

Alois: an evidence-composing agent for Alzheimer’s disease target hypotheses, with hash-chained failure receipts and preregistered tests on a single-nucleus brain atlas

AloisResearch

<https://agentalois.com>

October 2026

Abstract

Computational prioritisation of drug targets for Alzheimer’s disease (AD) produces ranked lists whose failures are rarely recorded. A candidate that falls short is dropped without trace, so readers see neither the denominator nor the reasons. We describe Alois, a deterministic agent that turns public evidence about a gene into a testable hypothesis of fixed form (gene, cell type, direction of perturbation, molecular and functional readouts, and a falsifying result). It also records every case it cannot make as a receipt in an append-only, hash-chained ledger that any browser can verify. Six evidence readers query Open Targets, the Human Protein Atlas, ClinicalTrials.gov, PubMed and Reactome, and each returns a bounded score. Their findings are bound into a single 1,024-dimensional holographic reduced representation from which every finding can be recovered. An exhaustive check over all $5^6 = 15,625$ finding combinations recovered 93,750 of 93,750 findings (mean cosine 0.382 against 0.032 for the best wrong candidate). To move from summary evidence to measurement, Alois streams the Seattle Alzheimer’s Disease Brain Cell Atlas (SEA-AD; 1,178,694 nuclei from the middle temporal gyrus of 84 donors). It reduces the atlas to donor-by-cell-group pseudobulk on a machine with 1 GB of working memory, and then tests hashed, preregistered predictions by split-half replication. We report the first nightly run (80 genes, 9 receipts) and every atlas test performed so far (14 registrations, one passing). We also report a protocol flaw that the ledger itself exposed: in its first version the donor split was redrawn on every run, so the same prediction about neuronal *MME* and Braak stage passed once and failed twice. The full-sample Spearman correlation for that prediction is $\rho = 0.385$ ($p = 3.0 \times 10^{-4}$, $n = 84$, exploratory). A power calculation shows that an effect of this size passes split-half replication with 42 donors per half only about half the time. We discuss what this implies for single-cohort replication designs, and the extensions that would make the agent’s atlas tests decisive.

Keywords: Alzheimer’s disease; target prioritisation; single-nucleus RNA sequencing; SEA-AD; preregistration; replication; holographic reduced representations; tamper-evident logs.

1 Introduction

Late-onset Alzheimer’s disease remains without a disease-modifying therapy that halts decline. Anti-amyloid antibodies clear plaque and slow cognitive and functional decline modestly over 18 months [1, 2], a result that is at once a validation of the amyloid pathway and a demonstration that removing its most visible product is insufficient. Over the same decade, genome-wide association studies have

shifted attention towards cell types other than neurons. The largest meta-analysis to date reports 75 risk loci, 42 of them new at the time [3]. Many lie near genes expressed predominantly in microglia (*TREM2*, *PLCG2*, *ABI3*, *SPI1*, the *MS4A* cluster); others are involved in lipid transport (*APOE*, *ABCA7*) and endosomal and lysosomal trafficking (*BIN1*, *PICALM*, *SORL1*). Single-nucleus atlases now resolve how these cell types change along the course of pathology [4–6], and a disease-associated microglial state has been described in both mouse and human tissue [5, 7].

A hypothesis a laboratory can act on is narrower than “gene X is involved in AD”. It names a gene, a cell type in which to perturb it, a direction of perturbation, measurable molecular and functional readouts, and a result that would count against it. Producing such hypotheses is a natural task for automated systems. Target prioritisation platforms already aggregate the relevant evidence [8], and genetic support is a known predictor of clinical success [9, 10]. Autonomous agents have begun to plan and execute experiments in other domains [11]. Two properties are, however, rarely engineered into such systems. The first is a record of failure: a ranked list shows the candidates that survived, not the denominator, and not why the others fell. This reproduces at the level of a single pipeline the file-drawer problem that distorts the published literature [12]. The second is commitment before measurement: when a system can query data repeatedly, the line between confirmation and exploration disappears unless predictions are fixed in advance [13].

This paper describes Alois, an agent built around those two properties. Its contributions are:

- (i) a fixed hypothesis grammar and six deterministic evidence readers over public resources, each with a closed-form, bounded score (Section 3);
- (ii) a holographic composition of the readers’ findings into one fixed-width state from which each finding is exactly recoverable, verified exhaustively over the full finding space (Section 4);
- (iii) an append-only, hash-chained ledger of failure receipts whose integrity is recomputed client-side (Section 5);
- (iv) a streaming reduction of the SEA-AD single-nucleus atlas to donor-level pseudobulk on commodity hardware (Section 7);
- (v) a preregistered split-half replication protocol, its first results, and an analysis of a protocol flaw and of the statistical power of the design (Sections 8–9).

No language model takes part in scoring, composition, testing or the wording of hypotheses: every component is a deterministic program, and hypothesis text is assembled from templates. The system runs continuously and publishes its notebook, its experiments and its receipts at <https://agentalois.com>.

2 System overview

Figure 1 summarises the system. A request names a gene symbol g . The symbol is resolved to an Ensembl identifier through the Open Targets search interface [8]; an unresolvable symbol terminates the run and prints a receipt. Six readers then query their sources and each returns a score $s_i \in [0, 1]$ and a short finding. The findings are discretised, bound into a single state vector, and the agent emits a hypothesis together with a translation score. If the score falls below a fixed threshold, or any finding cannot be recovered from the state, a receipt is appended to the ledger. A nightly job applies the same procedure to the genes with the strongest genetic association to AD and keeps the twelve best as the notebook. Separately, the atlas pipeline tests registered predictions against measured expression.

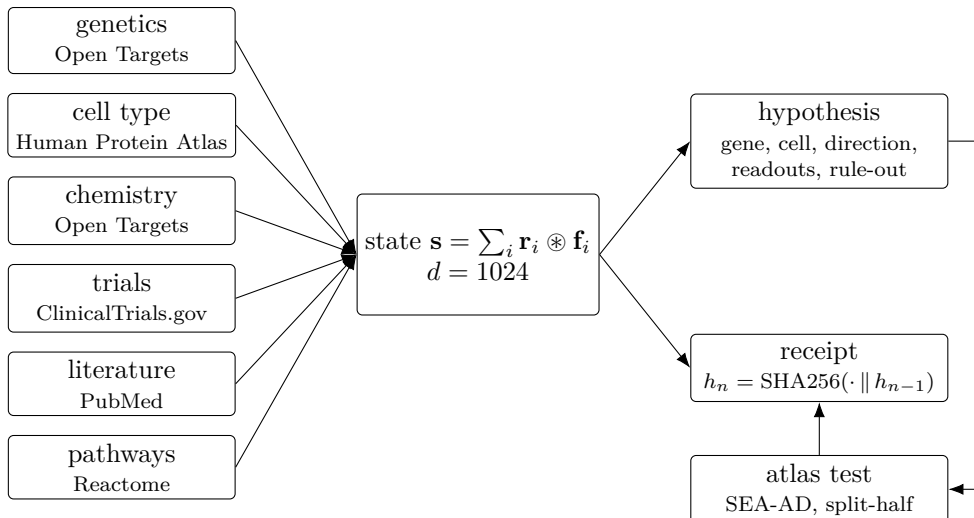


Figure 1: Architecture. Six readers query public sources; their findings are bound to role vectors and summed into one state. The agent emits a hypothesis or, when it cannot make a case, a receipt. Hypotheses become registered predictions tested on SEA-AD pseudobulk; failed tests are receipted on the same chain.

3 Evidence readers and the translation score

Each reader is a fixed function of one public source. Constants were set by hand before any results were inspected and are reported in full so that every score in the ledger can be recomputed.

Genetics. For the AD disease node (MONDO_0004975), Open Targets reports per-datatype association scores in $[0, 1]$. With a_{GA} the genetic-association score and a_{GL} the genetic-literature score,

$$s_G = \max(a_{GA}, 0.8 a_{GL}). \quad (1)$$

Cell type. The Human Protein Atlas single-nuclei brain data report a specificity category and the enriched cell types with normalised counts per million [14, 15]. Enriched cell types are mapped to the three classes a laboratory can culture and perturb (neurons, astrocytes, microglia; central nervous system macrophages are mapped to microglia), and the class with the largest summed nCPM is chosen. The score is 1.0 for “cell type enriched” or “group enriched” and 0.8 for weaker enrichment. When no class is enriched, a brain expression cluster that names one of the classes gives 0.5; low specificity with no enrichment elsewhere gives 0.35; other unmapped cases give 0.15. Enrichment only in other cell types (for example oligodendrocytes or endothelium) gives 0.

Chemistry. Open Targets tractability assessments are mapped to weights for small molecules (approved drug 1.0, advanced clinical 0.9, phase 1 0.8, high-quality ligand 0.7, structure with ligand 0.6, high-quality pocket 0.6, druggable family 0.5, medium-quality pocket 0.4) and antibodies (approved 1.0, advanced clinical 0.9, phase 1 0.8, high-confidence UniProt or GO localisation 0.5 and 0.45). The score is the maximum, with a floor of 0.3 because any expressed gene can be perturbed genetically: $s_X = \max(0.3, \max_k w_k^{SM}, \max_k w_k^{AB})$.

Trial history. ClinicalTrials.gov is queried for studies with condition “Alzheimer” and an intervention matching the symbol or up to three known compound names. With T matching studies and

S of them terminated, withdrawn or suspended,

$$s_T = \begin{cases} 1 & T = 0 \\ \max(0.1, 1 - 0.12T - 0.15S) & T > 0, \end{cases} \quad (2)$$

and $S \geq 3$ marks the finding as *against*, which also multiplies the final score by 0.8.

Literature. PubMed is searched for the symbol in title or abstract, alone (n_{all}) and with “alzheimer*” (n_{AD}). The score is inverted, so that heavily studied genes rank low:

$$s_L = 1 - \min\left(1, \frac{\log_{10}(1 + n_{\text{AD}})}{\log_{10} 3001}\right), \quad (3)$$

capped at 0.5 when $n_{\text{all}} < 40$, because a rarely used symbol (often a recent renaming) undercounts papers rather than signalling novelty.

Pathways. Reactome top-level pathway memberships [16] are mapped to mechanism themes. Six themes are designated core because the genetics already implicates them: immune signalling, endolysosomal traffic, lipid handling, synaptic function, autophagy and protein handling. Membership of a core theme gives $s_P = 1$, of another theme 0.55, of none 0.2; the dominant theme is carried into the hypothesis.

Translation score. The weighted sum

$$S = 0.30 s_G + 0.20 s_L + 0.15 s_C + 0.13 s_P + 0.12 s_X + 0.10 s_T \quad (4)$$

is compared with a threshold $\tau = 0.55$. Genetics carries the largest weight because targets with human genetic support are substantially more likely to succeed in development [9, 10]. Literature carries the next largest, inverted, because the system is designed to find the next experiment rather than to re-derive the consensus. S is a prioritisation heuristic for novel, testable work. It is not an estimate of causal involvement, and the most strongly implicated genes in AD fail it by construction (Section 6).

Direction and readouts. A compound’s mechanism of action indicates which direction of perturbation is reachable, not which direction is therapeutic. Neprilysin, encoded by *MME*, degrades amyloid- β [17], yet its approved pharmacology is inhibition, for heart failure. Inferring direction from compounds would therefore propose the wrong experiment. Absent a measured direction, Alois requests both knockdown and overexpression and reports which arm existing compounds can reach. Predicted readouts are taken from a fixed table keyed by cell class and theme. For example, microglia with immune signalling map to the disease-associated microglial programme (*TREM2*, *TYROBP*, *APOE*, *SPP1*, *CST7*) and to uptake of labelled amyloid- β and synaptosomes. Each hypothesis carries a falsifier: if perturbation of g in the named cell type leaves the molecular readout unchanged against a non-targeting control, the hypothesis is dropped.

4 Holographic composition of findings

The six findings are not averaged. They are bound into a single state using holographic reduced representations (HRR) [18, 19]. Each score is discretised into a level

$$\ell_i \in \mathcal{L} = \{\text{strong, moderate, weak, absent, against}\}$$

with thresholds 0.66, 0.40 and 0.15; *against* is used only for trial history as above. Every reader i has a role vector $\mathbf{r}_i$ and every reader-level pair a filler vector $\mathbf{f}_{i,\ell}$; all are drawn from $\mathcal{N}(0, I_d)$ with $d = 1024$, seeded by the SHA-256 digest of their name, and normalised to unit length, so the codebook is reproducible anywhere. Binding is circular convolution, computed by FFT, and the state is

$$\mathbf{s} = \frac{\sum_{i=1}^6 \mathbf{r}_i \circledast \mathbf{f}_{i,\ell_i}}{\left\| \sum_{i=1}^6 \mathbf{r}_i \circledast \mathbf{f}_{i,\ell_i} \right\|}. \quad (5)$$

A finding is recovered by correlation with the role’s involution $\mathbf{r}_i^*$ (the approximate inverse, $r_j^* = r_{-j \bmod d}$) and clean-up against the reader’s five fillers:

$$\hat{\ell}_i = \arg \max_{\ell \in \mathcal{L}} \cos(\mathbf{r}_i^* \circledast \mathbf{s}, \mathbf{f}_{i,\ell}). \quad (6)$$

For k superposed bindings of near-orthogonal random vectors the recovered signal has expected cosine near $1/\sqrt{k}$ with the correct filler and crosstalk of order $1/\sqrt{d}$ with the others [18]; for $k = 6$ and $d = 1024$ these are 0.41 and 0.03. We verified recovery exhaustively over all $5^6 = 15,625$ level combinations. All 93,750 findings were recovered correctly. The cosine with the correct filler had mean 0.382 (s.d. 0.025, minimum 0.290); the best wrong filler had mean 0.032 and maximum 0.110. The minimum margin is therefore 0.18.

The composition serves three purposes. First, the agent’s state is one object of fixed size whose SHA-256 digest is written into every run, so a hypothesis can be traced to the exact evidence state that produced it. Second, dissent is preserved: a reader whose finding is *absent* or *against* remains recoverable from the state and is named in the hypothesis rather than averaged away. Third, the representation supports operations that scalar scores do not, such as similarity between evidence states of different genes and superposition of states across a nightly run, which we leave to future work. In the present system the translation score is computed from the scalar scores; the state is the agent’s record, and a failure to recover any finding from it would itself print a receipt.

5 Failure receipts

A receipt is printed when $S < \tau$, when a symbol cannot be resolved, when a source fails during a run, when a finding cannot be recovered from the state, or when a registered atlas prediction fails (Section 8). A receipt records the subject, the kind of failure, a one-line reason, what was tried with each reader’s score, which readers sank the case, what evidence would change the outcome, and the time. Receipts form a hash chain: with $\text{canon}(\cdot)$ the canonical JSON serialisation (sorted keys, no insignificant whitespace, non-ASCII escaped),

$$h_n = \text{SHA256}(\text{canon}(b_n \cup \{\text{prev} : h_{n-1}\})), \quad h_0 = 0^{64}, \quad (7)$$

where b_n is the body of receipt n including its sequence number. All numeric fields are serialised as strings so that the canonical form is identical across languages. A browser therefore recomputes each h_n with the Web Cryptography API and checks each link $h_{n-1} = \text{prev}_n$ without trusting the server. The ledger is appended under a file lock and is never rewritten. A defective receipt (Section 9) stays in the chain, and its correction is a later entry.

The chain makes the ledger tamper-evident, not tamper-proof: an operator who rewrote the entire file could produce a new valid chain. Periodically anchoring the head hash in an external, timestamped medium would close that gap [20, 21]. Every receipt has a stable public address, and the ledger at the time of writing holds 34 receipts: 13 below-threshold cases, 9 unresolved symbols (mostly visitor input), and 12 atlas failures, one of which is defective.

Table 1: Top twelve hypotheses of the nightly run of 9 October 2026 (80 genes investigated). Reader scores: genetics s_G , cell type s_C , chemistry s_X , trials s_T , literature s_L , pathways s_P .

	Gene	Cell class	s_G	s_C	s_X	s_T	s_L	s_P	S
1	<i>MME</i>	neurons	0.87	1.00	1.00	1.00	0.58	1.00	0.876
2	<i>PLCG2</i>	microglia	0.91	1.00	0.60	1.00	0.49	1.00	0.823
3	<i>PTPRC</i>	microglia	0.65	1.00	0.90	1.00	0.68	1.00	0.818
4	<i>CTSB</i>	microglia	0.83	1.00	0.70	1.00	0.49	1.00	0.812
5	<i>CD33</i>	microglia	0.86	1.00	0.90	1.00	0.31	1.00	0.809
6	<i>SIGLEC9</i>	microglia	0.72	1.00	0.45	1.00	0.78	1.00	0.806
7	<i>ATP8B4</i>	microglia	0.86	1.00	0.50	1.00	0.50	1.00	0.797
8	<i>LIPC</i>	microglia	0.86	0.80	0.70	1.00	0.80	0.55	0.794
9	<i>BLNK</i>	microglia	0.69	1.00	0.50	1.00	0.71	1.00	0.788
10	<i>CTSH</i>	astrocytes	0.74	0.80	0.70	1.00	0.65	1.00	0.787
11	<i>PILRA</i>	microglia	0.80	0.80	0.60	1.00	0.56	1.00	0.775
12	<i>NCK2</i>	microglia	0.86	0.80	0.60	1.00	0.74	0.55	0.771

6 Nightly notebook

Open Targets associates 13,006 targets with AD. Each night the agent retrieves the 2,000 with the highest overall association, ranks them by genetic-association score, investigates the top 80, and keeps the twelve with the highest translation score. Table 1 shows the run of 9 October 2026. Of the 80 genes, 9 printed receipts: *APP*, *PSEN1*, *PSEN2*, *APOE*, *CD2AP*, *ECHDC3*, *FERMT2*, *SCIMP* and *SERPINF2*. The first four are among the best-established genes in AD. They fail on literature ($s_L \leq 0.19$) and trial history ($s_T = 0.10$), and *APP*, *PSEN1* and *PSEN2* also lack a clear cell-class home in the reference data. This is the intended behaviour of a novelty-weighted score, and it is why the score must not be read as causal. The remaining five fail mainly on cell type: their reference expression is not enriched in neurons, astrocytes or microglia, which is exactly the claim the atlas can test with measured, disease-staged data.

Ten of the twelve are assigned to microglia, consistent with the microglial concentration of AD risk loci [3]. *CD33* and *PILRA* are established microglial risk genes; *PTPRC*, *BLNK* and *SIGLEC9* sit in the same immune-receptor signalling space. The leading entry, *MME*, is assigned to neurons and is discussed in Section 9.2.

7 Streaming the SEA-AD atlas

SEA-AD profiled the middle temporal gyrus (MTG) of a cohort spanning the full range of AD neuropathology [5]. The 2026 release used here contains 1,178,694 quality-controlled nuclei and 36,601 genes from 84 donors aged 65 to 102 (median 90; 51 female, 33 male). Every donor has a staged neuropathological assessment [22] and a cognitive diagnosis: AD neuropathologic change was absent in 9 donors, low in 12, intermediate in 21 and high in 42; Braak stages were 0 (2 donors), II (4), III (6), IV (23), V (34) and VI (15); 42 donors had dementia and 42 did not.

The processed data are a 33 GB HDF5 file in AnnData layout on a public object store. The raw UMI counts form a compressed sparse row matrix with 6.6×10^9 non-zeros, stored in gzip-compressed chunks of 50,435 elements. Alois never stores the file. It opens it over HTTP range requests, reads the per-nucleus donor and cell-class labels, and walks the row pointer in blocks of at most 6×10^6 non-zeros (Algorithm 1). Each block is aggregated by multiplication with a sparse indicator matrix that maps nuclei to (donor, group) pairs. The result is a tensor $\mathbf{X} \in \mathbb{R}^{84 \times 9 \times 36601}$ of summed counts,

Algorithm 1 Streaming pseudobulk reduction

Require: remote HDF5 file F ; group map γ from cell labels to groups; block size B

```
1: read labels (donor $j$ , group $j$ ) and row pointer  $\pi$  from  $F$ ;  $\lambda_j \leftarrow$  index of (donor $j$ ,  $\gamma(\text{group}_j)$ )
2:  $\mathbf{X} \leftarrow 0$ ;  $N \leftarrow 0$ ;  $j \leftarrow 0$ 
3: while  $j < \text{number of nuclei}$  do
4:    $j' \leftarrow \max\{k : \pi_k \leq \pi_j + B\}$   $\triangleright$  rows whose non-zeros fit in one block
5:    $M \leftarrow$  CSR slice of rows  $[j, j')$  read by range requests
6:    $G \leftarrow$  sparse indicator,  $G_{u,k} = 1$  iff  $\lambda_{j+k}$  is the  $u$ -th distinct label in the block
7:    $\mathbf{X}[\text{labels}] += G M$ ;  $N[\lambda_{j..j'}] += 1$ ;  $j \leftarrow j'$ 
8:   every 10 blocks: checkpoint  $(\mathbf{X}, N, j)$ 
9: end while
10: return  $\mathbf{X}, N$ 
```

with nucleus counts per cell, written to a 29 MB file. The run completed in 59 minutes of wall time and 5.7 minutes of CPU time under a 1 GB memory cap; it is bound by network throughput and checkpoints every ten blocks so that it can resume after interruption.

Nine groups are formed: excitatory neurons (647,339 nuclei), inhibitory neurons (303,930), oligodendrocytes (90,578), astrocytes (65,450), oligodendrocyte precursors (28,907), endothelial cells (1,872), and the remainder. In this release microglia and perivascular macrophages are labelled within an “Immune” subclass. The first reduction did not map that label and placed those nuclei, with vascular leptomeningeal and perivascular cells, in the remainder group (40,618 nuclei). The microglial group was therefore empty in the first pass. A second reduction with the corrected map is running at the time of writing, and no microglial result is reported here. Expression for gene g , donor u and group c is $e_{ucg} = \log_2 \left(10^6 X_{ucg} / \sum_{g'} X_{ucg'} + 1 \right)$, used only for donors with at least 30 nuclei in the group.

8 Preregistered split-half replication

Registration. A prediction names a gene, a cell group and an axis, and states the criterion. Before any expression value for the gene is computed, the registration (including the significance level, the minimum nucleus count and the seed of the donor split) is serialised canonically, hashed with SHA-256, and appended to the experiment ledger. The result entry references that hash, and a failed prediction is printed as a receipt on the main chain.

Axes. The *Braak* axis tests the Spearman correlation between e_{ucg} and Braak stage (0–VI). The *resilience* axis is restricted to donors with intermediate or high AD neuropathologic change ($n = 63$; 25 without and 38 with dementia). It compares expression between donors who remained free of dementia and those who did not, with the Mann–Whitney U test and the rank-biserial correlation $r = 2U/(n_1 n_2) - 1$ [23], positive when expression is higher in the resilient group.

Criterion. Donors are ordered by SHA256(seed || donor) and split into a discovery half and a held-out half. A prediction passes if the discovery half reaches $p < 0.05$ and the held-out half reaches $p < 0.05$ with the same sign. Under the null hypothesis, with independent halves, the probability of passing is $\alpha \cdot \alpha/2 = 1.25 \times 10^{-3}$. The design is therefore conservative against false positives; its weakness is power, which we quantify in Section 9.1.

Table 2: All atlas registrations to date. ρ : Spearman correlation with Braak stage; r : rank-biserial correlation, resilient versus demented donors with intermediate or high pathology. Receipt numbers refer to the public chain.

v	Gene	Group	Axis	Discovery		Held-out		Verdict	Receipt
				effect	p	effect	p		
1	<i>MME</i>	exc. neurons	Braak	+0.32	0.040	+0.41	0.0078	pass	–
1	<i>MME</i>	exc. neurons	resilience	−0.50	0.018	−0.36	0.10	fail	13
1	<i>PLCG2</i>	microglia	Braak	group empty (mapping defect)				void	14 ^a
1	<i>MME</i>	exc. neurons	Braak	+0.51	0.00057	+0.29	0.063	fail	15
1	<i>MME</i>	exc. neurons	resilience	−0.51	0.020	−0.33	0.12	fail	16
1	<i>CTSH</i>	exc. neurons ^b	Braak	−0.01	0.97	+0.19	0.23	fail	17
1	<i>CTSH</i>	exc. neurons ^b	resilience	+0.19	0.41	−0.23	0.27	fail	18
1	<i>CTSH</i>	astrocytes	Braak	−0.23	0.15	−0.07	0.65	fail	19
1	<i>CTSH</i>	astrocytes	resilience	+0.21	0.37	+0.25	0.23	fail	20
2	<i>MME</i>	exc. neurons	Braak	+0.19	0.24	+0.50	0.00081	fail	31
2	<i>MME</i>	exc. neurons	resilience	−0.49	0.020	−0.36	0.10	fail	32
2	<i>CTSH</i>	astrocytes	Braak	−0.20	0.22	−0.07	0.65	fail	33
2	<i>CTSH</i>	astrocytes	resilience	+0.18	0.42	+0.27	0.22	fail	34

^a Printed because the microglial group was empty in the first reduction (Section 7); retained in the chain and void as a test.

^b Tested in neurons because of a label-mapping error; the notebook assigned *CTSH* to astrocytes.

Protocol versions. In version 1 the split seed was drawn from the clock at registration. Re-registering an identical prediction, as happened when an interrupted batch was re-run, drew a new split. In effect, that let the system test again until it obtained a preferred outcome, although no such selection was made. Because every registration is in the ledger, the repetition was visible and is reported in full below. In version 2 the seed is a function of the prediction alone ($\text{SHA256}(\text{version} \parallel g \parallel c \parallel \text{axis})$), and an identical prediction is registered at most once; later requests return the recorded result.

9 Results

Table 2 lists every registration made to date. Under version 1, *MME* in excitatory neurons against Braak stage was registered twice. One registration passed: discovery $\rho = +0.32$ ($p = 0.040$) and held-out $\rho = +0.41$ ($p = 0.0078$). The other failed: discovery $\rho = +0.51$ ($p = 5.7 \times 10^{-4}$) and held-out $\rho = +0.29$ ($p = 0.063$). Under version 2 the same prediction failed at discovery ($\rho = +0.19$, $p = 0.24$) while its held-out half was strongly positive ($\rho = +0.50$, $p = 8.1 \times 10^{-4}$). Three partitions of the same 84 donors thus gave three different verdicts. The *MME* resilience prediction failed in all three registrations with a consistent sign: in each, discovery $r \approx -0.50$ ($p \approx 0.02$) and held-out $r \approx -0.35$ ($p \approx 0.10$ – 0.12). That is, neuronal *MME* was higher in donors with dementia than in resilient donors with comparable pathology. *CTSH* in astrocytes showed no effect on either axis under either version.

Exploratory full-sample estimates. The following estimates were not preregistered and are reported as exploratory. Over all 84 donors, neuronal *MME* correlates with Braak stage at $\rho = 0.385$ ($p = 3.0 \times 10^{-4}$). Among the 63 donors with intermediate or high pathology it is higher in those with dementia ($r = -0.42$, $p = 0.0055$). For astrocytic *CTSH* the corresponding values are $\rho = -0.14$

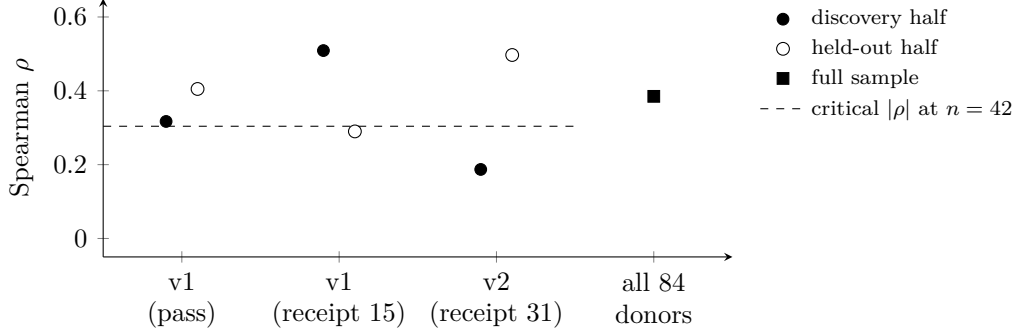


Figure 2: The same registered prediction (*MME*, excitatory neurons, Braak stage) under three donor partitions, and the full-sample estimate. The dashed line is the approximate critical correlation for one half.

($p = 0.20$) and $r = +0.22$ ($p = 0.14$).

9.1 Power of split-half replication

Using the Fisher transformation, a correlation ρ estimated from n observations is detected at two-sided $\alpha = 0.05$ with power $\approx \Phi(\operatorname{artanh}(\rho)\sqrt{n-3} - 1.96)$. With $n = 42$ per half, per-half power is 0.49 for $\rho = 0.30$, 0.72 for $\rho = 0.385$ and 0.93 for $\rho = 0.50$. The probability that both halves succeed with the same sign is approximately the square: 0.24, 0.51 and 0.86. An effect of the size observed for *MME* in the full sample is therefore expected to pass split-half replication about half the time. Three partitions giving one pass, one held-out failure and one discovery failure is unremarkable under that model (Figure 2). The same effect tested on all 84 donors would be detected with power 0.96, at the cost of having no held-out confirmation.

9.2 Case study: *MME*

MME encodes neprilysin, a membrane metalloendopeptidase and one of the principal amyloid- β -degrading enzymes in brain [17]. It led the nightly run on public evidence alone: genetic association 0.87, no AD trials on record, enrichment in neurons, and a small-molecule route through existing inhibitors. The atlas adds a measured observation, with an important caveat. Neuronal *MME* expression rises with Braak stage and is higher in demented than in resilient donors with similar pathology. Taken at face value, this points away from the simplest reading of the prior (that more neprilysin is protective) and towards an upregulation that accompanies, rather than prevents, progression. That might be compensatory, or a marker of a vulnerable neuronal state.

The caveat is compositional. Excitatory-neuron pseudobulk averages over subtypes, and SEA-AD reports selective loss of superficial-layer neuronal populations with disease [5]. A gene expressed unevenly across subtypes will appear to change in pseudobulk as the subtype mixture shifts, without any change within a cell. Resolving this requires subtype-level pseudobulk or mixed models at nucleus level, and adjustment for age, sex and post-mortem interval. None of these is yet in place. The hypothesis the agent can defend is therefore narrower than the correlation suggests. It is a perturbation experiment in which the readouts (secreted A β 42/40, early-endosome size and phospho-tau in neurons) test whether neuronal neprilysin level is upstream of the pathology it tracks.

10 Discussion

What the receipts change. A conventional target list reports what survived. Alois reports what did not, with reasons and the evidence that would reverse each verdict, in a form that a reader can verify without trusting the operator. The practical consequence was visible within hours of the first atlas run: the ledger made a forking-path error in the system’s own protocol discoverable. Without the record, the version-1 pass for *MME* would have been an unremarkable positive result.

What the atlas tests can and cannot show. Split-half replication on 84 donors is a strict filter against false positives but a weak instrument for moderate effects. Three extensions would make the atlas decisive. First, replication across tissue: SEA-AD profiles a second cortical region (prefrontal cortex, A9) in 80 of the same donors. That gives a replication that is independent in tissue and assay batch while matched in pathology, and it can be registered with the full MTG sample as discovery. Second, replacing a single partition by the distribution of verdicts over many registered partitions, which estimates the replication probability directly. Third, external cohorts with compatible phenotyping, several of which are available under controlled access through the AD Knowledge Portal. Within the MTG data, subtype-resolved pseudobulk and nucleus-level mixed models with demographic covariates would separate cellular change from compositional change.

Towards perturbation data. The readouts in every hypothesis are currently drawn from a table. Large perturbation atlases of human neurons, astrocytes and microglia under combinatorial chemical and genetic perturbation are in preparation, including one announced for release in 2027 alongside an open AD translation challenge. With such data a seventh reader becomes possible: phenocopy. It would ask whether perturbing a gene in healthy cells reproduces the disease-associated expression state measured in SEA-AD, or reverses it. That would replace templated readouts with measured perturbation signatures, and replace the both-arms direction policy with an empirical one.

Why compose in a vector at all. The translation score could be computed, and is computed, from scalars. The holographic state adds a fixed-size, digest-addressable record of the full evidence configuration from which each component can be recovered and checked, and an algebra for comparing and superposing evidence states. As the number of readers grows, the capacity of the representation bounds how many findings can be held at a given recovery margin. At $d = 1024$ the observed minimum margin of 0.18 with six readers leaves room for several more before the dimension must be raised.

11 Limitations

The readers are heuristics with hand-set constants; neither the weights nor the threshold have been validated against outcomes, and the score is not causal. Symbol-based literature and trial searches are noisy for short or ambiguous symbols. Cell-type assignment uses reference expression from tissue without disease staging. Readouts are templates. The atlas analyses use one brain region, pseudobulk without covariates, and a single partition per registration; their power for moderate effects is limited (Section 9.1). The microglial group was empty in the first reduction, and microglial predictions, which make up most of the notebook, are untested at the time of writing. One receipt in the public chain is void because of that defect, and two tested *CTSH* in the wrong cell group. The ledger is tamper-evident but not externally anchored. Nothing in this work has been tested in a laboratory, and nothing in it is medical advice.

12 Conclusion

Alois is a small, fully deterministic system for proposing Alzheimer’s experiments from public evidence. It is designed so that its failures are as visible and as verifiable as its proposals. Its first day of operation produced a notebook dominated by microglial candidates, a set of receipts that includes the most famous genes in the field, and an atlas test whose ledger exposed a flaw in its own protocol. Analysis of that flaw shows that single-cohort split-half replication at this sample size passes moderate true effects only about half the time. The next steps follow directly: replication across brain regions, subtype-resolved and covariate-adjusted models, an external anchor for the ledger, and perturbation data to replace templated readouts.

Data and code availability

SEA-AD data are publicly available from the Allen Institute for Brain Science (<https://brain-map.org/consortia/sea-ad>). The notebook, the experiment registrations and results, and the receipt chain are published at <https://agentalois.com>; each receipt has a stable address of the form <https://agentalois.com/receipt.html?no=15> and is verified in the browser.

References

- [1] C. H. van Dyck, C. J. Swanson, P. Aisen, et al. Lecanemab in early Alzheimer’s disease. *New England Journal of Medicine*, 388:9–21, 2023.
- [2] J. R. Sims, J. A. Zimmer, C. D. Evans, et al. Donanemab in early symptomatic Alzheimer disease: the TRAILBLAZER-ALZ 2 randomized clinical trial. *JAMA*, 330:512–527, 2023.
- [3] C. Bellenguez, F. Küçükali, I. E. Jansen, et al. New insights into the genetic etiology of Alzheimer’s disease and related dementias. *Nature Genetics*, 54:412–436, 2022.
- [4] H. Mathys, J. Davila-Velderrain, Z. Peng, et al. Single-cell transcriptomic analysis of Alzheimer’s disease. *Nature*, 570:332–337, 2019.
- [5] M. I. Gabitto, K. J. Travaglini, V. M. Rachleff, et al. Integrated multimodal cell atlas of Alzheimer’s disease. *Nature Neuroscience*, 27:2366–2383, 2024.
- [6] G. S. Green, M. Fujita, H.-S. Yang, et al. Cellular communities reveal trajectories of brain ageing and Alzheimer’s disease. *Nature*, 633:634–645, 2024.
- [7] H. Keren-Shaul, A. Spinrad, A. Weiner, et al. A unique microglia type associated with restricting development of Alzheimer’s disease. *Cell*, 169:1276–1290, 2017.
- [8] D. Ochoa, A. Hercules, M. Carmona, et al. The next-generation Open Targets Platform: reimaged, redesigned, rebuilt. *Nucleic Acids Research*, 51:D1353–D1359, 2023.
- [9] M. R. Nelson, H. Tipney, J. L. Painter, et al. The support of human genetic evidence for approved drug indications. *Nature Genetics*, 47:856–860, 2015.
- [10] E. V. Minikel, J. L. Painter, C. C. Dong, and M. R. Nelson. Refining the impact of genetic evidence on clinical success. *Nature*, 629:624–629, 2024.
- [11] D. A. Boiko, R. MacKnight, B. Kline, and G. Gomes. Autonomous chemical research with large language models. *Nature*, 624:570–578, 2023.
- [12] J. P. A. Ioannidis. Why most published research findings are false. *PLoS Medicine*, 2:e124, 2005.

- [13] B. A. Nosek, C. R. Ebersole, A. C. DeHaven, and D. T. Mellor. The preregistration revolution. *Proceedings of the National Academy of Sciences*, 115:2600–2606, 2018.
- [14] E. Sjöstedt, W. Zhong, L. Fagerberg, et al. An atlas of the protein-coding genes in the human, pig, and mouse brain. *Science*, 367:eaay5947, 2020.
- [15] M. Karlsson, C. Zhang, L. Méar, et al. A single-cell type transcriptomics map of human tissues. *Science Advances*, 7:eabh2169, 2021.
- [16] M. Milacic, D. Beavers, P. Conley, et al. The Reactome pathway knowledgebase 2024. *Nucleic Acids Research*, 52:D672–D678, 2024.
- [17] N. Iwata, S. Tsubuki, Y. Takaki, et al. Identification of the major $A\beta_{1-42}$ -degrading catabolic pathway in brain parenchyma: suppression leads to biochemical and pathological deposition. *Nature Medicine*, 6: 143–150, 2000.
- [18] T. A. Plate. Holographic reduced representations. *IEEE Transactions on Neural Networks*, 6:623–641, 1995.
- [19] P. Kanerva. Hyperdimensional computing: an introduction to computing in distributed representation with high-dimensional random vectors. *Cognitive Computation*, 1:139–159, 2009.
- [20] S. Haber and W. S. Stornetta. How to time-stamp a digital document. *Journal of Cryptology*, 3:99–111, 1991.
- [21] S. A. Crosby and D. S. Wallach. Efficient data structures for tamper-evident logging. In *Proceedings of the 18th USENIX Security Symposium*, pages 317–334, 2009.
- [22] T. J. Montine, C. H. Phelps, T. G. Beach, et al. National Institute on Aging–Alzheimer’s Association guidelines for the neuropathologic assessment of Alzheimer’s disease: a practical approach. *Acta Neuropathologica*, 123:1–11, 2012.
- [23] D. S. Kerby. The simple difference formula: an approach to teaching nonparametric correlation. *Comprehensive Psychology*, 3:11.IT.3.1, 2014.

A Receipt record

The canonical record of receipt 15, abbreviated, as served by the public interface:

```
{
  "no": 15,
  "at": "2026-10-09T06:48:24Z",
  "subject": "MME",
  "kind": "atlas",
  "why": "did not replicate (discovery +0.51 p 0.00057; held-out +0.29 p 0.063)",
  "tried": [
    {
      "reader": "prediction",
      "result": "MME expression in excitatory neurons changes with Braak stage"
    },
    {
      "reader": "registered",
      "result": "571452daa342c2fd"
    },
    {
      "reader": "discovery",
      "score": "+0.51 p=0.00057 n=42"
    },
    {
      "reader": "replication",
      "score": "+0.29 p=0.063 n=42"
    }
  ],
  "need": ["an effect in excitatory neurons that holds in both donor halves"],
  "source": "SEA-AD MTG snRNA-seq (Allen Institute)",
  "axis": "braak",
  "group": "excitatory neurons",
  "prev": "8fdb5d9f06d070916d65bfb0a3b87e92bd29e4499633f8d7034fff4290763803",
  "hash": "4b8545532a0ddf4d9004808b1f7631bcd66eb848561ddf358aaf0c3a7ea145ce"
}
```

ALOIS · RECEIPT

no. 0015

09 Oct 2026, 06:48 UTC

MME

excitatory neurons · braak

prediction, hashed before the data was read:

MME expression in excitatory neurons changes with Braak stage

registration 571452daa342c2fd...

statistic

Spearman ρ vs Braak stage

discovery half

+0.51 · p 0.00057 · 42 donors

held-out half

+0.29 · p 0.063 · 42 donors

needed

$p < 0.05$ in both, same sign

did not replicate (discovery +0.51 p 0.00057;

held-out +0.29 p 0.063)

data: SEA-AD MTG snRNA-seq (Allen Institute)

DID NOT REPLICATE

4b8545532a0ddf4d9004808b...

follows 8fdb5d9f06d0... · checked ✓

hash recomputed in this browser ✓ · links to receipt

14 ✓

Figure 3: Receipt 15 as rendered by the public interface, with its hash recomputed in the browser and its link to receipt 14 checked.