target set using a hypergeometric test with false discovery rate correction. It also includes similarity and divergence tools that compare multimodal fingerprints using cosine similarity, Jensen Shannon divergence, or Euclidean distance, and correlation tools that compute Spearman correlations and optionally adjust for covariates through residualization.

AIPOM-CoT performance evaluation

Benchmark design and question Set

We designed a controlled benchmark to evaluate how different systems support knowledge graph driven neuroscience analysis under free text prompting. The benchmark consisted of 23 queries that were selected to cover a range of analytical demands and to match the task categories reported in Results 3. Queries were grouped into three complexity tiers based on the number of required reasoning steps and the extent of multimodal integration. The simple tier contained five queries that can be answered by a single step graph lookup or a direct traversal in the knowledge graph. The medium tier contained eleven queries that require coordinated retrieval across more than one modality and typically involve region level comparisons or cross-referenced evidence. The deep tier contained seven queries that require systematic whole brain integration and explicit statistical validation across regions. Each query was annotated in advance with its expected target entities, the intended analytical depth, the modalities required for a complete answer, and predefined success criteria that specify what constitutes a correct and sufficiently supported response.

Baseline Methods

We compared AIPOM CoT against four baseline approaches that represent common strategies for natural language access to structured resources. The first baseline was direct prompting with GPT 4o, in which the model generated an answer end to end without specialized tool abstractions. The second baseline was a template driven knowledge graph method that relied on predefined Cypher query templates for common question patterns. The third baseline was a retrieval augmented generation pipeline that retrieved relevant context using embedding based search and then used GPT 4o to synthesize a response. The fourth baseline was a reasoning and acting approach namely ReAct that alternated between natural language reasoning and tool execution for up to ten iterations. All methods used the same GPT 4o backbone and the same Neo4j execution environment so that observed differences reflected the method design rather than differences in model capacity or database access.

Evaluation Metrics

We evaluated each method along four axes that correspond to Results 3. Planning quality was scored by coherence, efficiency, and adaptability. Reasoning capability was scored by logical consistency and evidence integration, where evidence integration was computed as the fraction of applicable modalities among molecular, morphological, and projection information that were actually retrieved and used. Self-correction capability was scored by error detection, correction implementation, and iterative refinement, with insufficiency flagged when results were empty, when $p>0.1$, when confidence intervals exceeded 50% of the effect size, or when modality coverage was below two thirds of the applicable modalities. Query understanding was scored by matching the inferred intent and required modalities to the benchmark annotations.

All sub metrics were scored per query on a 0 to 1 scale using a predefined rubric and were averaged across questions within each tier and overall. Composite scores were computed as the mean of sub metrics within each axis, with planning quality defined as the mean of coherence, efficiency, and adaptability, reasoning capability defined as the mean of logical consistency and evidence integration, and self-correction capability defined as the mean of error detection, correction implementation, and iterative refinement.

Multi-modality fingerprints representation for brain regions

For each region r, we constructed modality specific fingerprints for molecular composition, morphology, and projection. These fingerprints define a shared representation space for cross modal similarity and divergence analyses and were used in the data analysis of **Figure 4** and **Figure 5**.

We defined the molecular fingerprint as $g_r \in R^{K_G}$. For each region, we computed transcriptomic subclass composition vectors and removed dimensions that were zero across all regions to avoid redundant features. The remaining features were standardized by z score normalization across regions. We then applied principal component analysis and retained the minimum number of components that together explained at least 95% of the variance, yielding g_r.

We defined the morphology fingerprint as $m_r \in R^{K_M}$. For each region, we aggregated a total of 64 morphology features, including 32 axonal features and 32 dendritic features. Feature values were transformed using $\log(1 + \mathrm{x})$ to reduce skewness and then standardized by z score normalization across regions. We applied principal component analysis and retained components explaining at least 95% of the variance, yielding m_r.

We defined the projection fingerprint as $p_r \in R^{K_P}$. For each region, we constructed a target region weight vector from atlas registered projection summaries. We removed dimensions that were zero across all regions and applied a \log(1+x) transform followed by z score normalization across regions. Principal component analysis was then applied and components explaining at least 95% of the variance were retained, yielding p_r.

Similarities, distances, and cross-modality divergence

We quantified region to region similarity separately for morphological, molecular, and projection fingerprints. For each modality, pairwise similarity was computed on the PCA reduced vectors using cosine similarity, and a corresponding distance was defined as one minus similarity. This design matches the similarity matrices shown in **Figure 4A** and provides a common scale for defining cross-modality divergence in **Figure 4B**.

For molecule, we defined the cosine similarity between regions i and j as

$$S_G(i,j) = \mathrm{cosine}(g_i, g_j) = (g_i \cdot g_j)/(\|g_i\|\|g_j\|),$$

where g_i and g_j are the PCA-reduced molecular vectors for regions i and j, shown as column vectors. The molecular distance was defined as

$$D_G(i,j) = 1 - S_G(i,j).$$

For morphology, we used cosine similarity as the primary metric on the PCA reduced morphology fingerprints. The cosine similarity was defined as

$$S_M(i,j) = \text{cosine}\big(m_i, m_j\big),$$

and the corresponding morphology distance was

$$D_M(i,j) \;=\; 1 \;-\; S_M(i,j).$$

As a robustness check, we also computed Euclidean distances on z score standardized morphology features, and the qualitative conclusions were unchanged.

For connectivity, we defined cosine similarity on the PCA reduced projection fingerprints as

$$S_P(i,j) = \text{cosine}\big(p_i, p_j\big),$$

and we defined the connectivity distance as

$$D_P(i,j) = 1 - S_P(i,j).$$

To quantify cross-modality divergence, we defined mismatch indices that measure disagreement between structures across modalities. Specifically, for each region pair (i,j), we used the absolute distance difference to construct two divergence measures, where larger values indicate stronger cross modal mismatch. The molecule–morphology mismatch was defined as

$$\text{MM}_{\text{GM}}(i,j) = |D_G(i,j) - D_M(i,j)|,$$

and the molecule–projection mismatch was defined as

$$\text{MM}_{\text{GP}}(i,j) = |D_G(i,j) - D_P(i,j)|.$$

Here, the subscripts GM and GP indicate the modality pairs (Gene–Morphology and Gene–Projection), respectively. These two indices correspond to the molecule versus morphology divergence map and the molecule versus projection divergence map in **Figure 4B**. We further used these indices to rank region pairs and to select exemplar comparisons for the evidence panels in **Figure 4C** and **Figure 4D**.

Machine-learning classifiers for regional identity prediction tasks

We evaluated how well each modality captures regional identity by training supervised classifiers to predict brain region labels from modality specific representations. The single modality baseline was implemented with a Random Forest classifier trained on one input modality. For multimodal prediction, we used a late fusion strategy in which modality specific classifiers were trained independently and their predicted class probabilities were combined by a meta classifier. The fused predictor was defined as

$$\widehat{y_{\text{fusion}}} = f_{\text{meta}}\big(f_1(x_1), f_2(x_2)\big),$$

where f_1 and f_2 are Random Forest classifiers trained on different modalities and f_{meta} is a logistic regression model trained on the probability outputs of the base classifiers. Hyperparameters for all classifiers were selected using randomized search under stratified five-fold cross validation, and the same training and evaluation protocol was applied across modalities to ensure comparability. This setup corresponds to the regional identity prediction tasks summarized in **Figure 5A** and **Figure 5C**.

Brain region separability and confusion-score-based analysis

To quantify region separability in a way that reflects local confusability rather than global averages across all region pairs, we computed a Fisher style local confusion score based on within region dispersion and between region centroid separation, all measured in a given feature space (molecule, morphology, projection, or fused). For each region r, we defined a dimension normalized within-region noise as

$$s_r^2 = \frac{1}{n_r}\sum_x \frac{|x-\mu_r|^2}{d},$$

where x denotes the feature representation of a neuron assigned to region r, μ_r is the centroid of region r in the same feature space, n_r is the number of neurons in region r, and d is the feature dimensionality. For any pair of regions i and j, we defined the dimension normalized between region distance as

$$b_{ij} = \frac{|\mu_i-\mu_j|^2}{d}.$$

We then defined a local confusion score for a region i by restricting comparisons to its k nearest neighboring regions. Specifically, $N_k(i)$ denotes the set of k regions whose centroids are closest to region i (excluding i itself) under centroid distances computed in the morphology-derived space. The local confusion score was defined as

$$C_i(k) = \frac{s_i^2}{\text{mean}_{j\in N_k(i)}(b_{ij})}.$$

Lower values indicate better separability because they reflect smaller within region dispersion relative to separation from nearby region centroids. We summarized overall separability by averaging the local confusion scores across regions, which yields a global confusion index defined as

$$\text{Confusion}(k) = \text{mean}_i\, C_i(k).$$

Unless otherwise noted, we used k=5 for local evaluation to emphasize confusable region neighborhoods rather than averaging over all region pairs. This analysis corresponds to the region separability and confusion results shown in **Figure 5D**.

To identify specific region pairs that remain difficult to distinguish, we additionally computed a pairwise confusion matrix. For each region pair i and j, the pairwise confusion score was defined as

$$C_{ij} = \frac{s_i^2+s_j^2}{b_{ij}+\varepsilon},$$

where ε is a small constant included to avoid numerical instability when centroid distances are close to zero. Higher C_{ij} values indicate stronger confusability between the two regions under the same feature representation.

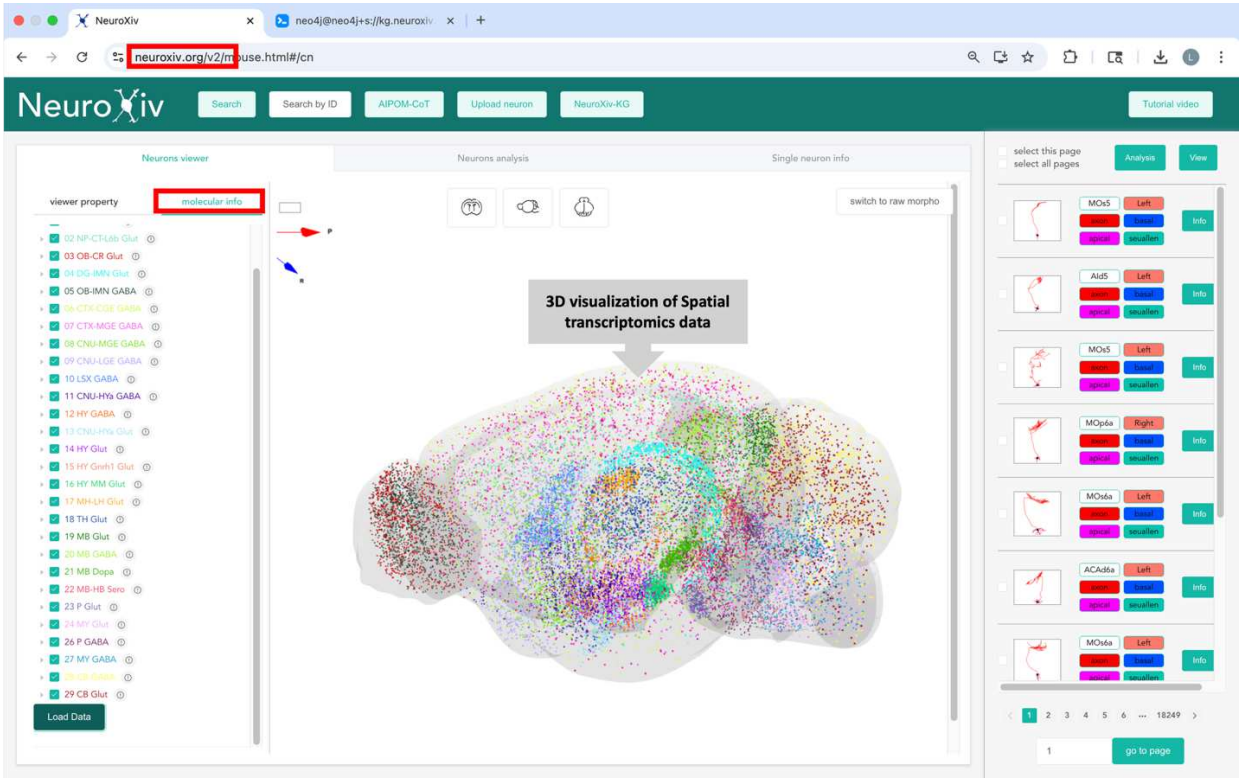
Spatial homogeneity within brain regions

To characterize spatial homogeneity within each brain region, we quantified within-region dispersion using the trace of the covariance matrix computed from neuron level features. For region r, we defined

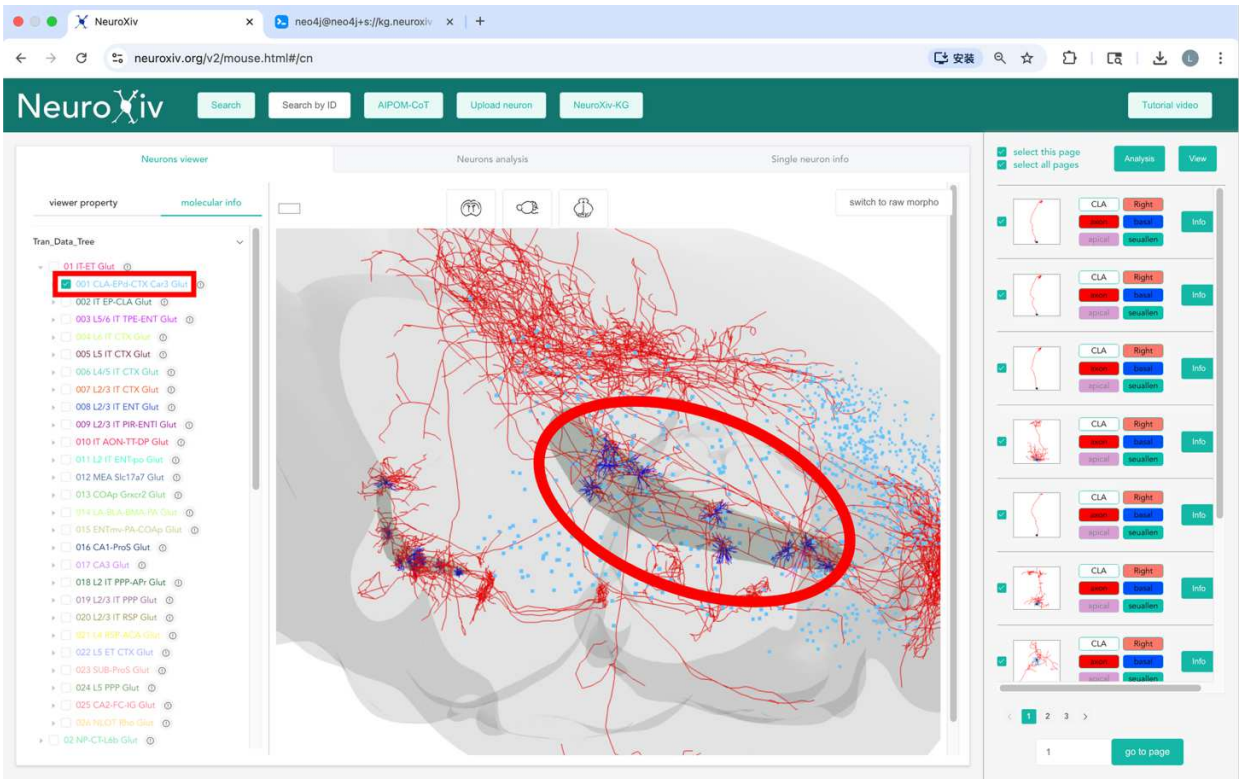
$$H(r) = \frac{\text{trace}\big(\text{Cov}(X_r)\big)}{d}$$

where X_r is the feature matrix of neurons in region r, with rows representing neurons and columns representing features, $\mathrm{Cov}(X_r)$ is its covariance matrix, and d is the feature dimensionality. The dimension normalization allows direct comparison across modalities that may differ in feature scale or representation. Lower values of $H(r)$ indicate higher spatial homogeneity, meaning that neurons within the same region occupy a more compact distribution in the feature space.

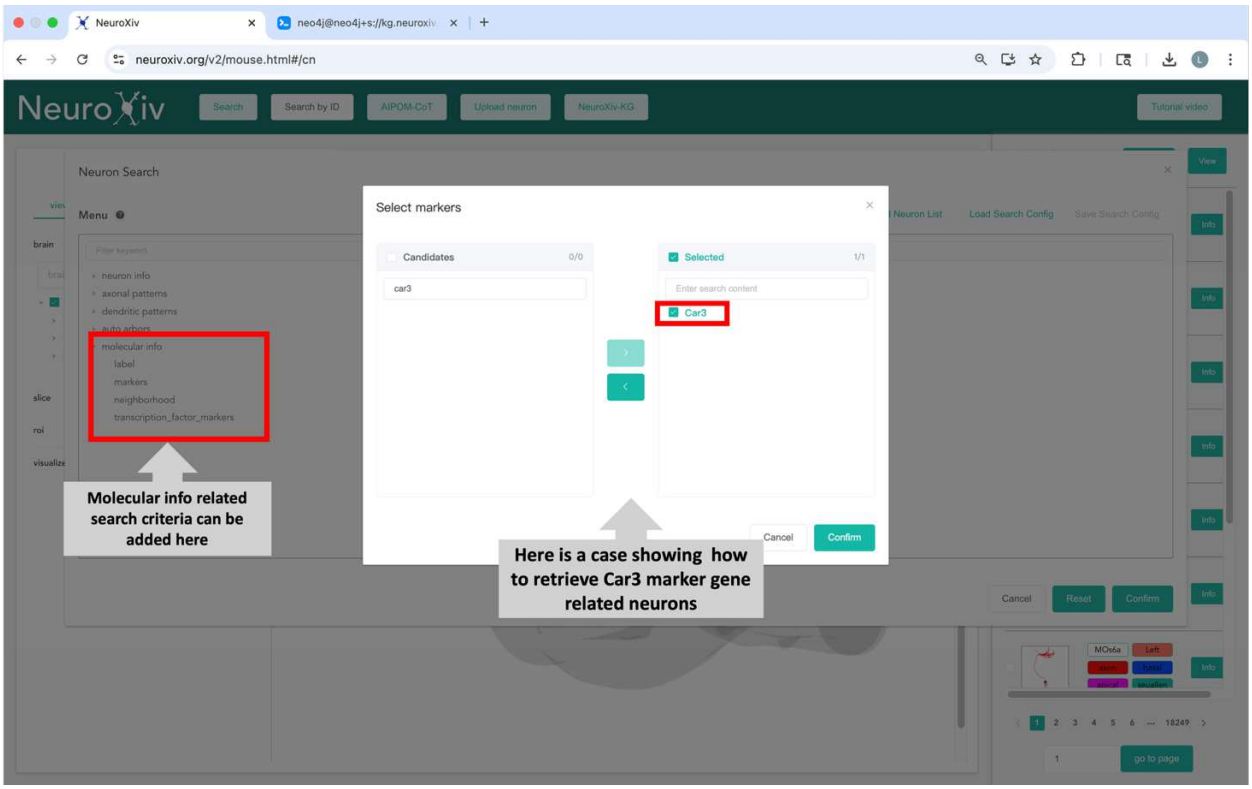
Supplementary Figures



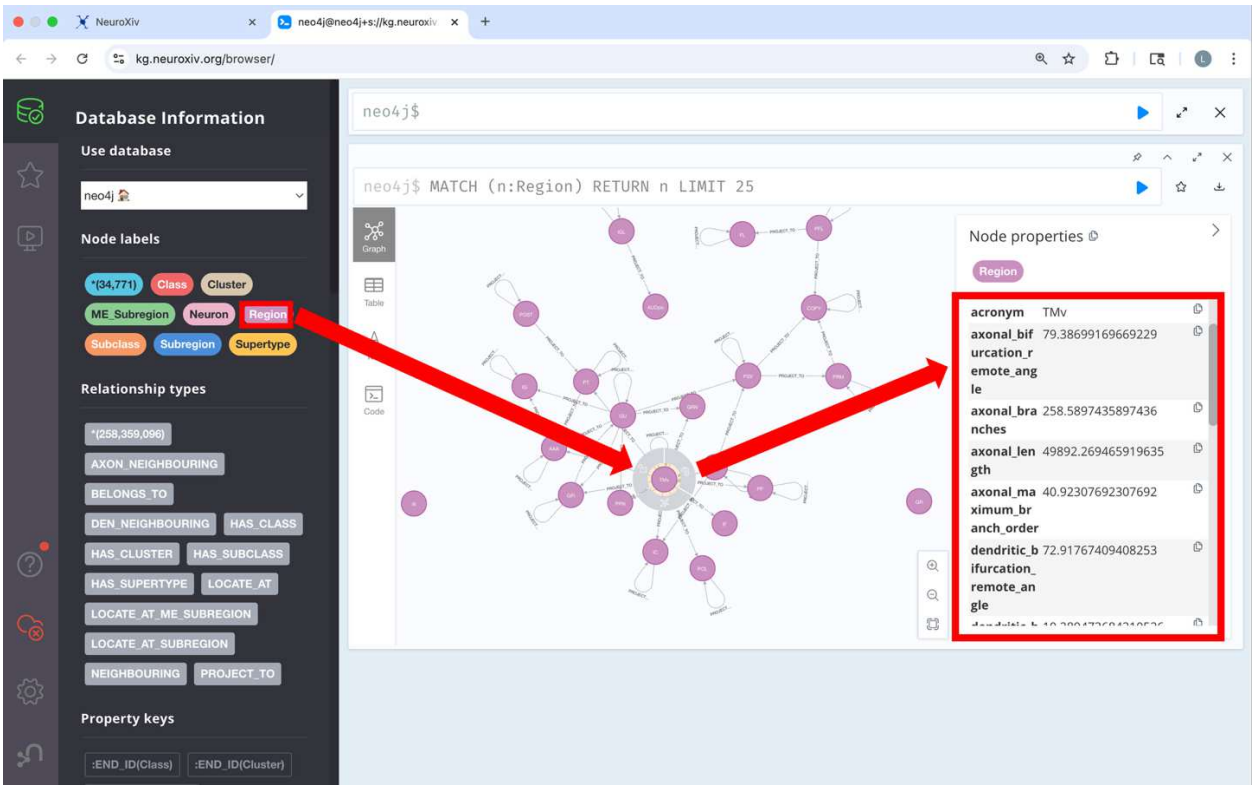
Supplementary Figure 1. *A screenshot of the 3D visualization of Spatial transcriptomics data.*



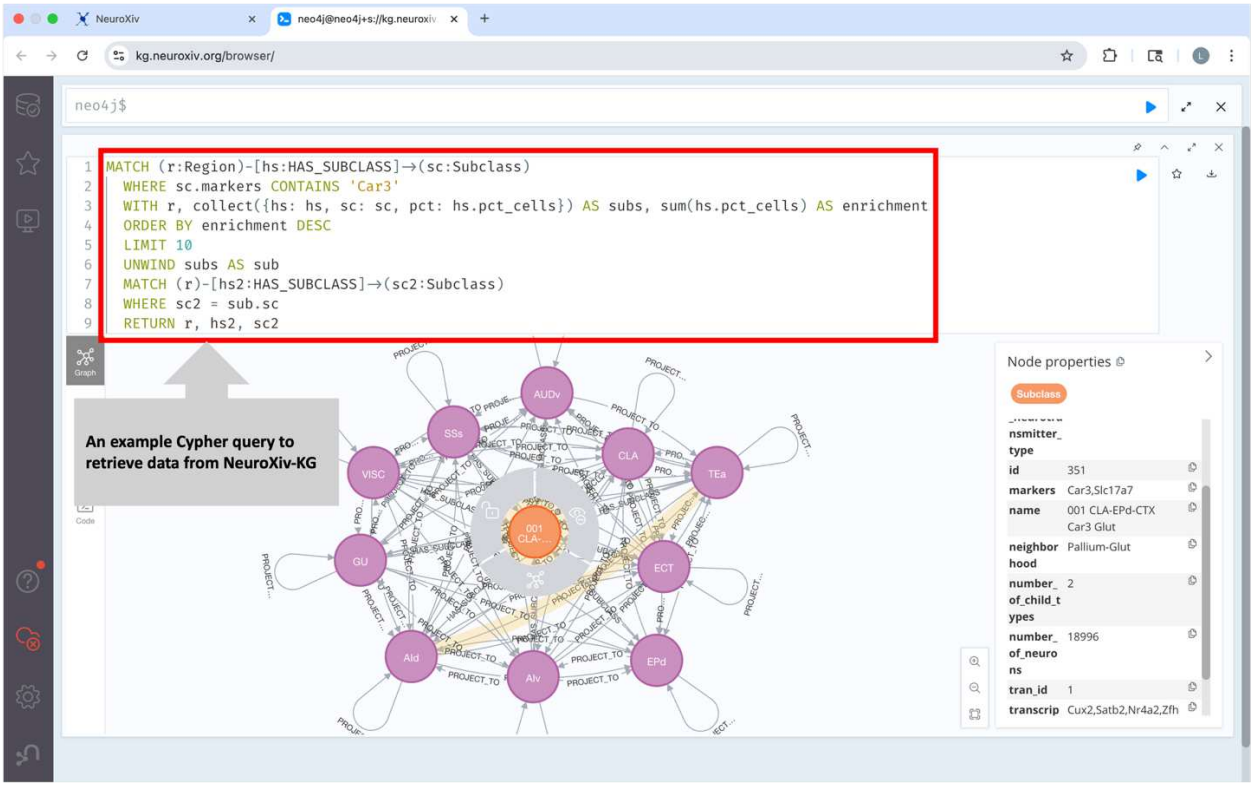
Supplementary Figure 2. *A screenshot showing the 3D visualization of tri-modal data (Morphology, Connectivity and Molecule) together. The tri-modal data are anchored together through spatial position.*



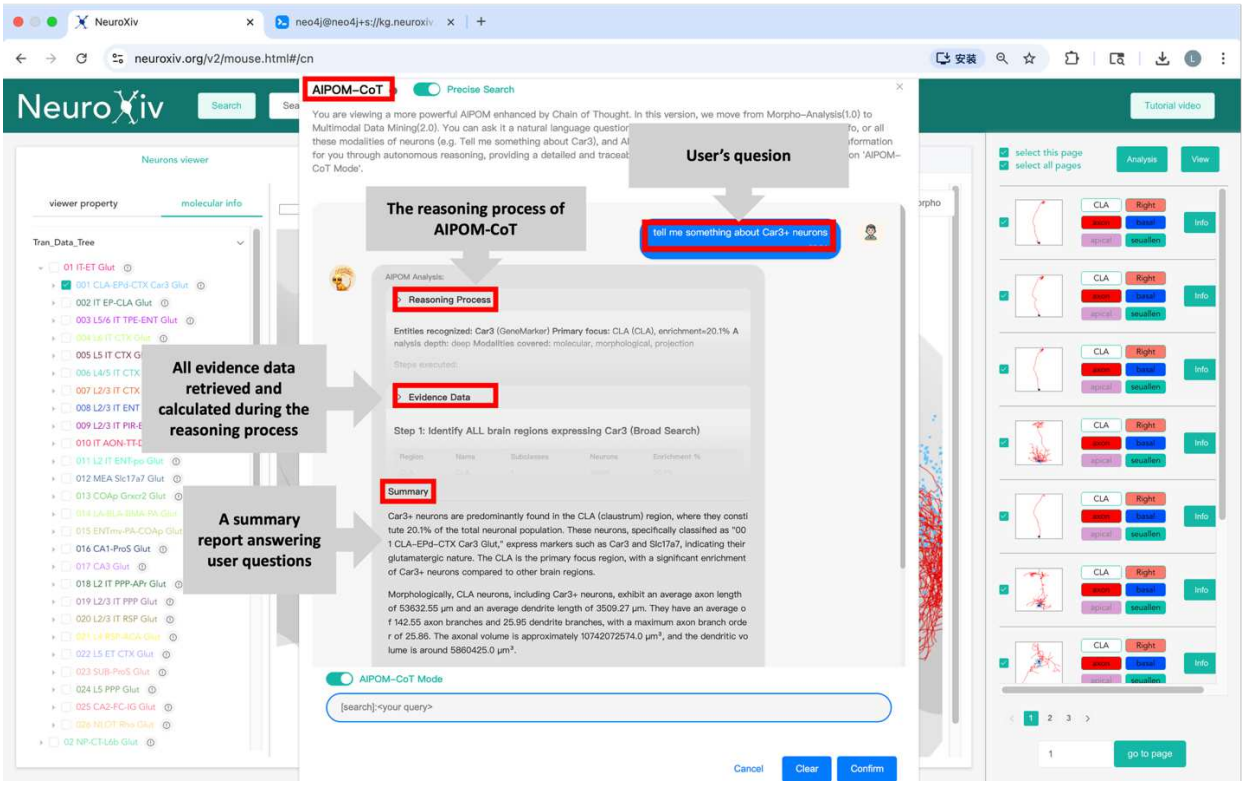
Supplementary Figure 3. *A screenshot demonstrating how to retrieve Car3 marker gene related neurons.*



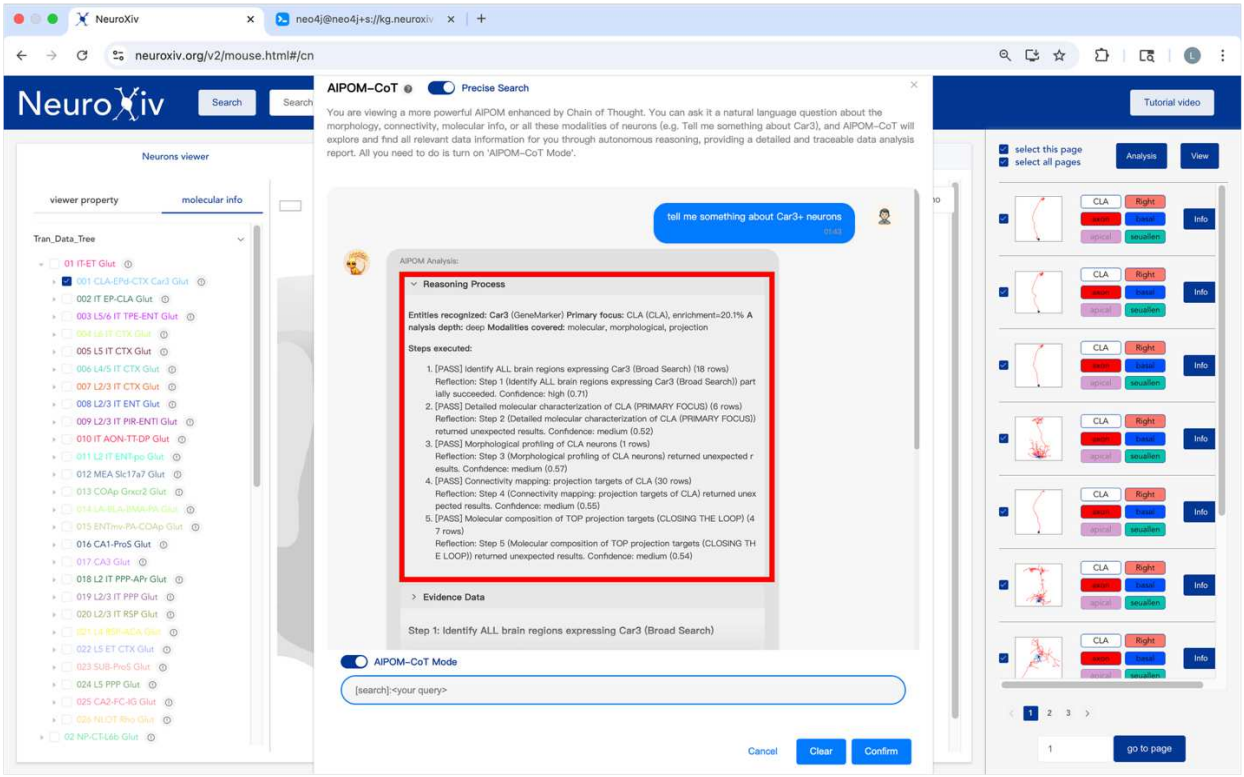
Supplementary Figure 4. *A screenshot of the overview of NeuroXiv-KG. Users can retrieve all entity nodes in the left list, the relationship between entities and the metadata of the entity itself*



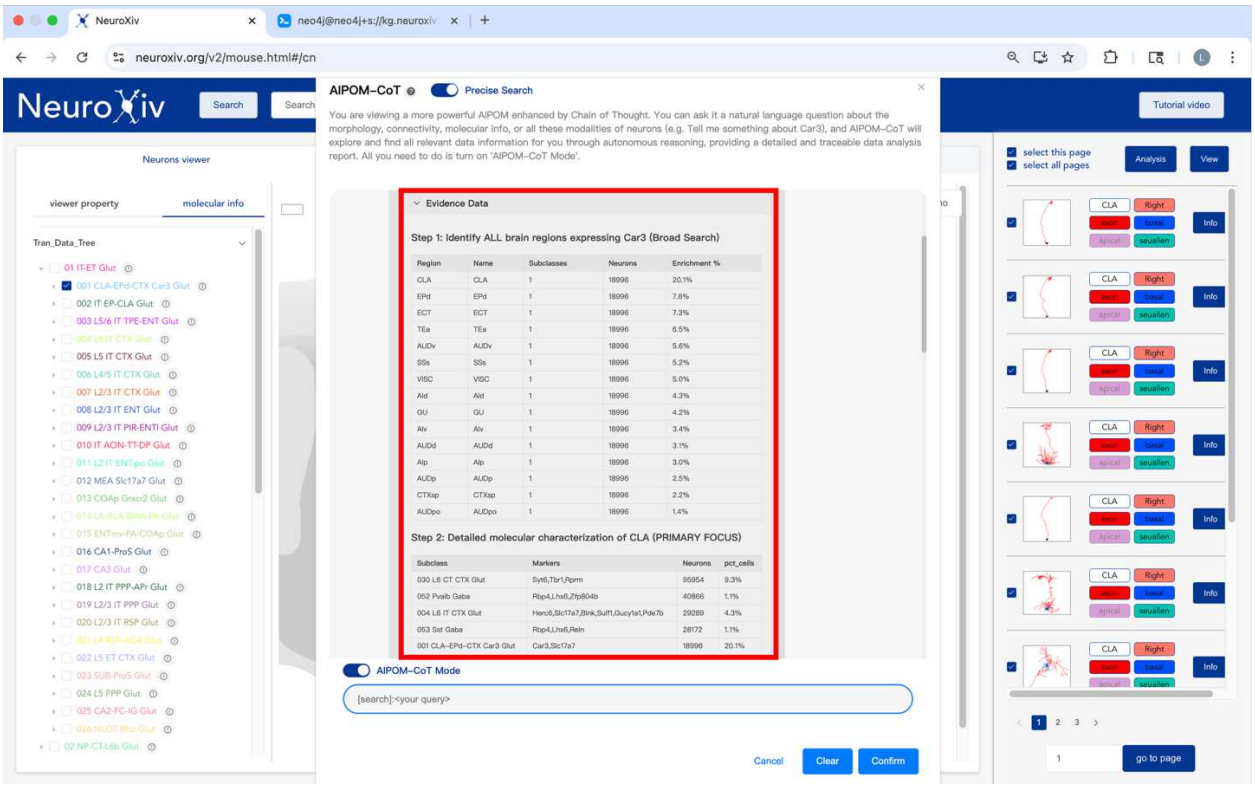
Supplementary Figure 5. *A screenshot of an example Cypher query to retrieve data from NeuroXiv-KG. Users can use the example cypher query to retrieve the brain region with the highest proportion of Car3-associated neurons from the NeuroXiv-KG.*



Supplementary Figure 6. *A screenshot demonstrating how to use AIPOM-CoT to do the multimodal data mining automatically. A text box or input field allows users to enter their questions. Then AIPOM-COT will parse user intent, plan execution, automatically complete data exploration, retrieval and calculation, and finally return all intermediate key data evidence and a detailed and reliable data analysis summary report.*



Supplementary Figure 7. *A screenshot showing the reasoning process of AIPOM-CoT.*



Supplementary Figure 8. *A screenshot showing the complete evidence data offered by AIPOM-CoT.*