

Interpreting Omics Data Analysis with Large Language Models for Disease Target and Drug Discovery

Zixi Xu^{1*}, Weihang Chen^{1,2*}, Wuyu Ren^{1*}, Tianqi Xu¹,
Somadina Amaechina¹, Raad Khan^{1,4}, Yixin Chen², Michael
Province⁴, Philip Payne¹, Fuhai Li^{1,2,5}

¹Institute for Informatics, Data Science and Biostatistics (I2DB), ²Computer Science & Engineering, ³Department of Biology, ⁴Department of Genetics, ⁵Department of Pediatrics, Washington University in St Louis, MO, USA.

* Equal contributions; #Correspondence: Fuhai.Li@wustl.edu

Abstract

In biomedical scientific discovery, synthesizing prior knowledge from the literature is an essential component of interpreting numerical omics data analyses for disease target identification and drug discovery. Large language models (LLMs) alone can rapidly retrieve disease mechanisms from biomedical text, but text-only outputs are general and unreliable for target and drug prioritization without cohort-specific quantitative evidence. Herein, we propose a provenance-aware Text-to-Target framework that couples schema-constrained multi-model LLM retrieval with numeric omics data analysis. The key design is a modality-aware fusion step: candidates are partitioned into overlap-supported anchors, retrieval-only hidden hubs, and network-emergent novelty nodes, then propagated into staged hypothesis and strategy generation under topology constraints. We evaluate the model in Alzheimer’s disease (AD) and pancreatic ductal adenocarcinoma (PDAC). In PDAC, the workflow produced a balanced 75-gene candidate universe and a 23-strategy portfolio, with significant DepMap support at both target level and strategy level. In AD, stricter candidate controls yielded a compact 34-gene universe and 14 strategies; under an expanded CRISPRbrain registry, both target-level axes were significant, with strong strategy-level enrichment. Across both diseases, final strategies preserved full provenance closure to the candidate pool, enabling end-to-end auditability from retrieval

artifacts to validation outputs. These results support a transferable discovery architecture in which omics evidence constrains biological activity, LLM retrieval expands mechanistic search space, and network-aware fusion preserves interpretability. The framework provides a reproducible basis for dual-disease target prioritization and motivates continuous literature-mechanism concordance with agentic evidence-refresh loops.

Keywords: Large Language Models (LLMs), Single Cell, Omics, Precision Medicine, Network Biology, Target Discovery, Drug Discovery.

1 Introduction

1.1 Motivation

Large-scale single-cell omics data can potentially characterize key disease dysfunctional signaling mechanisms for precision target and drug discovery. Although all human cells share a common genome, disease phenotypes emerge from context-dependent regulatory programs and multicellular interactions [1, 2]. In practice, the key bottleneck is not detecting expression shifts, but translating noisy and descriptive omics contrasts into mechanistically prioritized, testable targets and effective drugs.

In biomedical target and drug discovery, literature synthesis remains essential for translating numerical omics contrasts into mechanism-level interpretation [3–9]. Current precision-medicine workflows therefore split into two weakly connected stages. First, computational omics pipelines quantify disease-vs-reference shifts [10–14]. Second, experts manually interpret those outputs using literature synthesis and biological reasoning. This handoff is slow, difficult to standardize, and vulnerable to omission. LLMs improve literature-scale retrieval speed, but text-only reasoning is not sufficient for reliable prioritization. LLM outputs can hallucinate links, miss context-specific dependencies, and drift from cohort-level transcriptional evidence [15]. The core challenge is therefore integration: how to combine broad semantic priors from text with quantitative cohort constraints from omics while preserving interpretability and auditability.

In this study, we propose to address this challenge with a Text-to-Target framework that couples LLM retrieval, Single-cell-resolved differential signals, and PathFinder-weighted network inference in a single provenance-aware pipeline. The framework is designed to produce mechanism-grounded candidates that can be traced from retrieval and network evidence to final validation outcomes.

1.2 Single-cell Omics and LLM Priors

Our framework assigns explicit roles to each evidence modality. Single-cell or meta-cell (average of about 200–300 single cells) omics serves as the quantitative backbone: we start from integrated single-cell and single-nucleus atlases, then analyze Single-cell-level profiles to reduce sparsity and dropout while preserving biologically meaningful cell states [16]. This representation improves the stability of differential-expression estimation and downstream network inference.

LLMs serve as structured literature-retrieval engines rather than standalone decision makers. Schema-constrained prompts retrieve cell-type-specific targets, pathway context, and mechanism statements in machine-parseable tables. Prompt templates are disease-matched: PDAC prompts focus on malignant epithelial and tumor-microenvironment signaling, whereas AD prompts emphasize neuroimmune coupling, glial state transitions, and intercellular propagation logic.

This division of labor is the central design principle. Omics constrains what is active in the cohort, LLM retrieval expands mechanistic search space, and network-based fusion resolves their intersection and disagreement. Candidates retained after fusion carry explicit provenance tags, enabling transparent traceability from final strategy outputs back to both quantitative and text-derived evidence.

1.3 AD and PDAC Testbeds

We evaluate the framework in two complementary diseases that stress different aspects of target discovery. PDAC is a high-mortality solid tumor with limited long-term benefit from current multi-agent regimens and strong unmet need for mechanism-guided target prioritization [17–19]. AD is a complex neurodegenerative disease affecting millions of people in the United States, and no curative therapy is currently available because key molecular mechanisms remain incompletely resolved [20–24].

PDAC provides a high-contrast benchmark for lineage-matched analysis. Disease progression from normal acinar or ductal states to malignant epithelial states is shaped by recurrent oncogenic programs, including frequent *KRAS* activation, but therapeutic responses remain constrained by resistance and microenvironment complexity [25]. To support robust expression contrasts, we built a high-confidence acinar-lineage reference from HCA, GEO, and PDAC single-nucleus resources [11, 26–28]. Acinar-lineage cells in PDAC cohorts were annotated through a two-stage process: lineage classification followed by disease-state stratification into normal-like and malignant-like subsets. This lineage-matched design reduces confounding from cell-identity imbalance and enables cleaner interpretation of malignant transformation-associated signaling. After consistent preprocessing and filtering, retained matrix dimensions differ modestly across groups by design, reflecting group-aware denoising rather than forced symmetry. AD provides a complementary low-contrast benchmark. Disease programs in AD are distributed across interacting neuronal and glial states, including astrocytes, microglia, oligodendroglial lineages, and vascular-associated compartments. Compared with PDAC, AD therefore requires tighter candidate-space control to preserve interpretability under weaker and more heterogeneous signal geometry. This dual-disease setup allows us to test both transferability and boundary conditions of the same pipeline. PDAC probes performance in a stronger malignant-vs-reference regime, whereas AD probes robustness under broader cell-state heterogeneity and diffuse effects.

1.4 Text-to-Target Overview

The Text-to-Target workflow is a staged integration cycle that alternates expansion and constraint. We use an ensemble of frontier LLMs (GPT-5, Gemini 2.5-pro, and DeepSeek-r1) to retrieve disease- and cell-context priors in schema-constrained form [29–32]. Single-cell differential-expression analysis then injects cohort-specific quantitative evidence, and PathFinder transforms these signals into weighted network structure.

Candidate fusion is performed through provenance-aware quadrants: *Q1* (jointly supported anchors), *Q2* (retrieval-only hidden hubs), and *Q3* (network-emergent novelty). Hypothesis upgrading and multi-target strategy generation are then constrained by explicit topology metadata, and final outputs are evaluated with disease-matched perturbation resources (DepMap for PDAC, CRISPRbrain for AD).

This design moves beyond direct list overlap and yields high-confidence targets supported by convergent text and omics evidence with explicit provenance closure. Our goal is not only stronger prioritization, but also an auditable discovery process in which each final strategy can be traced to concrete retrieval, network, and validation artifacts.

2 Methods

2.1 Workflow Overview

The Text-to-Target framework unifies semantic priors from biomedical text with quantitative evidence from disease-state omics while preserving the distinct epistemic role of each modality. Rather than treating the literature branch and expression branch as two lists to be overlapped, the workflow first converts transcriptomic contrast into a weighted signaling graph and then projects language-derived targets into this graph. This design separates jointly supported candidates from text-only and network-emergent candidates and retains provenance for each candidate class during downstream hypothesis generation and validation.

PathFinder is used as the primary omics integration engine that translates DEG evidence into topological context. Candidate construction is therefore not defined by direct LLM-DEG overlap alone. DEG-informed PathFinder networks provide weighted nodes and edges, and these are fused with curated LLM targets to form a quadrant candidate pool that supports both conservative and novelty-oriented discovery.

The full pipeline can therefore be read as a gradual information-constraining process. Literature retrieval starts broad and mechanistic, expression analysis constrains this space with disease-state measurements, network propagation organizes measured signals into pathway structure, quadrant construction labels candidates by evidence provenance, and final hypothesis/therapy generation is conditioned on this structured candidate pool rather than on an untyped merged list. The same provenance-aware design also supports disease-matched downstream validation at both single-target and multi-target strategy resolution, implemented in DepMap for PDAC and CRISPRbrain for AD.

2.2 Data Curation and Harmonization

Expression inputs were retrieved from the OmniCellTOSG ecosystem through CellTOSG Loader workflows [6]. For PDAC, we focused on a lineage-matched contrast by selecting malignant epithelial meta-cells as the disease state and acinar-lineage meta-cells as the reference state. Loader-level query constraints preserved key meta-data fields needed for traceability and were configured to reduce avoidable cohort imbalance across major descriptors when feasible.

For AD, we used the same loading and harmonization principles but instantiated disease-control contrasts across brain cell-state contexts relevant to neurodegeneration. In downstream hypothesis generation, a glia-centric emphasis (astrocyte and microglial disease programs with linked neuronal context) was prioritized to reduce diffuseness of whole-brain candidate space and improve mechanistic specificity of cross-cell signaling hypotheses.

Raw matrices were harmonized before inferential analysis. Gene symbols were standardized to a common symbol space, non-numeric artifacts were coerced and audited, duplicate symbols were collapsed at gene level, and genes with zero aggregate expression within a group were removed. This harmonization avoids inflation of low-information features and improves comparability across the disease and reference matrices. Because filtering is group aware, final matrices are comparable but not forced to have identical gene counts; this preserves biologically meaningful asymmetry after denoising.

The resulting PDAC matrices are summarized in Table 1. Disease-specific AD matrices were processed under the same normalization and filtering logic and then passed to AD PathFinder and candidate-pooling modules. These harmonized matrices provide the numerical substrate for differential-expression statistics and graph-based signaling inference in both disease branches.

Table 1: Summary of processed PDAC and AD expression matrices used for downstream analysis.

Dataset	Cell population	Genes	Samples	Matrix shape
PDAC	Acinar (normal-like)	17,250	201	$17,250 \times 201$
PDAC	Malignant (cancer-like)	18,790	214	$18,790 \times 214$
AD	Astrocyte (Disease)	41,149	176	$41,149 \times 176$
AD	Astrocyte (Normal)	41,149	176	$41,149 \times 176$
AD	Microglia (Disease)	41,149	139	$41,149 \times 139$
AD	Microglia (Normal)	41,149	139	$41,149 \times 139$
AD	Glutamatergic Neuron (Disease)	41,149	128	$41,149 \times 128$
AD	Glutamatergic Neuron (Normal)	41,149	128	$41,149 \times 128$

2.3 Differential Expression Priors

For each retained gene g , group means over malignant and acinar meta-cells were calculated and summarized by

$$\log_2 \text{FC}_g = \log_2 \left(\frac{\bar{x}_g^{(M)} + \epsilon}{\bar{x}_g^{(A)} + \epsilon} \right),$$

where ϵ is a small stabilization constant. Statistical significance was tested using two-sided Welch t -tests, followed by Benjamini-Hochberg correction across genes. Genes satisfying $|\log_2 \text{FC}_g| > 1$ and $\text{FDR}_g < 0.05$ were treated as differential-expression signals.

In this framework, DEG signals are not consumed directly as a final target list. Instead, they serve as quantitative priors for network learning in PathFinder. This distinction is important: the objective is not only to identify strongly shifted genes, but also to infer which parts of the signaling graph organize these shifts into coherent disease-associated routes. The same inferential principle was applied to AD with disease-vs-control contrasts defined in a neurodegeneration setting, while allowing disease-specific downstream filtering rules in candidate pooling.

2.4 LLM Retrieval and Curation

The text branch uses multiple LLM families as independent retrieval engines to reduce model-specific idiosyncrasy. Retrieval is implemented as a chained prompt workflow with strict schema constraints. The process begins by eliciting a disease-specific cell-type and subpopulation map at atlas-compatible granularity, with subpopulations grouped into biologically coherent compartments and accompanied by concise functional descriptions. This taxonomy output is then partitioned into chunks so that subsequent retrieval can be executed with bounded context per branch while preserving full coverage of major compartments in each disease setting.

For each cell-type chunk, the model is asked to retrieve three linked target classes in disease context: therapeutic targets, surface markers, and key driver genes. The output format is fixed as a structured table containing cell type, standardized target symbol, target class, short disease-function text, and evidence citation fields. Prompt constraints explicitly require literature support and disallow unsupported entries. Citation reporting is normalized to either PMID or author-year style so that each row can be audited independently. In a subsequent linked pass, the retrieved target table is used as input for pathway mapping, where targets are assigned to standardized pathway resources (for example KEGG, Reactome, or Hallmark-style nomenclature) with associated cell type, key member targets, biological effect, and citation fields.

Prompting is therefore schema-oriented rather than free-form, and each phase produces machine-parseable markdown tables instead of narrative prose. Branch-level outputs are merged, deduplicated, and normalized to HGNC-style symbols; unresolved or malformed entities are removed from the discovery layer. The final normalized text-derived set is denoted T_{LLM} . Each retained target keeps provenance attributes including model family, prompt lineage, cell-type annotation, target class, and linked citation metadata, so that downstream integration and hypothesis generation remain fully traceable to retrieval evidence. In AD instantiation, additional emphasis was placed on astrocyte-microglia-neuron interaction logic and state-transition language so that retrieval outputs better support intercellular mechanism construction rather than only intracellular cascade descriptions.

2.5 PathFinder Network Inference

PathFinder [11] is applied as the core network-inference engine in both disease branches. For PDAC, malignant epithelial meta-cells were used as case state and acinar cells as reference state. For AD, disease-control contrasts were instantiated over neurodegeneration-relevant cellular states and then integrated into a unified weighted intra-network view. In both settings, PathFinder operates on a prior biological interaction graph and learns path importance through graph-transformer supervision.

Let $G = (V, E)$ denote the prior graph and $\mathcal{P} = \{p_k\}$ the candidate signaling paths. In our implementation, path length is constrained to biologically plausible ranges (up to 10 hops) to avoid diffuse routes that are hard to interpret mechanistically. Node features are derived from expression profiles over meta-cells and encoded through graph-aware transformer layers; the original PathFinder configuration is preserved with six transformer layers, eight attention heads, and hidden dimension 16. Optimization follows the established training setup (learning rate 1×10^{-5} , L2 weight decay 3×10^{-7} , gradient clipping at norm 5.0, batch size 4, 25 epochs, fixed seed).

Path-level importance parameters α_k are learned jointly with phenotype classification. To incorporate expression contrast as a soft biological prior, each path is assigned

$$w_k = \frac{1}{|p_k|} \sum_{g \in p_k} |\log FC(g)|,$$

and training minimizes

$$\mathcal{L} = \mathcal{L}_{\text{cls}} + \lambda \sum_k (\alpha_k - w_k)^2.$$

The regularization term encourages consistency between learned path salience and DEG-derived signal magnitude while preserving predictive discrimination through

$\mathcal{L}_{\text{cls}}$. Importantly, the exported node weight is an integrated PathFinder importance score learned under DEG-informed constraints, rather than a direct one-to-one remapping of a single differential-expression statistic. As a result, the network weights retain quantitative linkage to disease-vs-reference transcriptomic contrast while also encoding pathway-level topological context learned during model fitting.

After fitting, high-importance paths are aggregated to produce graph artifacts in JSON format (full network and reduced-radius subnetworks). Exported nodes carry importance weights and exported edges carry relation weights. These weighted artifacts are not auxiliary visualizations; they are the primary omics representation consumed during candidate construction.

For AD specifically, we retained three cell-branch PathFinder exports and built an integrated weighted graph using observed cell-distribution proportions (glutamatergic : microglial : astrocyte = 128 : 139 : 176). Because branch topologies were identical while weights differed, weighted integration preserved shared structure and produced a single graph for downstream quadrant construction and bridge analysis.

2.6 Quadrant Candidate Construction

Candidate construction proceeds by integrating T_{LLM} with the set of genes represented in the PathFinder export, denoted V_{PF} . The integration is written as

$$Q1 = T_{\text{LLM}} \cap V_{\text{PF}}, \quad Q2 = T_{\text{LLM}} \setminus V_{\text{PF}}, \quad Q3 = V_{\text{PF}} \setminus T_{\text{LLM}}.$$

The meaning of each subset is biologically and methodologically distinct. Genes in $Q1$ are anchors jointly supported by semantic priors and network activation. Genes in $Q2$ retain text-prior support but are absent from the current network realization, so they are preserved as potentially context-conditional or underrepresented candidates rather than removed. Genes in $Q3$ are network-emergent novelty nodes that are not directly retrieved by the LLM branch.

For $Q1$, each entry carries both network and text metadata, including PathFinder node weight, LLM cell-type annotation, and functional class. For $Q2$, curation emphasizes entity validity and interpretable class assignment so these candidates can be incorporated explicitly during mechanism expansion. For $Q3$, ranking does not rely solely on node weight; it also quantifies bridge behavior relative to $Q1$ anchors.

Given a novelty node $u \in Q3$, we compute two bridge descriptors from network connectivity patterns: the number of anchor pairs linked through u (reported as `q1_pair_count`) and the number of unique anchors participating in those links (reported as `q1_unique_count`). We further preserve relation-level direction class (directed versus undirected-only) and compact path representations in detail tables. This produces a novelty ranking that favors nodes that are both weight-supported and topologically relevant to known anchor structure.

In AD, direct transfer of PDAC-style broad $Q2$ construction produced an overly large retrieval-only set relative to downstream interpretability. We therefore constrained $Q2$ through a stricter refinement trajectory. A connectivity-aware core was first extracted by requiring non-overlap retrieval targets to remain linked to $Q1$ anchors through pathway-supported reachability, prioritizing directed support where available and using undirected support as a controlled fallback. Because this strict core under-covered several biologically central AD programs, we then applied a capped layered supplement that prioritized targets recurring in astrocyte-, microglial-, and glutamatergic-relevant pathway contexts and recurring across

merged retrieval outputs. This preserved interpretability while avoiding broad re-expansion of retrieval-only noise.

Likewise, AD $Q3$ construction was based on explicit threshold profiling rather than a single hard cutoff. Raw novelty candidates were assembled from $Q1$ first/second-neighborhood nodes and intermediate nodes on $Q1$ - $Q1$ bridge paths. Each candidate was scored by PathFinder node weight, direct anchor-neighbor support, and bridge frequency. We compared strict, balanced, and loose regimes over these metrics and retained the balanced regime (high-weight with non-zero bridge support) as the final AD setting, because it preserved bridge explainability and yielded a tractable novelty set under weaker effect concentration than in PDAC.

The final candidate universe used for hypothesis generation is

$$\mathcal{C}_{\text{pool}} = Q1 \cup Q2 \cup Q3,$$

with explicit provenance tags retained for all elements. This provenance tagging enables every downstream hypothesis or therapy proposal to be traced back to its evidentiary origin.

2.7 Hypothesis and Strategy Generation

Mechanistic generation is conditioned on provenance-labeled candidates rather than on a flat merged list, and is executed through a progressive prompt chain aligned with the quadrant design. The initial mechanism pass receives $Q1$ anchors together with $Q2$ hidden-hub candidates and asks the model to assemble candidate signaling networks without introducing therapy content at this stage. Outputs are constrained to a structured mechanism table that records hypothesis identifiers, directed network sequences, and role decomposition into upstream nodes, core nodes, and downstream nodes, plus short biological rationale fields.

After this baseline mechanism layer is established, the model is supplied with $Q3$ novelty-node details and bridge-path constraints derived from the network artifacts. At this point, prompt instructions explicitly require consistency with provided topology metadata, including bridge direction class and path length. The model is asked to insert $Q3$ nodes at biologically plausible positions within existing signal flow and to treat longer bridge paths as potentially indirect links requiring mechanistic interpretation of intermediate processes. Updated hypotheses are again returned in structured table format so that pre- and post-novelty mechanism versions can be compared directly.

Only after upgraded mechanism tables are finalized does the workflow transition to strategy design. Therapy generation is constrained to select target combinations directly from the upgraded network sequences, typically using compact 2-4 node sets, and to provide a rationale for combinational logic and expected synergistic effect. Prompt constraints favor existing or repurposable drug classes when available, while preserving explicit linkage between each strategy and its parent mechanism hypothesis. Because provenance is retained throughout the chain, final strategy rows can be audited as intersection-supported, text-prior-supported, or network-emergent-supported, which is essential for transparent prioritization and downstream statistical validation.

For AD, the same staged prompting template was retained but with disease-tailored emphasis. Prompts de-emphasized purely tumor-style linear intracellular chains and instead encouraged mechanistic narratives centered on glial state transitions, neuroinflammatory feed-forward loops, and astrocyte-microglia-neuron

coupling. This adjustment was important because AD candidates are distributed across multiple interacting cell states; forcing a rigid linear style reduced biological plausibility and strategy traceability in pilot runs.

2.8 Disease-Matched Validation

Because available perturbation resources differ between oncology and neurodegeneration settings, validation was implemented as two disease-specific modules under a shared permutation-testing logic.

For PDAC, validation is conducted in DepMap using two complementary statistical views that address different levels of evidence. At target resolution, PDAC model identifiers are derived from `Model.csv` using pancreas/PDAC matching over Oncotree metadata fields. For each gene g , we compute

$$\Delta_g^{\text{effect}} = \mu_{\text{effect},g}^{\text{PDAC}} - \mu_{\text{effect},g}^{\text{all}}, \quad \Delta_g^{\text{dep}} = \mu_{\text{dep},g}^{\text{PDAC}} - \mu_{\text{dep},g}^{\text{all}}.$$

For a target set S , summary statistics are set means over Δ^{effect} and Δ^{dep} . Significance is estimated by same-size permutation against a shared gene universe (default 5000 draws). Effect is evaluated with a left-tail empirical test (more negative indicates stronger PDAC-selective vulnerability), and dependency is evaluated with a right-tail empirical test (more positive indicates stronger PDAC-selective dependency), both using $(+1)/(N+1)$ correction.

At PDAC strategy resolution, each therapy row is treated as one multi-gene strategy unit s with gene set G_s . We compute

$$E_s = \text{mean}(\Delta_g^{\text{effect}}, g \in G_s), \quad D_s = \text{mean}(\Delta_g^{\text{dep}}, g \in G_s),$$

then map to directional standardized scores

$$Z_E = -\frac{E_s - \mu_E}{\sigma_E}, \quad Z_D = \frac{D_s - \mu_D}{\sigma_D},$$

and combine them as

$$\text{Score}_s = w_E Z_E + w_D Z_D,$$

with default $w_E = w_D = 0.5$. Local significance is estimated via same-size permutation for each strategy, while global significance is estimated by permutation over the full strategy collection with observed strategy-size profile preserved.

For AD, validation is conducted in CRISPRbrain using AD-labeled versus nonAD-labeled screen registries. Screen-level gene scores are standardized within screen (default `abs.z` transform) and aggregated across included screens. For each gene g , we compute

$$\Delta_g^{\text{score}} = \mu_{\text{score},g}^{\text{AD}} - \mu_{\text{score},g}^{\text{nonAD}}, \quad \Delta_g^{\text{hit}} = \mu_{\text{hit},g}^{\text{AD}} - \mu_{\text{hit},g}^{\text{nonAD}},$$

where hit is derived from screen hit-class annotations (with threshold fallback when needed). Target-level significance is then assessed by same-size right-tail permutation for set means of Δ^{score} and Δ^{hit} .

At AD strategy resolution, each strategy s with gene set G_s is summarized by

$$E_s^{\text{AD}} = \text{mean}(\Delta_g^{\text{score}}, g \in G_s), \quad H_s^{\text{AD}} = \text{mean}(\Delta_g^{\text{hit}}, g \in G_s),$$

followed by global-universe standardization and composite scoring:

$$\text{Score}_s^{\text{AD}} = w_E z(E_s^{\text{AD}}) + w_H z(H_s^{\text{AD}}), \quad (w_E = w_H = 0.5 \text{ by default}).$$

As in PDAC, local same-size permutation and global profile-preserving permutation are both reported.

Together, these validation modules separate broad target-pool enrichment from combination-level actionability while preserving disease-appropriate perturbation semantics.

2.9 Drug-enabled Network Construction and Interactive Evidence Visualization

To make downstream outputs directly usable for therapeutic triage, we introduced a drug-enabled concordance layer that augments the mechanism backbone with evidence-linked compounds. The design preserves the same provenance-aware principle used throughout the framework: signaling routes remain the primary mechanistic object, while drug associations are attached as explicit evidence records rather than replacing route-level interpretation.

For each disease branch, construction integrates three synchronized inputs: (i) a curated Mermaid signaling topology (core target-target routes), (ii) a deduplicated disease-specific target-drug evidence ledger, and (iii) PathFinder-derived target weights from the corresponding weighted network export. Target nodes retain their signaling-role classes (for example overlap-supported, literature-emergent, LLM-supported, or contradicted) and inherit quantitative size scaling from PathFinder support. Because these PathFinder weights are learned under DEG-informed priors, the rendered node sizes remain anchored to disease-state transcriptomic contrast while preserving network-level context. When gene-level p -value tables are available, the same rendering interface can optionally provide a sensitivity-view using $-\log_{10}(p)$ scaling without changing edge semantics.

Drug nodes are generated by matching ledger targets to topology aliases, followed by deduplication at the target-drug-citation level. Each target-to-drug edge stores explicit evidence attributes, including evidence direction (support/mixed/contradict), development status (approved/investigational/preclinical/unknown), citation text, and hypothesis linkage fields. This schema keeps therapeutic interpretation auditable at edge resolution, avoiding loss of evidence polarity in flat target summaries.

The resulting graph bundle is exported in machine- and human-readable formats: structured JSON for downstream analysis, interactive HTML for node/edge inspection, and publication-oriented static renders. The interactive layer exposes target-level network evidence and drug-level therapeutic metadata in hover fields, allowing direct tracing from signaling context to literature-backed drug support.

In this study, we intentionally evaluated two evidence-input regimes. The core outputs (PDAC core and AD core) were seeded by manual literature-review tests before graph construction. PDAC core used one manually curated resistance-focused benchmark study [33]. AD core used four manually curated mechanistic exemplars spanning cGAS-STING neuroinflammation, ABCA7/NLRP3 glial signaling, CD74/MIF immune-metabolic modulation, and hippocampal glucose rescue [34–37].

We then stress-tested an automated expansion mode in PDAC. The extended PDAC output used LLM-assisted retrieval to aggregate a larger, continuously updateable literature pool into standardized node-evidence-edge-node records, and then merged that evidence graph with prior LLM mechanism hypotheses plus the deduplicated drug evidence ledger. This manual-seed versus automated-expansion contrast is retained as an explicit provenance axis in downstream result interpretation.

2.10 Reproducibility and Audit Trail

The full workflow is script-driven, seed-controlled, and artifact-oriented. Intermediate outputs are persisted in machine-readable formats at each integration boundary, including normalized LLM retrieval tables, PathFinder weighted JSON networks, quadrant tables (Q1/Q2/Q3), bridge-detail summaries, structured hypothesis and therapy tables, and disease-specific validation reports (DepMap for PDAC; CRISPRbrain registries and score summaries for AD) at both target and strategy levels.

This artifact lineage design supports independent re-analysis under alternative thresholds, model settings, or ranking criteria without rerunning the full pipeline from scratch. It also ensures that any final claim can be traced backward to the exact computational objects that generated it, which is particularly important in multi-module pipelines where evidence from language models, networks, and perturbation screens is fused.

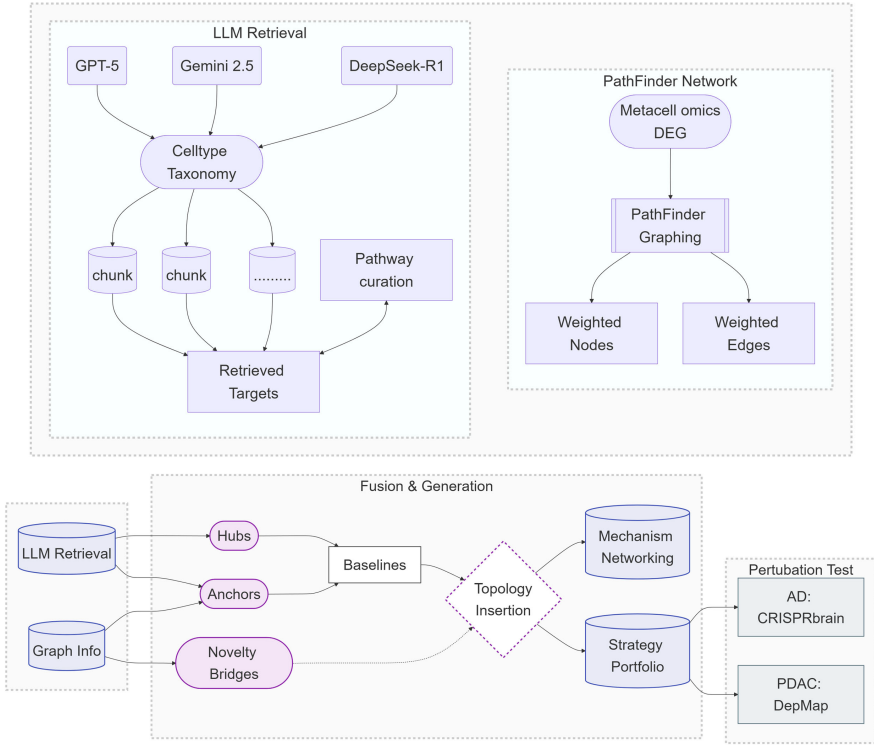


Fig. 1: Overview of the LLM-Omics provenance-aware target discovery framework. The workflow integrates semantic priors with cohort-level omics evidence across three decoupled phases. **(A) Dual-Modality Extraction:** The language branch (LLM Engine) constructs a structural prior (T_{LLM}) via schema-constrained chunked retrieval and pathway curation. In parallel, the omics branch (Omics Engine) transforms differential expression signals into a weighted biological network (V_{PF}) using PathFinder as an inferential analysis engine. **(B) Fusion and Generation:** Bypassing simple list intersection, the framework partitions candidates into distinct provenance quadrants: dual-supported anchors (Q1), retrieval-only hidden hubs (Q2), and network-emergent novelty bridges (Q3). During hypothesis generation, the LLM first assembles baseline mechanistic sequences using Q1 and Q2, and subsequently upgrades them by inserting Q3 nodes under explicit network topology constraints, yielding a structured strategy portfolio. **(C) Disease-Matched Validation:** The generated multi-target strategies are evaluated through permutation testing using DepMap for the PDAC cohort and CRISPRbrain registries for the AD cohort, ensuring independent functional support for the prioritized combinations.

3 Results

We report results in the same order as the pipeline: retrieval quality, provenance-aware candidate fusion, topology-constrained hypothesis expansion, and disease-matched perturbation validation. This organization separates stage-wise soundness checks from final actionability evidence and makes cross-disease comparisons explicit.

3.1 Retrieval Quality and Upstream Prior

The prompt-chained retrieval workflow generated a structured upstream prior that could be handed to the network-integration stage without additional manual rewriting. After normalization and curation, the retrieval-facing pool (defined as $Q1 \cup Q2$) contained 49 unique targets, each represented in table-form outputs that preserved biologically interpretable fields rather than free-text fragments. This structure is important for reproducibility: each target remains linked to a specific retrieval lineage and can be followed downstream into candidate construction, hypothesis expansion, and final strategy validation.

At this stage, we focused on soundness checks that test whether the retrieval layer is computationally usable before asking whether it is therapeutically useful. The key checkpoint was symbol-level interoperability with the downstream PathFinder node space. The curated retrieval pool could be projected into the network fusion step in a deterministic manner, indicating that this prior was sufficiently clean for provenance-aware integration rather than ad hoc post-processing.

A second checkpoint examined whether the retrieval layer was overly narrow. Instead of collapsing onto only one PDAC signaling axis, the curated set retained both canonical malignant regulators and context-linked components that become relevant when hypotheses are later expanded under bridge constraints. This balance prevented the candidate pipeline from becoming a simple confirmation exercise and created room for network-emergent novelty in later stages.

Table 2: Stage-wise cardinalities.

Pipeline layer	Count
Curated upstream prior ($Q1 \cup Q2$)	49
Network novelty reservoir ($Q3$)	26
Integrated candidate universe ($Q1 \cup Q2 \cup Q3$)	75
Upgraded mechanism coverage	19
Final strategy portfolio size	23
Final strategy target diversity	23

3.2 Candidate Pool Composition

Integrating the curated retrieval pool with PathFinder-derived network nodes yielded a 75-target candidate universe that remained balanced across provenance classes: 27 overlap-supported anchors in $Q1$, 22 retrieval-only hidden hubs in $Q2$, and 26 network-emergent novelty nodes in $Q3$. In relative terms, these correspond to 36.0%, 29.3%, and 34.7% of the final candidate universe, respectively. This near-balanced composition indicates that neither modality dominated candidate construction.

The biological profile of each quadrant also aligned with its intended role. The overlap quadrant ($Q1$) contained highly weighted disease-relevant anchors such as *MYC* (13871.35), *STAT3* (9115.33), and *EGFR* (2280.81), indicating convergence between text-derived priors and network-activated structure. The retrieval-only quadrant ($Q2$) retained plausible but currently non-projected entities (for example *KRAS*, *IL6*, *TBK1*, and *AURKA*), preserving potentially context-dependent hypotheses rather than discarding them prematurely. The network-only quadrant

(Q3) provided a dedicated novelty reservoir that was not reducible to simple retrieval overlap.

From a stage-wise reliability perspective, this candidate construction result is central to the method logic. The fused set is explicitly topology-aware and provenance-labeled, enabling downstream mechanism generation to reason over anchor certainty, hidden priors, and emergent bridges as distinct evidence types.

Table 3: Quadrant composition of the integrated candidate universe.

Quadrant	Count	Share	Representative nodes
Q1 (LLM $\cap$ PathFinder)	27	36.0%	MYC, STAT3, EGFR, VIM, TGFBR2
Q2 (LLM-only hubs)	22	29.3%	KRAS, IL6, TBK1, AURKA, ZEB1
Q3 (PathFinder-only)	26	34.7%	MAX, EGR1, CTCF, ELK1, MAZ

3.3 Q3 Bridge Topology

To test whether Q3 novelty nodes were mechanistically informative rather than peripheral artifacts, we quantified bridge statistics against Q1 anchors. Across all Q3 targets, the median `q1_pair_count` was 18.5 (range 3–94), and the median `q1_unique_count` was 15.5 (range 4–42). These values indicate that many novelty nodes participate in multiple anchor-anchor relationship patterns and therefore occupy topologically central, not merely terminal, positions in the inferred network.

The highest-ranked examples reinforced this interpretation. *MAX* combined the largest weight in Q3 (10415.66) with broad anchor bridging (63 pairs, 41 unique anchors), while *MAZ* achieved the strongest pair-level bridge signal (94 pairs, 42 unique anchors). *EGR1*, *CTCF*, and *ELK1* similarly showed mixed weight-topology support. In the context of hypothesis upgrading, these nodes are valuable because they provide concrete insertion points where network-informed novelty can be introduced into previously anchor-centered mechanistic narratives.

A practical implication is that novelty integration can be constrained rather than speculative. Because each Q3 candidate is accompanied by bridge statistics, downstream prompts can be instructed to prioritize nodes with high joint support (weight plus bridge connectivity), which reduces the risk of arbitrary expansion.

Table 4: Top Q3 novelty nodes prioritized by weight and bridge connectivity.

Q3 node	Node weight	Representative top anchor hubs
MAX	10415.66	APEX1, RARA, KIT
EGR1	9442.34	CXCR6, EGFR, SDC1
CTCF	7118.87	IL2RA, ADRB2, STAT3
ELK1	7101.45	ADRB2, IL2RA, EGFR
MAZ	6323.25	KIT, NCR1, VIM

3.4 Hypothesis Upgrading and Traceability

Mechanism generation began with $Q1 + Q2$ and was then upgraded by introducing $Q3$ bridge constraints. This sequence changed the role of novelty from optional decoration to structured mechanism expansion. The upgraded corpus referenced 19 unique **Upgraded-Hypo-*** identifiers and yielded 23 final strategy rows, indicating that expanded mechanism space was translated into concrete design outputs rather than remaining purely descriptive.

A critical stage-wise soundness criterion at this boundary is provenance closure: final strategy targets should remain within the integrated candidate universe. This criterion was fully satisfied. All 23 strategy targets were members of $Q1 \cup Q2 \cup Q3$, with a provenance mix of 12 from $Q1$ (52.2%), 2 from $Q2$ (8.7%), and 9 from $Q3$ (39.1%). The resulting pattern is informative. The strategy layer remained anchored to overlap-supported biology, but still incorporated substantial network-emergent novelty, consistent with the intended role of the upgraded hypothesis stage.

This balance suggests that the upgraded prompts did not over-privilege either conservative anchors or novelty nodes. Instead, provenance classes were translated into complementary mechanism roles, allowing strategies to remain interpretable while expanding beyond direct overlap candidates.

Table 5: Provenance composition of targets used in finalized strategies.

Source quadrant	Targets used	Share
$Q1$ anchors	12	52.2%
$Q2$ hidden hubs	2	8.7%
$Q3$ novelty bridges	9	39.1%

3.5 Strategy Portfolio Structure

The final therapy corpus produced 23 valid strategies from 23 parsed rows, so no candidate strategy was lost during structural parsing. Strategies were contributed by all three generation blocks (9, 7, and 7 rows), reducing sensitivity to any single prompt branch and improving internal robustness.

Strategy size distribution indicated a deliberately compact design space: 12 two-target strategies (52.2%), 10 three-target strategies (43.5%), and one four-target strategy (4.3%), with an average of 2.52 targets per strategy. This profile is useful for translation because it balances mechanistic coverage with practical combinational tractability.

Target recurrence patterns were coherent but not degenerate. *MYC* appeared in 8 strategies, while *EGFR* and *STAT3* each appeared in 7, and *AXL* appeared in 4. This indicates a consistent mechanistic center around receptor-transcription and stress-adaptation axes, yet without complete collapse into one repeated combination. Additional recurrence among *MAX* and *IL6* (three each) further supports a multi-axis architecture linking proliferative control, stromal signaling, and transcriptional adaptation.

Another stage-wise checkpoint asked whether statistically stronger strategies were concentrated in only one generation block. They were not. Locally significant

strategies (right-tail $p < 0.05$) were distributed across all blocks (4, 4, and 5), supporting the conclusion that the upgraded hypothesis layer produced broadly usable mechanism families rather than one dominant template.

Table 6: Therapy portfolio structure and recurrence profile.

Composition statistic	Value
Total valid strategies	23
Source blocks (Table-1 / Table-2 / Table-3)	9 / 7 / 7
Strategy size distribution (2 / 3 / 4 targets)	12 / 10 / 1
Average targets per strategy	2.52
Local significant strategies by block ($p < 0.05$)	4 / 4 / 5
Most frequent target	MYC (8 strategies)
Next recurrent targets	EGFR (7), STAT3 (7)

3.6 PDAC DepMap Validation

DepMap validation was performed on the finalized 23-target strategy universe across 69 PDAC models, and evaluated separately at target and strategy levels. This two-resolution design is important because it distinguishes general pool quality from combination-level therapeutic prioritization.

At target resolution, the effect metric showed a significant left shift relative to random same-size sets (left-tail empirical $p = 0.0039992$, $z = -3.2523$), while the dependency metric showed a significant right shift (right-tail empirical $p = 0.0361928$, $z = 2.0780$). The directionality is consistent with the intended interpretation of stronger PDAC-selective vulnerability and dependency in the curated target set.

At strategy resolution, enrichment was stronger and more decisive. The observed mean strategy score was 1.1400, compared with a permutation mean near zero (-0.0021), yielding $z = 8.9883$ and right-tail $p = 0.00019996$. Local testing identified 13 significant strategies out of 23 (56.5%), demonstrating that signal concentration was not limited to only a few outliers.

Top-ranked strategies were enriched for receptor-transcription couplings and dual-axis interventions, led by **T2_R1** (EGFR+MYC), **T1_R1** (ERBB2+EGFR+MYC), and **T3_R4** (AXL+MYC). At target directionality level, the strongest negative effect deltas were seen for *MYC*, *PPARG*, *CTCF*, and *MAX*, while strongest positive dependency deltas included *PPARG*, *ERBB2*, *EGFR*, and *MAX*. These directional leaders are concordant with the recurrent mechanism cores observed in the strategy portfolio.

3.7 PDAC Synthesis

Taken together, the PDAC pipeline produced a coherent evidence trajectory from structured retrieval to network fusion, from provenance-constrained hypothesis expansion to strategy-level functional validation. Each stage passed a local soundness checkpoint before feeding the next stage: retrieval outputs were computationally interoperable, candidate fusion remained modality-balanced, novelty nodes showed measurable bridge centrality, upgraded hypotheses preserved provenance closure, and final strategies showed both structural diversity and concentrated validation signal.

Table 7: Condensed DepMap validation summary.

Validation level	Statistic	Result
Target-level (effect)	Left-tail empirical p	0.0039992
Target-level (effect)	Directional z -score	-3.2523
Target-level (dependency)	Right-tail empirical p	0.0361928
Target-level (dependency)	Directional z -score	2.0780
Strategy-level (global)	Right-tail empirical p	0.00019996
Strategy-level (global)	Global z -score	8.9883
Strategy-level (local)	Significant strategies ($p < 0.05$)	13/23 (56.5%)

Table 8: Top strategies by strategy-level score.

Strategy ID	Target set	Score	Local p (right-tail)
T2_R1	EGFR, MYC	3.523	0.00340
T1_R1	ERBB2, EGFR, MYC	3.142	0.00240
T3_R4	AXL, MYC	2.741	0.00980
T1_R8	PPARG, MAP2K1, MYC	2.540	0.00520
T3_R1	EGFR, EGR1, MYC	2.320	0.00600

Table 9: Leading target-level directional signals from DepMap summaries.

Signal type	Top targets
Most negative effect deltas	MYC, PPARG, CTCF, MAX, ERBB2, EGFR, AXL, HSP90AA1
Most positive dependency deltas	PPARG, ERBB2, EGFR, MAX, MYC, HSP90AA1, AXL, CD4

These outcomes support the central methodological claim for the PDAC branch: PathFinder is most effective when treated as an analysis engine that reshapes candidate construction, rather than as a downstream verification add-on. Under this design, LLM retrieval contributes broad mechanistic priors, network inference contributes topology-aware evidence organization, and validation serves as a quantitative stress test of the jointly constructed strategy space.

3.8 AD Candidate Construction

Applying the same backbone workflow to AD produced a distinct but internally coherent signal profile. The merged AD retrieval layer contained 151 unique targets after deduplication and audit. PathFinder network artifacts for glutamatergic, microglial, and astrocyte branches were topology-consistent and were integrated into a weighted intra-network representation. Unlike PDAC, where direct fusion already produced a well-sized candidate universe, AD required additional control of candidate-space expansion to preserve interpretability under a weaker and more distributed disease signal.

The stage-wise contraction pattern makes this explicit. A broad retrieval-only AD pool remained large after initial pathway-aware projection, but strict connectivity filtering reduced it to a compact core that was then selectively expanded by a capped layered supplement. In parallel, novelty extraction started from a large bridge-candidate reservoir and was reduced by balanced topology constraints before final selection. This is reflected in final quadrant cardinalities: the AD branch produced a compact candidate universe of 34 genes, composed of 10 overlap-supported anchors, 14 retrieval-prior hidden hubs, and 10 network-emergent novelty nodes. No cross-quadrant overlap was retained in the finalized lists, preserving strict provenance separation. In AD context, this compactness is a feature rather than a deficit because it limits combinational explosion in downstream hypothesis and strategy generation.

Table 10: AD stage-wise cardinalities.

Pipeline layer	Count
Merged AD retrieval targets (deduplicated)	151
Integrated AD PathFinder graph size	469 nodes / 467 edges
Broad AD Q_2 (pre-refinement)	106
Strict AD Q_2 core	4
Q_1 anchors	10
Q_2 hidden hubs (strict+layered)	14
Q_3 novelty nodes (balanced)	10
Final AD candidate universe ($Q_1 \cup Q_2 \cup Q_3$)	34
Final AD strategy rows	14
Unique AD strategy targets	13

3.9 AD Novelty Bridge Structure

The AD novelty pool behaved differently from PDAC in both scale and topology. Bridge-candidate extraction initially produced 134 potential Q_3 nodes. Threshold profiling over strict, balanced, and loose regimes showed that strict filtering over-pruned mechanistically informative candidates, while loose filtering diluted interpretability with low-support tails. The balanced regime selected a final set of 10 nodes and preserved a high-signal novelty layer that remained manageable during hypothesis upgrading.

Relative to PDAC, AD novelty nodes showed lower bridge density overall, consistent with broader but weaker coupling structure in neurodegenerative state transitions. Within the selected AD Q_3 set, bridge-count median was 5 (range 1–13) and direct anchor-neighbor median was 0 (range 0–1), indicating that many novelty nodes contribute through intermediate bridges rather than immediate anchor adjacency. Even under this sparser topology, the selected nodes remained mechanistically useful because they formed reproducible bridge motifs linking inflammatory and stress-response anchors.

3.10 AD Hypothesis-to-Strategy Translation

The AD hypothesis layer produced a compact but diverse strategy portfolio. Fourteen strategy rows were retained after parsing, covering 13 unique targets and referencing

Table 11: AD selected novelty nodes and bridge statistics (balanced Q3).

Q3 node	Node weight	bridge_count	q1_neighbor_count
MYC	10169.20	1	0
MAX	8133.49	13	1
EGR1	7623.48	5	1
ELK1	4839.19	1	0
TP53	2796.30	5	1

11 unique **Upgraded-Hypo-*** identifiers. This indicates that AD hypotheses did not collapse into a single repetitive template despite tighter candidate-space control.

Provenance closure was fully maintained: every strategy target belonged to the finalized AD candidate universe, with target-source composition of 6 from Q1, 1 from Q2, and 6 from Q3. Thus, AD strategies remained anchored in overlap-supported biology while still incorporating network-emergent novelty. Recurrence patterns were biologically coherent, with *STAT3* as the dominant recurrent node (7 strategies), followed by *C3* and *PPARG* (5 each), and *EGR1/TP53* (4 each), consistent with an inflammatory-transcriptional core in the generated AD mechanism set.

Table 12: AD strategy architecture summary.

Composition statistic	Value
Valid AD strategies	14
Unique AD strategy targets	13
Referenced upgraded hypothesis IDs	11
Target provenance mix (Q1 / Q2 / Q3)	6 / 1 / 6
Strategy size distribution (2 / 3 / 4 targets)	5 / 6 / 3
Most recurrent targets	STAT3 (7), C3 (5)

3.11 AD CRISPRbrain Validation

AD validation was performed using CRISPRbrain with two registry settings: a default astrocyte ITC-focused profile and an expanded profile that additionally included neuron PSAP-KO contrasts and microglial immune-activation context. The expanded setting is treated as the primary AD validation view because it recovers full target coverage and better reflects AD-relevant cellular heterogeneity.

Under the expanded registry, target-level tests were significant on both axes (**delta_score** right-tail $p = 0.0346$, $z = 1.9152$; **delta_hit** right-tail $p = 0.0402$, $z = 1.7247$), while strategy-level global enrichment remained strongly significant ($p = 0.00019996$, $z = 6.4864$), with 7 of 14 strategies locally significant. Compared with default, expanded coverage increased from 10/13 to 13/13 targets. As expected, effect-size concentration on **delta_score** decreased when heterogeneity increased, while hit-consistency became stronger and remained significant.

At strategy resolution, the leading combinations were centered on an inflammatory-transcriptional axis and remained coherent with the upgraded hypothesis lineage. Top-ranked entries included EPHA4+STAT3 (**AD_THERAPY_014**), PPARG+STAT3 (**AD_THERAPY_012**), and STAT3+MYC (**AD_THERAPY_010**), each with

local right-tail significance below 0.01. The concentration of high-scoring strategies around STAT3-linked combinations is consistent with recurrence patterns observed in the AD strategy table and supports a non-random, mechanism-consistent prioritization signal despite attenuated target-level effect concentration relative to PDAC.

Table 13: AD CRISPRbrain validation summary: default vs expanded registry.

Metric	Default	Expanded
AD / nonAD screens	3 / 3	9 / 9
Usable targets / total	10 / 13	13 / 13
Target <code>delta_score</code> right-tail p	0.0062	0.0346
Target <code>delta_hit</code> right-tail p	0.2879	0.0402
Strategy global right-tail p	0.0002	0.0002
Strategy global z	5.6778	6.4864
Local significant strategies	7 / 13	7 / 14

3.12 Drug-enabled Network Outputs and Evidence Traceability

To evaluate whether the framework can move beyond target prioritization toward therapeutic triage, we generated three drug-enabled signaling outputs under two evidence regimes: PDAC core (manual-review seed), PDAC extended (automated retrieval expansion), and AD core (manual-review seed). All regenerated outputs remained topologically consistent with their mechanism backbones while adding explicit target-to-drug evidence edges. PDAC core contained 11 target nodes, 11 drug nodes, and 11 target-to-drug edges; PDAC extended expanded to 29 targets, 15 drugs, and 16 target-to-drug edges; and AD core contained 18 targets, 17 drugs, and 17 target-to-drug edges.

Evidence polarity was preserved at edge resolution rather than collapsed into binary support labels. In extended PDAC, target-to-drug edges were distributed as 12 support, 2 mixed, and 2 contradict links, while AD contained 8 support, 7 mixed, and 2 contradict links. This direction-aware representation is important for translational triage because it separates convergence-supported candidates from context-dependent or conflicting evidence patterns.

Layout diagnostics indicated that interactive outputs remained visually tractable. Estimated target-to-drug edge crossings were 0 in PDAC core, 2 in PDAC extended, and 0 in AD core despite increased extended-edge density. Representative renders are shown in Figures 2, 3, and 4. Together, these results establish an operational text-to-target-to-drug bridge in which signaling structure, quantitative weighting, and literature evidence are jointly inspectable in a provenance-preserving artifact, while also exposing a practical migration path from manual evidence seeding to automated literature expansion.

3.13 Cross-Disease Synthesis

Viewed jointly, PDAC and AD results delineate two operating regimes of the same workflow. In PDAC, the malignant-versus-reference contrast is sharper and

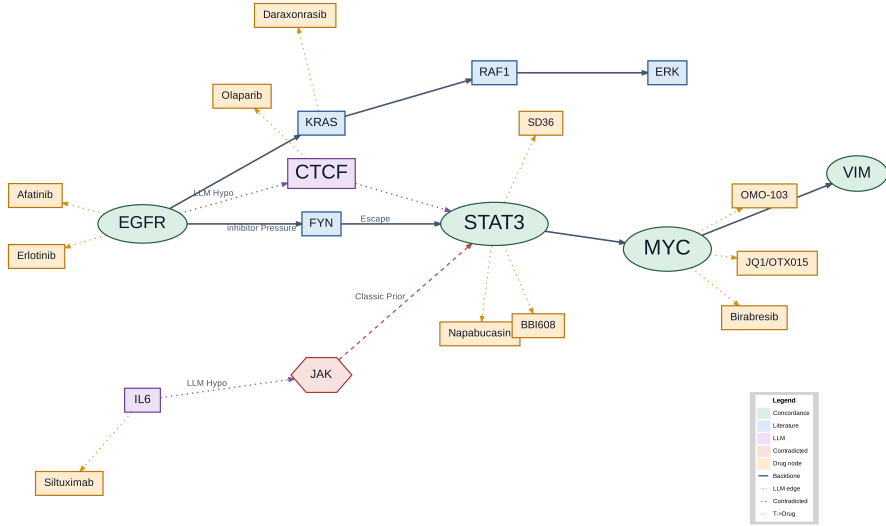


Fig. 2: PDAC core drug-enabled concordance network (manual-review-seeded baseline). This baseline output was constructed from manually curated literature-seed evidence. Target node size follows quantitative network support, and target-to-drug links preserve direction-aware literature evidence and development-stage metadata.

leads to a larger integrated candidate universe (75 genes), whereas AD requires stronger compression to maintain interpretability (34 genes). Despite this difference in scale, quadrant structure remains comparably balanced in both diseases (PDAC: $Q1/Q2/Q3 = 27/22/26$; AD: $10/14/10$), indicating that text priors, overlap-supported anchors, and network-emergent novelty all continue to contribute meaningful evidence rather than collapsing into one dominant source.

The downstream mechanism layer shows a similar pattern of shared structure with disease-specific intensity. PDAC generated 23 final strategies, and AD generated 14, both within a tractable combinational range. In both branches, provenance closure was preserved from candidate pool to strategy output, supporting the claim that hypothesis upgrading did not introduce untracked targets. The practical implication is that the framework can preserve mechanistic traceability even when disease-specific filtering pressure differs substantially.

Validation further clarifies where the two diseases diverge and where they converge. At target resolution, PDAC shows stronger directional separation, consistent with concentrated oncogenic dependence, while AD shows moderate but significant effects under the expanded CRISPRbrain setting, consistent with distributed neurodegenerative signal geometry. At strategy resolution, however, both branches exhibit strong non-random enrichment with identical global right-tail significance ($p = 0.00019996$). The corresponding global z -scores differ in magnitude (PDAC: 8.9883; AD: 6.4864), but both remain in a high-confidence range. Local significance rates are also comparable in practical terms (PDAC: 13/23; AD: 7/14), indicating that each branch yields an actionable strategy subset rather than isolated outliers.

supporting a KRAS–EGFR–STAT3 triple-axis strategy for suppressing refractory tumor growth and resistance [19, 25, 33]. This alignment indicates that the framework can identify clinically relevant control backbones before full experimental consolidation.

Concordance also appears at the mechanism level, not only at the target level. Generated sequences (for example, Upgraded-Hypo-15 and Table 3, Hypo-7) captured an EGFR-to-STAT3 relay and positioned STAT3 as a hub linked to mesenchymal plasticity and EMT-associated states through factors such as VIM and CTCF [2, 27]. These links are consistent with desmoplastic and invasive phenotypes in resistant PDAC, supporting the biological coherence of the inferred signaling backbone.

At the same time, retrospective comparison identifies an important resolution gap. Our earlier hypotheses emphasized the canonical IL6–JAK–STAT3 bridge, which is common in steady-state priors, whereas newer perturbation-focused studies in KRAS/EGFR-inhibited settings indicate an alternative SRC-family route, including FYN-mediated Tyr705 activation of STAT3 [19, 33, 38]. This mismatch illustrates “canonical inertia”: LLM priors can preserve broadly valid motifs while missing context-specific rewiring under treatment pressure.

The same retrospective lens becomes more informative when projected onto a drug-enabled layer. Instead of asking only whether a target reappears in later literature, we can evaluate whether target-linked compounds carry convergent, mixed, or contradictory support in the same mechanistic context. This extension improves translational interpretability: backbone-consistent routes with support-dominant drug links become higher-priority intervention candidates, whereas contradiction-heavy links are explicitly flagged for perturbation-specific re-evaluation before strategy escalation.

Topological Robustness in AD. In AD, retrospective mapping shows strong macro-level concordance between inferred topology and later mechanistic evidence. The workflow prioritized *STAT3* as a central transcriptional regulator linked to reactive astrogliosis and *C3* induction [23, 39, 40], consistent with studies describing IL-6/STAT3 as an amplifier of neuroinflammatory feed-forward loops [23, 24, 41]. It also highlighted a *PPARG*-centered metabolic-inflammatory coupling, concordant with reports connecting energy-state reprogramming to both astroglial toxicity and tau-related pathology [39, 42, 43].

However, as in PDAC, concordance at the backbone level does not guarantee molecular-level completeness. The model identified “Proteostasis” and “Redox” junctures as critical gaps but represented them as abstract placeholders. Recent studies provide finer-grained fillers, including mtDNA-associated *cGAS-STING/NLRP3* signaling for proteostatic stress and an *IDO1-KYN-AhR* axis for metabolic immune hijacking [23, 41, 44]. These comparisons suggest that Text-to-Target is effective for topology-aware triage, while perturbation-resolved updating remains necessary for high-resolution mechanism specification.

4.2 Prospective Design: A Continuous Literature-Mechanism Concordance Framework

The retrospective analysis identifies a consistent failure mode across diseases: stable backbone recovery with incomplete perturbation-specific detail. To address this gap, we propose a continuous Literature-Mechanism Concordance layer that replaces one-time retrospective checking with ongoing, perturbation-aware updating [45].

The key design choice is to use ordered mechanism edges, rather than flat target sets, as the operational unit. Each hypothesis is serialized into directed relations (for example, $A \rightarrow B$, $B \rightarrow C$) with explicit context tags (disease, cell type, perturbation mode, and directionality). A retrieval-augmented evidence pipeline continuously collects claim-level support for each edge [46, 47], and stores each evidence object in a normalized schema that records support polarity (support/contradict/mixed), experimental modality (in vitro, in vivo, cohort, perturbation screen), and temporal metadata.

Edge-level evidence is then aggregated into sequence-level concordance profiles using interpretable metrics such as support coverage, contradiction burden, and recency-weighted support strength. This scoring separates broadly supported sequences from conflict-prone or weakly evidenced alternatives, enabling explicit triage instead of binary keep/drop decisions. Mechanism sequences are also versioned; when new studies change edge support, updated releases include transparent change logs and citation deltas.

Under this design, hypothesis validation becomes a living process rather than a static endpoint. For fast-moving domains such as neuroinflammation and combination oncology, continuous 7-day/24-hour operation is not an engineering preference but a methodological requirement for maintaining translational relevance and mechanistic precision over time.

In this prospective setting, edge-level concordance can be propagated directly into target-to-drug evidence maintenance. When an updated study strengthens, weakens, or reverses support for a mechanism edge, linked therapeutic annotations can be refreshed in the same version cycle by updating evidence direction and citation fields on corresponding target-to-drug edges. This creates a unified update surface in which mechanistic and therapeutic revisions are synchronized rather than managed as separate post hoc layers.

Our current PDAC/AD comparison offers an initial design signal for this prospective loop: manual-review-seeded core outputs provide a high-precision bootstrap, while automated retrieval expansion increases coverage and route diversity when normalized into the same node-evidence-edge-node schema. This staged pattern directly motivates an agentic operating model in which human-curated seeds initialize the system and autonomous retrieval-refresh agents maintain breadth over time.

4.3 Agentic Outlook

A practical implementation is an always-on 7-day/24-hour agentic loop with explicit role separation across evidence acquisition, interpretation, and release control. In this architecture, literature watchers monitor new outputs; retrieval and extraction services convert documents into structured edge evidence; adjudication services update support status; hypothesis revision services regenerate mechanism sequences when contradictions accumulate; validation orchestration services trigger scheduled re-analysis; and release governance services publish only versioned snapshots that pass predefined quality gates. OpenClaw-style computer-use/tool-use modules can be added at the acquisition boundary to automate source monitoring, parsing, and citation normalization.

To reduce drift, this loop should run at multiple cadences. Fast cycles (6–24 hours) update evidence ledgers and contradiction flags, medium cycles (weekly) revise or prune hypothesis branches, and slow cycles (monthly) freeze stable releases for benchmarking. Progress should be tracked longitudinally with concordance gain per cycle,

hypothesis stability under refresh, novelty-to-support conversion, and incremental changes in downstream validation statistics.

This outlook aligns with the broader shift from one-shot prompting to tool-grounded iterative scientific agents [48–50]. In our setting, continuous operation is valuable only insofar as it improves mechanism fidelity and translational utility as the evidence landscape changes.

4.4 Limitations and Future Work

The current framework has clear scope limitations. Transcriptomic signals support association-driven prioritization but do not establish causal sufficiency; post-transcriptional regulation, protein activity states, and spatial microenvironment effects remain incompletely modeled. In addition, AD validation depends on registry composition and assay heterogeneity, whereas PDAC validation may underrepresent microenvironment-mediated vulnerabilities when epithelial contrasts dominate.

Continuous 7-day/24-hour operation also creates new methodological risks, including retrieval-noise accumulation, citation leakage, policy drift across model updates, and self-reinforcing loops around highly cited but context-mismatched mechanisms. For this reason, governance is not optional infrastructure: it is part of the method definition.

The drug-enabled layer has additional practical boundaries in its current form. Node-size rendering is currently based on PathFinder importance weights, which are learned under DEG-informed constraints and therefore remain quantitatively linked to transcriptomic differential signals, but are not a direct visualization of a single gene-level statistic such as $-\log_{10}(p)$. Interactive displays also expose citation granularity more strongly at edge level than at node summary level. Although full evidence objects are retained in JSON, manuscript-facing evidence digest tables still require explicit export. These limitations are tractable but should be addressed to strengthen reviewer-facing translational reporting.

Future work will focus on four directions: perturbation-coupled prospective feedback, spatial/proteomic augmentation, refreshable edge-level literature concordance, and governance with deterministic checkpoints and release policies. These upgrades would move Text-to-Target from high-quality hypothesis triage toward continuously improving, mechanism-auditable therapeutic discovery.

Acknowledgment

Funding: This research was partially supported by NIA 4R33AG078799-02, NLM 1R01LM013902-01A1, NIA R56AG065352, NIA 1R21AG078799-01A1.

References

- [1] Rood, J.E., Wynne, S., Robson, L., Hupalowska, A., Randell, J., Teichmann, S.A., Regev, A.: The human cell atlas from a cell census to a unified foundation model. *Nature*, 1–2 (2024)
- [2] Hosein, A.N., Brekken, R.A., Maitra, A.: Pancreatic cancer stroma: an emerging therapeutic target. *Nature Reviews Clinical Oncology* **17**, 487–505 (2020)

- [3] Gottweis, J., Weng, W.-H., Daryin, A., Tu, T., Palepu, A., Sirkovic, P., Myaskovsky, A., Weissenberger, F., Rong, K., Tanno, R., et al.: Towards an ai co-scientist. arXiv preprint arXiv:2502.18864 (2025)
- [4] Mitchener, L., Yiu, A., Chang, B., Bourdenx, M., Nadolski, T., Sulovari, A., Landsness, E.C., Barabasi, D.L., Narayanan, S., Evans, N., et al.: Kosmos: An ai scientist for autonomous discovery. arXiv preprint arXiv:2511.02824 (2025)
- [5] Li, F., Zhang, H., Huang, D., Li, H., Li, W., Xu, T., Chen, Y., Province, M., Payne, P.R.: Ai for scientific discovery in omics data-driven precision medicine. *Missouri Medicine* **123**(1), 55 (2026)
- [6] Zhang, H., Xu, T., Cao, D., Liang, S., Schimmelpfennig, L., Kaster, L., Huang, D., Cruchaga, C., Li, G., Province, M., Chen, Y., Payne, P., Li, F.: Omnicelltosg: The first cell text-omic signaling graphs dataset for joint llm and gnn modeling. arXiv preprint arXiv:2504.02148 (2025) [cs.AI]. <https://doi.org/10.48550/arXiv.2504.02148>
- [7] Huang, D., Li, H., Li, W., Zhang, H., Dickson, P., Zhan, M., Miller, J.P., Cruchaga, C., Province, M., Chen, Y., Payne, P., Li, F.: Omnicellagent: Towards ai co-scientists for scientific discovery in precision medicine. bioRxiv (2025) <https://www.biorxiv.org/content/early/2025/08/04/2025.07.31.667797.full.pdf>. <https://doi.org/10.1101/2025.07.31.667797>
- [8] Zhang, H., Liang, S., Xu, T., Li, W., Huang, D., Dong, Y., Li, G., Miller, J.P., Goedegebuure, S.P., Sardiello, M., et al.: Biomedgraphica: An all-in-one platform for biomedical prior knowledge and omic signaling graph generation. bioRxiv, 2024–12 (2024)
- [9] Zhang, H., Huang, D., Li, W., Province, M., Chen, Y., Payne, P., Li, F.: Galax: Graph-augmented language model for explainable reinforcement-guided subgraph reasoning in precision medicine. arXiv preprint arXiv:2509.20935 (2025)
- [10] Zhang, H., Goedegebuure, S.P., Ding, L., DeNardo, D., Fields, R.C., Province, M., Chen, Y., Payne, P., Li, F.: M3netflow: A multi-scale multi-hop graph ai model for integrative multi-omic data analysis. *iScience*
- [11] Feng, J., Song, H., Province, M., Li, G., Payne, P.R., Chen, Y., Li, F.: Pathfinder: a novel graph transformer model to infer multi-cell intra-and inter-cellular signaling pathways and communications. *Frontiers in Cellular Neuroscience* **18**, 1369242 (2024)
- [12] Dong, Z., Zhao, Q., Payne, P.R., Province, M.A., Cruchaga, C., Zhang, M., Zhao, T., Chen, Y., Li, F.: Highly accurate disease diagnosis and highly reproducible biomarker identification with pathformer. *Research Square*, 3 (2023)
- [13] Zhang, H., Cao, D., Xu, T., Chen, E., Li, G., Chen, Y., Payne, P., Province, M., Li, F.: mosgraphflow: a novel integrative graph ai model mining disease targets from multi-omic data. bioRxiv (2024)

- [14] Regev, A., Teichmann, S.A., Lander, E.S., *et al.*: The human cell atlas. *eLife* **6**, 27041 (2017)
- [15] Bender, E.M., Gebru, T., McMillan-Major, A., Shmitchell, S.: On the dangers of stochastic parrots: Can language models be too big? In: *Proceedings of FAccT*, pp. 610–623 (2021)
- [16] Persad, S., Choo, Z.-N., Dien, C., Sohail, N., Masilionis, I., Chaligné, R., Nawy, T., Brown, C.C., Sharma, R., Pe’er, I., *et al.*: Seacells infers transcriptional and epigenomic cellular states from single-cell genomics data. *Nature Biotechnology* **41**(12), 1746–1757 (2023)
- [17] Siegel, R.L., Giaquinto, A.N., Jemal, A.: Cancer statistics, 2024. *CA: A Cancer Journal for Clinicians* **74**(1), 12–49 (2024). <https://doi.org/10.3322/caac.21820>. PMID: 38230766
- [18] Rahib, L., Smith, B.D., Aizenberg, R., Rosenzweig, A.B., Fleshman, J.M., Matrisian, L.M.: Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the united states. *Cancer Research* **74**(11), 2913–2921 (2014). <https://doi.org/10.1158/0008-5472.CAN-14-0155>. PMID: 24480245
- [19] Khan, N., *et al.*: Drugging the ‘undruggable’ kras: Breakthroughs, challenges, and opportunities in pancreatic cancer. *Cancer Biol Med* (2025). <https://doi.org/10.20892/j.issn.2095-3941.2025.0122>
- [20] 2025 alzheimer’s disease facts and figures. *Alzheimer’s & Dementia* **21**(4), 70235 (2025)
- [21] Jia, J., Ning, Y., Chen, M., Wang, S., Yang, H., Li, F., Ding, J., Li, Y., Zhao, B., Lyu, J., *et al.*: Biomarker changes during 20 years preceding alzheimer’s disease. *New England Journal of Medicine* **390**(8), 712–722 (2024)
- [22] Li, F., Chen, Y., Western, D., Ali, M., Liu, M., Gong, K., Xu, Y., Lowery, J., Holtzman, D.M., Robins, C., *et al.*: Apoe-stratified proteomic and metabolomic analysis reveals mitochondrial dysfunction inflammation and lipid dysregulation in alzheimer’s disease. *Advanced Science*, 13872 (2026)
- [23] Li, F., Eteleeb, A.M., Buchser, W., Sohn, C., Wang, G., Xiong, C., Payne, P.R., McDade, E., Karch, C.M., Harari, O., *et al.*: Weakly activated core neuroinflammation pathways were identified as a central signaling mechanism contributing to the chronic neurodegeneration in alzheimer’s disease. *Frontiers in Aging Neuroscience* **14**, 935279 (2022)
- [24] Li, F., Oh, I., Kumar, S., Eteleeb, A., Gupta, A., Buchser, W., Xiong, C., Cole, S.F., McDade, E., Karch, C.M., *et al.*: Loss of estrogen unleashing neuro-inflammation increases the risk of alzheimer’s disease in women. *bioRxiv*, 2022–09 (2022)
- [25] Drizyte-Miller, K., Talabi, T., Somasundaram, A., Cox, A.D., Der, C.J.: Kras: the achilles’ heel of pancreas cancer biology. *Journal of Clinical Investigation*

- 135**(16), 191939 (2025). <https://doi.org/10.1172/JCI191939>
- [26] Human Cell Atlas Consortium: Human Cell Atlas Data Portal: Pancreas Atlas. <https://explore.data.humancellatlas.org/projects/b3938158-4e8d-4fdb-9e13-9e94270dde16>. Accessed: 2025-01-01 (2024)
- [27] Zhou, J., Melms, J.C., Ghabri, G., *et al.*: Single-nucleus and spatial transcriptome profiling of pancreatic cancer identifies multicellular dynamics associated with neoadjuvant treatment. *Nature Genetics* **54**, 1368–1378 (2022). <https://doi.org/10.1038/s41588-022-01134-8>
- [28] Feng, J., Goedegebuure, S.P., Zeng, A., Bi, Y., Wang, T., Payne, P., Ding, L., DeNardo, D., Hawkins, W., Fields, R.C., *et al.*: sc2menetdrug: A computational tool to uncover inter-cell signaling targets and identify relevant drugs based on single cell rna-seq data. *PLoS computational biology* **20**(1), 1011785 (2024)
- [29] OpenAI: Early experiments in accelerating science with gpt-5. OpenAI Research Report (2025). Available from OpenAI website
- [30] Li, X., Wang, Y., *et al.*: Capabilities of gpt-5 on multimodal medical reasoning. arXiv preprint arXiv:2508.08224 (2025) [cs.CL]
- [31] Comanici, G., Bieber, E., Schaekermann, M., *et al.*: Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. arXiv preprint arXiv:2507.06261 (2025) [cs.CL]
- [32] DeepSeek-AI, Guo, D., *et al.*: Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. arXiv preprint arXiv:2501.12948 (2025) [cs.CL]
- [33] Liaki, V., *et al.*: A targeted combination therapy achieves effective pancreatic cancer regression and prevents tumor recurrence. *Proceedings of the National Academy of Sciences* (2025). <https://doi.org/10.1073/pnas.2523039122>
- [34] Fazal, F., Dar, N.J., Ahamad, S., Khan, S., Bano, N., Saha, S., Nazir, A., Bhat, S.A.: cgas-sting signaling in alzheimer’s disease: Microglial mechanisms and therapeutic opportunities. *Molecular Aspects of Medicine* **107**, 101444 (2026). <https://doi.org/10.1016/j.mam.2025.101444>
- [35] Santos-Garcia, I., Bascunana, P., Brackhan, M., Villa, M., Eiriz, I., Bruning, T., Pahnke, J.: The abc transporter a7 modulates neuroinflammation via nlrp3 inflammasome in alzheimer’s disease mice. *Alzheimer’s Research & Therapy* **17**, 30 (2025). <https://doi.org/10.1186/s13195-025-01673-2>
- [36] Gladen-Kolarsky, N., Neff, C.J., Hack, W., Brandes, M.S., Wiedrick, J., Meza-Romero, R., Lockwood, D.R., Quinn, J.F., Offner, H., Vandenbark, A.A., Gray, N.E.: The cd74 inhibitor drhq improves short-term memory and mitochondrial function in 5xfad mouse model of $\alpha\beta$ accumulation. *Metabolic Brain Disease* **40**(1), 95 (2025). <https://doi.org/10.1007/s11011-024-01433-5>
- [37] Minhas, P.S., Jones, J.R., Latif-Hernandez, A., Sugiura, Y., Durairaj, A.S., Wang, Q., Mhatre, S.D., Uenaka, T., Crapser, J., Conley, T., Ennerfelt, H., Jung,

- Y.J., Liu, L., Prasad, P., Jenkins, B.C., Ay, Y.A., Matrongolo, M., Goodman, R., Newmeyer, T., Heard, K., Kang, A., Wilson, E.N., Yang, T., Ullian, E.M., Serano, G.E., Beach, T.G., Wernig, M., Rabinowitz, J.D., Suematsu, M., Longo, F.M., McReynolds, M.R., Gage, F.H., Andreasson, K.I.: Restoring hippocampal glucose metabolism rescues cognition across alzheimer disease pathologies. *Science* **385**(6711), 6131 (2024). <https://doi.org/10.1126/science.abm6131>
- [38] Mellenthin, C., Schaller, F., Vignot, L., Peisl, S., Vesce, R., Esposito, I., Gonelle-Gispert, C., Schneider, C., Adamina, M., Bühler, L.H.: Il-6/gp130/jak/stat3 pathway activation in pancreatic ductal adenocarcinoma and its association with clinical features: Protocol for a retrospective cross-sectional study. *JMIR Research Protocols* **14**(e76811) (2025). <https://doi.org/10.2196/76811>
- [39] Mathys, H., Peng, Z., Boix, C.A., Victor, M.B., Leary, N., Babu, S., Abdelhady, G., Jiang, X., Ng, A.P., Ghafari, K., *et al.*: Single-cell atlas reveals correlates of high cognitive function, dementia, and resilience to alzheimer’s disease pathology. *Cell* **186**(20), 4365–4385 (2023)
- [40] Gabitto, M.I., Travaglini, K.J., Rachleff, V.M., Kaplan, E.S., Long, B., Ariza, J., Ding, Y., Mahoney, J.T., Dee, N., Goldy, J., *et al.*: Integrated multimodal cell atlas of alzheimer’s disease. *Nature Neuroscience* **27**(12), 2366–2383 (2024)
- [41] Mackiewicz, J., Lisek, M., Boczek, T.: Targeting can/ncf1 in alzheimer’s brain degeneration. *Frontiers in Immunology* **Volume 14 - 2023** (2023). <https://doi.org/10.3389/fimmu.2023.1281882>
- [42] Wang, L., Nykänen, N.-P., Western, D., Gorijala, P., Timsina, J., Li, F., Wang, Z., Ali, M., Yang, C., Liu, M., *et al.*: Proteo-genomics of soluble trem2 in cerebrospinal fluid provides novel insights and identifies novel modulators for alzheimer’s disease. *Molecular neurodegeneration* **19**(1), 1 (2024)
- [43] Shen, Y., Timsina, J., Heo, G., Beric, A., Ali, M., Wang, C., Yang, C., Wang, Y., Western, D., Liu, M., *et al.*: Csf proteomics identifies early changes in autosomal dominant alzheimer’s disease. *Cell* **187**(22), 6309–6326 (2024)
- [44] Roy, E.R., Chiu, G., Li, S., Propson, N.E., Kanchi, R., Wang, B., Coarfa, C., Zheng, H., Cao, W.: Concerted type i interferon signaling in microglia and neural cells promotes memory impairment associated with amyloid plaques. *Immunity* **55**(5), 879–8946 (2022). <https://doi.org/10.1016/j.immuni.2022.03.018>
- [45] Gao, S., Fang, A., Huang, Y., Giunchiglia, V., Noori, A., Schwarz, J.R., Ektefaie, Y., Kondic, J., Zitnik, M.: Empowering biomedical discovery with ai agents. *Cell* **187**(22), 6125–6151 (2024). <https://doi.org/10.1016/j.cell.2024.09.022>
- [46] Liu, S., McCoy, A.B., Wright, A.: Improving large language model applications in biomedicine with retrieval-augmented generation: a systematic review, meta-analysis, and clinical development guidelines. *Journal of the American Medical Informatics Association* **32**(4), 605–615 (2025) <https://academic.oup.com/jamia/article-pdf/32/4/605/61442713/ocaf008.pdf>. <https://doi.org/10.1093/jamia/ocaf008>

- [47] Li, M., Kilicoglu, H., Xu, H., Zhang, R.: Biomedrag: A retrieval augmented large language model for biomedicine. *Journal of Biomedical Informatics* **162**, 104769 (2025). <https://doi.org/10.1016/j.jbi.2024.104769>
- [48] Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., Cao, Y.: ReAct: Synergizing Reasoning and Acting in Language Models (2023). <https://arxiv.org/abs/2210.03629>
- [49] Wu, Q., Bansal, G., Zhang, J., Wu, Y., Li, B., Zhu, E., Jiang, L., Zhang, X., Zhang, S., Liu, J., Awadallah, A.H., White, R.W., Burger, D., Wang, C.: AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation (2023). <https://arxiv.org/abs/2308.08155>
- [50] Moritz, M., Topol, E., Rajpurkar, P.: Coordinated ai agents for advancing healthcare. *Nature Biomedical Engineering* **9**(4), 432–438 (2025). <https://doi.org/10.1038/s41551-025-01363-2>

5 Supplementary Tables

Table S1: PDAC cell-specific targets

Cell Type	Target Symbol	Target Class	Evidence Level & Citation
Acinar	EGFR	Receptor Tyrosine Kinase	Genetic requirement in KRAS PDAC initiation;PMID: 22975374
Acinar	NR5A2	Nuclear Receptor (Transcription Factor)	Functional constraint of Kras-driven neoplasia;PMID: 23645620
Acinar	PRSS1	Secreted Enzyme (Marker)	PDAC spatial/single-cell acinar markers;PMID: 35995947
Acinar_Cell	AMY2A	Enzyme	PMID: 31739036 (Qian et al., 2020)
Acinar_Cell	BHLHA15	Transcription Factor	PMID: 23610129 (Shi et al., 2013)
Acinar_Cell	PTF1A	Transcription Factor	PMID: 21245155 (Krah et al., 2011)
Activated/ angiogenic EC	NRP1	Co-receptor	PMID: 26804176 (Zhou et al., 2016)
Activated/ angiogenic EC	TEK	Receptor tyrosine kinase	Roos-Mattila et al.,2023
Activated/ angiogenic EC	VEGFA	Secreted ligand	Craven et al.,2015
ADM-like/ acinar-like metaplastic	EXTL3	Surface Receptor/ Enzyme	REG3A/REG3B→EXTL3 axis; PMID: 34006485
ADM-like/ acinar-like metaplastic	NFATC4	Transcription Factor	Drives acinar-to-ductal conversion and initiation;PMID: 26697077
ADM-like/ acinar-like metaplastic	REG3A	Secreted Lectin-like Protein	Mechanistic ADM promotion; PMID: 34006485
ADM-like/ acinar-like metaplastic	REG3B	Secreted Lectin-like Protein	Mechanistic ADM promotion; PMID: 34006485
Adrenergic/ Neural-like Progenitor Cells	ADRB2	GPCR	PMID: 27624061 (Partecke et al., 2016)
Adrenergic/ Neural-like Progenitor Cells	NTRK2	Receptor Tyrosine Kinase	PMID: 31748746 (Maurer et al., 2019)
Adrenergic/ Neural-like Progenitor Cells	TH	Enzyme	PMID: 34808793 (Li et al.,2021)

Continued on next page

Cell Type	Target Symbol	Target Class	Evidence Level & Citation
Antigen-presentation-competent malignant	B2M	Antigen Presentation Component	Antigen-presentation malignant programs in PDAC atlases; PMID: 41278823
Antigen-presentation-competent malignant	CD74	MHC-II Chaperone (Surface)	Tumor-intrinsic MHC-II activation study;PMID: 41278823
Antigen-presentation-competent malignant	HLA-DRA	MHC-II (Surface)	scRNA+spatial + pharmacologic induction;PMID: 41278823
Antigen-presentation-competent malignant	MAP2K1	Kinase	Cobimetinib induction of MHC-II in PDAC;PMID: 41278823
Antigen-presentation-competent malignant	MAP2K2	Kinase	Cobimetinib induction of MHC-II in PDAC;PMID: 41278823
antigen-presenting Cancer-Associated Fibroblasts (apCAFs)	SLPI	Secreted Peptidase Inhibitor	Validated (Elyada et al.,2019; PMID: 31197017)
apCAF	IDO1	Enzyme	Preclinical;Li et al.,2021 (PMID: 34127627)
apCAF	TGFBR2	Receptor kinase	PMID: 35523176 (Huang et al., 2022)
Arterial EC	EFNB2	Ligand	Shiau et al.,2022
Arterial EC	GJA5	Gap junction	Shiau et al.,2022
Arterial EC	SOX17	Transcription factor	Shiau et al.,2022
B cell (naïve/ memory)	BCL6	Transcription factor (target)	PMID: 36099917
B cell (naïve/ memory)	CD19	Surface marker	PMID: 36099917
B cell (naïve/ memory)	EBI3	Cytokine sub-unit (IL-35)	PMID: 36099917
B cell (naïve/ memory)	IL12A	Cytokine sub-unit (IL-35)	PMID: 36099917
B_Cell	MS4A1	Surface Protein	PMID: 34101736
Basal-like/ squamous malignant	AXL	Receptor Tyrosine Kinase	Therapeutic inhibition in PDAC models;PMID: 292???? (Cancer Res 2018)(AACR Journals)
Basal-like/ squamous malignant	KRT5	Intermediate Filament (Marker)	Basal-like marker framework; TP63-linked squamous program; PMID: 30428345

Continued on next page

Cell Type	Target Symbol	Target Class	Evidence Level & Citation
Basal-like/ squamous malignant	TACSTD2	Surface Glycoprotein	Large-cohort PDAC expression/ targetability;PMID: 36878461
Basal-like/ squamous malignant	TP63	Transcription Factor	Mechanistic (enhancer repro- gramming);PMID: 30428345
Capillary EC	FLT1	Receptor tyrosine kinase	Craven et al.,2015
Capillary EC	KDR	Receptor tyrosine kinase	PMID: 22207048 (Doi et al., 2012)
Capillary EC	PECAM1	Surface adhesion	Shiau et al.,2022 (PDAC TME remodeling)
Capillary EC	VWF	Secreted glycoprotein	Shiau et al.,2022
CD4 helper T	CD4	Surface marker	PMID: 33803936
CD4 helper T	ICOS	Costimulatory receptor	PMID: 33803936
CD4 helper T	TBX21	Transcription factor	PMID: 33803936
CD4+ T Cells - Naive/Central Memory	LEF1	Transcription Factor	PMID: 32472114 (Sade-Feldman et al.,2019)
CD4+ T Cells - Regulatory T Cells (Tregs)	IL2RA	Cytokine Receptor	PMID: 28783719 (Jang et al., 2017)
CD4+ T Cells - Regulatory T Cells (Tregs)	TNFRSF18	TNF Receptor Superfamily	PMID: 32472114 (Sade-Feldman et al.,2019)
CD4+ T Cells - T helper 1 (Th1)	TNF	Cytokine	PMID: 29061364 (Balachandran et al.,2017)
CD8 cytotoxic effector T	CD8A	Surface marker	PMID: 33803936
CD8 cytotoxic effector T	PDCD1	Immune checkpoint receptor	PMID: 35480107
CD8 cytotoxic effector T	TNFRSF9	Costimulatory receptor	PMID: 35480107
CD8 exhausted T	HAVCR2	Immune checkpoint receptor	PMID: 34072549
CD8 exhausted T	LAG3	Immune checkpoint receptor	PMID: 33803936
CD8 exhausted T	PVR	Checkpoint ligand	PMID: 34358448
CD8 exhausted T	TIGIT	Immune checkpoint receptor	PMID: 37140899

Continued on next page

Cell Type	Target Symbol	Target Class	Evidence Level & Citation
CD8 exhausted T	VSIR	Immune checkpoint receptor	PMID: 34072549
CD8 Trm-like	CXCR6	Chemokine receptor	PMID: 36689623
CD8 Trm-like	ITGAE	Surface marker (integrin)	PMID: 36689623
CD8 Trm-like	ZNF683	Transcription factor	PMID: 36689623
CD8+ T Cells - Cytotoxic/Effector	IFNG	Cytokine	PMID: 26200351 (Stromnes et al.,2015)
CD8+ T Cells - Cytotoxic/Effector	PRF1	Pore-Forming Protein	PMID: 32472114 (Sade-Feldman et al.,2019)
CD8+ T Cells - Exhausted (Tex)	TOX	Transcription Factor	PMID: 33649582 (Sivakumar et al.,2023)
CD8+ T Cells - Naive/Memory-like	IL7R	Cytokine Receptor	PMID: 31331973 (Steele et al., 2020)
CD8+ T Cells - Naive/Memory-like	TCF7	Transcription Factor	PMID: 32472114 (Sade-Feldman et al.,2019);PMID: 35614219 (Bear et al.,2020)
cDC1	BATF3	Transcription factor	PDAC DC-centered therapy study;PMID: 40815670 (Hogg et al.,2025).
cDC1	CD40	Costimulatory receptor	Human+mouse PDAC;PMID: 21436454 (Beatty et al.,2011).
cDC1	CLEC9A	C-type Lectin Receptor	Validated (Elyada et al.,2019; PMID: 31570890)[Atlas]
cDC1	FLT3	Receptor Tyrosine Kinase	PMID: 32243813
cDC1	IRF8	Transcription Factor	Validated (Meyer et al.,2018; PMID: 29915268)[Mechanistic in PDAC]
cDC1	XCR1	Surface marker (chemokine receptor)	PDAC DC-centered therapy study;PMID: 40815670 (Hogg et al.,2025).
cDC2	CD1C	Surface marker	PDAC scRNA integration;PMID: 37633924 (Oh et al.,2023).
cDC2	CLEC10A	C-type Lectin Receptor	Validated (Elyada et al.,2019; PMID: 31570890)[Atlas]
cDC2	IRF4	Transcription factor	PDAC DC-centered therapy study;PMID: 40815670 (Hogg et al.,2025).
Classical monocyte	CCR2	Chemokine receptor	Phase 1b PDAC trial;PMID: 27055731 (Nywening et al.,2016).
Classical monocyte	S100A8	Surface marker (alarmin)	PDAC immunosuppression study;PMID: 23894703 (Basso et al.,2013).

Continued on next page

Cell Type	Target Symbol	Target Class	Evidence Level & Citation
Classical monocyte	SPI1	Transcription factor	PDAC scRNA-seq atlas;PMID: 31273297 (Peng et al.,2019).
Classical/Maintainer-like Malignant Cells	CEACAM6	GPI-anchored Protein	PMID: 35110444 (Baumgartner et al.,2022)
Classical/Maintainer-like Malignant Cells	MET	Receptor Tyrosine Kinase	PMID: 16533754 (Di Renzo et al.,2006)
Classical/progenitor malignant ductal	CLDN18	Tight Junction Protein (Surface)	Clinical/pathology + targeting rationale;PMID: 36051096
Classical/progenitor malignant ductal	GATA6	Transcription Factor	Mechanistic + clinical association;PMID: 27325420
Classical/progenitor malignant ductal	MSLN	GPI-anchored Surface Glycoprotein	Preclinical therapeutic targeting; PMID: 18281514
Classical/progenitor malignant ductal	MUC1	Transmembrane Mucin (Surface)	Phase I vaccine in pancreatic cancer;PMID: 15372205
Cycling malignant	AURKA	Ser/Thr Kinase	Preclinical PDAC study;PMID: 25632225
Cycling malignant	CDK1	Cyclin-Dependent Kinase	PDAC-focused synthesis/positioning;PMID: 34503199
Cycling malignant	MKI67	Proliferation Marker	Standard cycling-state marker; PDAC cycling annotation context;PMID: 35995947
Cycling malignant	PLK1	Ser/Thr Kinase	PDAC-focused preclinical/combination;PMID: 40188436
Ductal Epithelial Cells (Non-malignant)	CA2	Enzyme	PMID: 33649587 (Tosti et al., 2021)
Ductal_Normal	KRT19	Cytoskeletal Protein	PMID: 21323577 (Strobel et al., 2010)
EMT-high malignant	CDS2	Metabolic Enzyme	State-selective dependency mapping;PMID: 40848256
EMT-high malignant	TBK1	Ser/Thr Kinase	Mechanistic AXL→TBK1 axis; PMID: 30938713
Endocrine islet cells	CHGA	Secretory Granule Protein (Marker)	Endocrine markers in PDAC scRNA contexts;PMID: 385? (review)
Endocrine islet cells	GCG	Secreted Hormone (Marker)	Endocrine markers in PDAC scRNA contexts;PMID: 385? (review)

Continued on next page

Cell Type	Target Symbol	Target Class	Evidence Level & Citation
Endocrine islet cells	INS	Secreted Hormone (Marker)	Endocrine markers in PDAC scRNA contexts;PMID: 385? (review)
Endocrine islet cells	SST	Secreted Hormone (Marker)	Endocrine markers in PDAC scRNA contexts;PMID: 385? (review)
Endothelial Blood	DLL4	Ligand (Notch pathway)	Preclinical;Noguera-Troise et al., 2006 (PMID: 17189185)
Follicular B Cells	PAX5	Transcription Factor	PMID: 32341562 (Kandalaft et al.,2022)
γ δ T Cells ($\gamma\delta$ T Cells)	KLRB1	C-Type Lectin Receptor	PMID: 31562229 (Daley et al., 2019)
γ δ T Cells ($\gamma\delta$ T Cells)	TRGC1	T Cell Receptor (Constant Region)	PMID: 26808574 (Daley et al., 2016)
γ δ T Cells ($\gamma\delta$ T Cells)	TRGC2	T Cell Receptor (Constant Region)	PMID: 26808574 (Daley et al., 2016)
Hypoxia/glycolysis-high malignant	APEX1	DNA Repair/Redox Factor	APE1/Ref-1 inhibition decreases HIF1A-CA9 induction;PMID: 27535970
Hypoxia/glycolysis-high malignant	CA9	Enzyme (Cell-surface carbonic anhydrase)	Hypoxia axis dual-targeting in PDAC;PMID: 27535970
Hypoxia/glycolysis-high malignant	HIF1A	Transcription Factor	Targeted knockdown suppresses PDAC metastasis;PMID: 31100311
Hypoxia/glycolysis-high malignant	SLC2A1	Transporter (Surface)	Hypoxia/glycolysis hallmark in PDAC;HIF1A-linked program; PMID: 31100311
iCAF	CXCL12	Chemokine	PMID: 24277834 (Feig et al., 2013)
iCAF	IL1B	Cytokine	PMID: 36107485 (de Castro Silva et al.,2022)
iCAF	IL1R1	Receptor	PMID: 34518296 (Dosch et al., 2021)
iCAF	IL6	Cytokine	PMID: 34518296 (Dosch et al., 2021)
iCAF	IL6R	Cytokine Receptor	Preclinical;Lesina et al.,2011 (PMID: 21670415)
iCAF	JAK2	Kinase	Preclinical;Wormann et al.,2014 (PMID: 24917537)
iCAF	STAT3	Transcription factor	PMID: 36107485 (de Castro Silva et al.,2022)
IFN-stimulated TAM	TLR9	Pattern-recognition receptor	PDAC immunotherapy models; PMID: 34479922 (Carbone et al., 2021).
Inflammatory TAM	CEBPB	Transcription factor	PDAC scRNA-seq atlas;PMID: 31273297 (Peng et al.,2019).
Inflammatory TAM	CSF1R	Receptor tyrosine kinase	Preclinical PDAC models;PMID: 25082815 (Zhu et al.,2014).

Continued on next page

Cell Type	Target Symbol	Target Class	Evidence Level & Citation
Interferon-stimulated malignant	CD274	Immune Check-point Ligand (Surface)	ISG15 pathway regulates PD-L1 in PDAC;PMID: 31709456
Interferon-stimulated malignant	ISG15	Ubiquitin-like Modifier	Functional knockdown sensitizes to PD-1 therapy;PMID: 31709456
Interferon-stimulated malignant	JAK1	Kinase	Metastatic PDAC trial + IFN signaling targeting;PMID: 26351344
Interferon-stimulated malignant	JAK2	Kinase	Metastatic PDAC trial + IFN signaling targeting;PMID: 26351344
Interferon-stimulated malignant	STAT1	Transcription Factor	Prognostic/biology association in pancreatic cancer;PMID: 24658320
LAMP3+ migratory/activated DC	CCR7	Chemokine receptor	PDAC scRNA integration;PMID: 37633924 (Oh et al.,2023).
LAMP3+ migratory/activated DC	LAMP3	Surface/activation marker	PDAC scRNA integration;PMID: 37633924 (Oh et al.,2023).
Lymphatic EC	FLT4	Receptor tyrosine kinase	PMID: 16525637 (Schneider et al.,2006)
Lymphatic EC	LYVE1	Surface receptor	PMID: 27267807 (Gore et al., 2016)
Lymphatic EC	PDPN	Surface glycoprotein	PMID: 27267807 (Gore et al., 2016)
Lymphatic EC	VEGFC	Secreted ligand	PMID: 11291931 (Tang et al., 2001)
M-MDSC	ARG1	Enzyme (immunometabolic)	PDAC orthotopic model;PMID: 36727849 (Menjivar et al.,2023).
M-MDSC	IL4I1	Enzyme (Phenylalanine oxidase)	Emerging (Sadik et al.,2020; PMID: 32972994)[Pan-cancer incl. PDAC data]
M-MDSC	S100A8	Alarmin complex	PDAC immunosuppression study;PMID: 23894703 (Basso et al.,2013).
M-MDSC	S100A9	Alarmin complex	PDAC immunosuppression study;PMID: 23894703 (Basso et al.,2013).
Malignant_ Basal-like	KRT17	Cytoskeletal Protein	PMID: 21460848 (Collisson et al.,2011)
Malignant_ Basal-like	MYC	Transcription Factor	PMID: 29942284 (Brunton et al., 2018)
Malignant_ Basal-like	S100A2	Calcium-binding Protein	PMID: 26115632 (Moffitt et al., 2015)
Malignant_ Classical	KRAS	GTPase	PMID: 25482143 (Waters & Der, 2014)
Malignant_ Classical	MUC5AC	Mucin	PMID: 26115632 (Moffitt et al., 2015)

Continued on next page

Cell Type	Target Symbol	Target Class	Evidence Level & Citation
Malignant- Classical Mast Cell	TFF1	Trefoil Factor	PMID: 29915354 (Hayashi et al., 2018)
Mast cell	CPA3	Carboxypeptidase	Validated (Porcelli et al.,2019; PMID: 31305093)
Mast cell	KIT	Receptor tyrosine kinase	Human PDAC tissue study; PMID: 33669751 (Ammendola et al.,2021).
Mast cell	TPSAB1	Protease (marker/effector)	Human PDAC tissue study; PMID: 33669751 (Ammendola et al.,2021).
Mast Cell	TPSAB1	Serine Protease	Validated (Ma et al.,2019;PMID: 30944854)
Mast Cell	TPSB2	Serine Protease	Validated (Ma et al.,2019;PMID: 30944854)
Mesothelial-like stromal myCAF	WT1	Transcription factor	PMID: 35523176 (Huang et al., 2022)
myCAF	ACTA2	Cytoskeletal/ contractile	PMID: 28232471 (Öhlund et al., 2017)
myCAF	COL1A1	ECM structural	PMID: 28232471 (Öhlund et al., 2017)
myCAF	FAP	Surface enzyme	PMID: 24277834 (Feig et al., 2013)
myCAF	LOXL2	Enzyme (Oxidase)	Preclinical;Miller et al.,2015 (PMID: 26046445)
myCAF	LRRC15	Surface protein	PMID: 31699795 (Dominguez et al.,2020)
myCAF	PDGFRA	Receptor Tyrosine Kinase	Preclinical;Hwang et al.,2008 (PMID: 18424622)
myCAF	TGFB1	Cytokine	PMID: 31699795 (Dominguez et al.,2020)
Natural Killer (NK)Cells	KLRK1	Activating Receptor (C-Type Lectin)	PMID: 34670907 (Van Aude-naerde et al.,2021)
Natural Killer (NK)Cells	NCR1	Activating Receptor (IgSF)	PMID: 34670907 (Van Aude-naerde et al.,2021)
Neutrophil	BHLHE40	Transcription factor	PDAC scRNA-seq neutrophils; PMID: 35688610 (Wang et al., 2023).
Neutrophil	FCGR3B	Surface marker (Fc receptor)	PDAC scRNA-seq atlas;PMID: 31273297 (Peng et al.,2019).
NK	HLA-E	Checkpoint ligand	PMID: 36787695
NK	KLRC1	Immune checkpoint receptor	PMID: 36787695
NKT/MAIT-like	CD1D	Antigen-presenting molecule (axis)	PMID: 16772371

Continued on next page

Cell Type	Target Symbol	Target Class	Evidence Level & Citation
NKT/MAIT-like	IFNAR1	Cytokine receptor	PMID: 38980903
NKT/MAIT-like	TRAV10	Lineage marker (TCR gene)	PMID: 38980903
NKT/MAIT-like	TRAV11	Lineage marker (TCR gene)	PMID: 38980903
Non-classical monocyte	CX3CR1	Chemokine receptor	PDAC scRNA-seq atlas;PMID: 31273297 (Peng et al.,2019).
Non-classical monocyte	FCGR3A	Surface marker (Fc receptor)	PDAC scRNA-seq atlas;PMID: 31273297 (Peng et al.,2019).
Non-classical monocyte	NR4A1	Transcription factor	PDAC scRNA-seq atlas;PMID: 31273297 (Peng et al.,2019).
Normal ductal	ANXA4	Membrane-associated Protein (Marker)	PDAC spatial/single-cell ductal markers;PMID: 35995947
Normal ductal	CFTR	Ion Channel (Surface)	PDAC spatial/single-cell ductal markers;PMID: 35995947
Normal ductal	SLC4A4	Transporter (Surface)	scRNA + functional targeting in PDAC models;PMID: 36522548
Normal ductal	SOX9	Transcription Factor	Ductal gene required for acinar-derived lesions;PMID: 23201164
Partial EMT malignant	TGFBR1	Receptor Ser/Thr Kinase	Randomized clinical study with gemcitabine;PMID: 30318515
Partial EMT malignant	VIM	Intermediate Filament (Marker)	EMT marker usage in PDAC metastasis context;PMID: 28414315
Partial EMT malignant	ZEB1	Transcription Factor	Genetic depletion suppresses metastasis;PMID: 28414315
pDC	CLEC4C	C-type Lectin Receptor	Validated (Conrad et al.,2022; PMID: 35021034)
pDC	GZMB	Marker/effector gene	PDAC scRNA-seq atlas;PMID: 31273297 (Peng et al.,2019).
pDC	IRF7	Transcription factor	PDAC scRNA-seq atlas;PMID: 31273297 (Peng et al.,2019).
pDC	LILRA4	Immunoglobulin-like Receptor	Validated (Kenkel et al.,2017; PMID: 29087323)[Mouse PDAC, target validation]
pDC	TCF4	Transcription Factor	Validated (Kenkel et al.,2017; PMID: 29087323)
pDC	TLR7	Receptor	PMID: 29906447
Pericyte	CSPG4	Surface proteoglycan	Natarajan et al.,2022
Pericyte	PDGFRB	Receptor tyrosine kinase	PMID: 22403002 (Yuzawa et al., 2012)
Pericyte	RGS5	Signaling regulator	Natarajan et al.,2022
Plasma cell	MZB1	Secretory chaperone	PMID: 36099917

Continued on next page

Cell Type	Target Symbol	Target Class	Evidence Level & Citation
Plasma cell	PRDM1	Transcription factor	PMID: 36099917
Plasma cell	SDC1	Surface marker	PMID: 36099917
Plasma cell	XBP1	Transcription factor	PMID: 36099917
Plasma_Cell	CD38	Ectoenzyme/ Receptor	PMID: 33144837
PMN-MDSC	OLR1	Scavenger Receptor	Validated (Condamine et al., 2015;PMID: 26479788)[Extended to PDAC models]
PMN-MDSC	PROK2	Chemokine-like Protein	Validated (Miao et al.,2020; PMID: 32612001)
PMN-MDSC/ neutrophil-like suppressor	CXCR2	Chemokine receptor	PDAC metastasis/immunity study;PMID: 27265504 (Steele et al.,2016).
Quiescent stellate/resident fibroblast	CYGB	Redox/secreted protein	PMID: 28540429 (Nielsen et al., 2017);Hoang et al.,2022
Quiescent stellate/resident fibroblast	HYAL1	ECM pathway	PMID: 41681979 (Chiorean et al.,2026)
Quiescent stellate/resident fibroblast	HYAL2	ECM pathway	PMID: 41681979 (Chiorean et al.,2026)
Quiescent stellate/resident fibroblast	PLIN2	Lipid droplet protein	PMID: 28540429 (Nielsen et al., 2017)
Quiescent stellate/resident fibroblast	RARA	Nuclear receptor	PMID: 32973176 (Kocher et al., 2020)
Quiescent stellate/resident fibroblast	VDR	Nuclear receptor	PMID: 25259922 (Sherman et al.,2014)
retroperitoneal/ Stellate-like Fibroblasts	LRAT	Enzyme (Retinol Metabolism)	Validated (Helms et al.,2020; PMID: 32939096)
Schwann/ peripheral glia	GFAP	Intermediate filament	PMID: 37524695 (Xue et al., 2023)
Schwann/ peripheral glia	IL1A	Cytokine	PMID: 37524695 (Xue et al., 2023)
Schwann/ peripheral glia	MDK	Secreted growth factor	PMID: 37524695 (Xue et al., 2023)
Schwann/ peripheral glia	NGFR	Surface receptor	PMID: 37524695 (Xue et al., 2023)
Schwann/ peripheral glia	S100B	Calcium-binding protein	PMID: 37524695 (Xue et al., 2023)
Schwann_Cell	NTRK1	Receptor Tyrosine Kinase	Preclinical;Cavel et al.,2012 (PMID: 22116068)
Schwann_Cell	RET	Receptor Tyrosine Kinase	Preclinical;Gil et al.,2010 (PMID: 20566723)

Continued on next page

Cell Type	Target Symbol	Target Class	Evidence Level & Citation
SPP1-high TAM	CTHRC1	TME interaction node	PDAC TME study;PMID: 39075108 (Li et al.,2024).
SPP1-high TAM	MIF	Cytokine	PDAC scRNA integration;PMID: 37633924 (Oh et al.,2023).
SPP1-high TAM	SPP1	Secreted cytokine-like ECM protein	PDAC scRNA integration;PMID: 37633924 (Oh et al.,2023).
TAM - Immuno-suppressive	CD163	Scavenger Receptor	Validated (Puig-Kröger et al., 2020;PMID: 32376654)
TAM - Immuno-suppressive	IL10	Cytokine	Validated (Ruffell et al.,2014; PMID: 24703780)[Context: PDAC models]
TAM - Immuno-suppressive	MRC1	Scavenger Receptor	Validated (Kemp et al.,2021; PMID: 34461618)
TAM - Inflammatory	CXCL8	Chemokine	Validated (Chen et al.,2021; PMID: 34755856)
TAM - Lipid-Associated	LPL	Lipase	Emerging (Lobba et al.,2024; PMID: 38286473)
Treg	CCR8	Chemokine Receptor	PMID: 36796338
Treg	CTLA4	Immune checkpoint receptor	PMID: 38852289
Treg	FOXP3	Transcription factor	PMID: 36968070
Treg	ITGAV	Surface marker /therapeutic target	PMID: 37292693
Treg	ITGB5	Surface marker /therapeutic target	PMID: 37292693
TREM2-APOE-like TAM	APOE	Secreted/ lipid-transport protein	Human PDAC macrophage profiling;PMID: 33782087 (Kemp et al.,2021).
TREM2-APOE-like TAM	TREM2	Immunoreceptor	Human PDAC macrophage profiling;PMID: 33782087 (Kemp et al.,2021).
Venous EC	ACKR1	Atypical chemokine receptor	Shiau et al.,2022
Venous EC	ANGPT2	Secreted ligand	PMID: 37907568 (Roos-Mattila et al.,2023)
Venous EC	SELE	Surface adhesion	Craven et al.,2015
Venous EC	VCAM1	Surface adhesion	Craven et al.,2015
vSMC	MYH11	Contractile motor	Shiau et al.,2022
vSMC	TAGLN	Actin-binding	Shiau et al.,2022
$\gamma\delta$ T	IL17A	Cytokine	PMID: 27569912
$\gamma\delta$ T	RORC	Transcription factor	PMID: 27569912

Continued on next page

Cell Type	Target Symbol	Target Class	Evidence Level & Citation
$\gamma\delta$ T	TRDC	Surface/TCR marker	PMID: 27569912

Table S2: PDAC cell-specific pathways

Pathway Name	Source	Associated Cell Type(s)	Key Targets Involved	Citation
Acinar Cell Differentiation	Developmental Biology	Acinar_Cell	PTF1A; BHLHA15	PMID: 23610129
Antigen processing and presentation	KEGG; KEGG (hsa04612)	Antigen-presentation-competent malignant; apCAF;NKT/MAIT-like;antigen-presenting CAFs (apCAFs);cDC2	HLA-DRA; CD74;B2M; CD1D;CD1C	PMID: 41278823;Huang 2022 (PMID: 35523176); PMID: 16772371;Elyada et al.,2019 (PMID: 31197017);Hey et al.,2023
Arginine and proline metabolism	KEGG; KEGG (hsa00330)	M-MDSC	ARG1	Menjivar et al.,2023 (PMID: 36727849);PMID: 21466986;Gabitass et al., 2011
B cell receptor signaling pathway	KEGG; KEGG (hsa04662);Reactome (R-HSA-983695)	B cell (naïve/memory);B_Cell;Follicular B Cells	CD19;MS4A1; PAX5;BCL6	PMID: 36099917;PMID: 34101736;PMID: 27455446 (Castino et al.,2016)
C-type lectin receptor signaling	KEGG (hsa04625)	cDC1;cDC2;pDC	CLEC9A; CLEC10A; CLEC4C	Elyada et al.,2019
Calcineurin-NFAT signaling	Reactome	ADM-like/acinar-like metaplastic	NFATC4	PMID: 26697077
Calcium signaling pathway	KEGG (hsa04020)	Basal-like/Squamous Malignant Cells	S100A2	PMID 17145868 (Ohuchida et al.,2007)
Carbonic anhydrase-mediated bicarbonate transport	Reactome (R-HSA-1475029)	Ductal Epithelial Cells (Non-malignant)	CA2	PMID 33649587 (Tosti et al.,2021)
Catecholamine biosynthesis	Reactome (R-HSA-209931)	Adrenergic/Neural-like Progenitor Cells	TH	PMID 34808793 (Li et al., 2021)
CD40 signaling	Reactome	cDC1;cDC2;LAMP3+ migratory/activated DC;NKT/MAIT-like	CD40	Beatty et al.,2011 (PMID: 21436454);Beatty et al., 2013 (PMID: 23983255); PMID: 38980903;PMID: 21336340
CDS2-dependent phospholipid metabolism supporting mesenchymal state	Reactome	EMT-high malignant	CDS2	PMID: 40848256
Cell adhesion molecules	KEGG; KEGG (hsa04514)	CD8 Trm-like; Classical/Maintainer-like Malignant Cells	ITGAE; CEACAM6	PMID: 36689623;PMID 35110444 (Baumgartner et al.,2022)
Cell cycle (G2/M checkpoint control)	KEGG	Cycling malignant	PLK1;AURKA; CDK1;MKI67	PMID: 40188436;PMID: 25632225;PMID: 34503199
Chemokine receptors bind chemokines	Reactome	iCAF	CXCL12	Feig 2013 (PMID: 24277834)
Chemokine signaling pathway	KEGG; KEGG (hsa04062)	iCAF;Classical monocyte;M-MDSC; Non-classical monocyte;PMN-MDSC/ neutrophil-like suppressor;Neutrophil; LAMP3+ migratory/activated DC; CD8 Trm-like; γδ T;LAMP3+ DC;Monocyte; PMN-MDSC;Treg; inflammatory CAFs (iCAFs);TAM - Inflammatory	CXCL12;CCR2; CX3CR1; CXCR2;CCR7; CXCR6;CCR8; CXCL8;PROK2	Feig 2013 (PMID: 24277834);Nywening et al., 2016 (PMID: 27055731); Peng et al.,2019 (PMID: 31273297);Steele et al., 2016 (PMID: 27265504); Ijichi et al.,2011 (PMID: 21926469);Oh et al.,2023 (PMID: 37633924);PMID: 27569912;Steele et al.,2016 (PMID: 26900958);PMID: 23644553;Elyada et al., 2019 (PMID: 31197017); Chen et al.,2021;Miao et al.,2020
CSF1/CSF1R signaling	Reactome; Custom/ Literature	Inflammatory TAM; TAM_C1QC+	CSF1R	Zhu et al.,2014 (PMID: 25082815);PMID: 25219499
CTLA4 inhibitory signaling	Reactome; Custom/ Literature	Treg	CTLA4	PMID: 38852289;PMID: 34789575

Continued on next page

Pathway Name	Source	Associated Cell Type(s)	Key Targets Involved	Citation
Cytokine Signaling (IFN- γ)	Reactome (R-HSA-877300); Hallmark Interferon γ Response	CD8+ Cytotoxic/ Effector; CD4+ Th1	IFNG; TNF; TBX21	PMID: 26200351 (Stromnes et al.,2015)
Cytokine Signaling (IL-7)	Reactome (R-HSA-449147); KEGG (hsa04060)	CD8+ Naive/Memory-like; CD8+ Cytotoxic/ Effector	IL7R; IFNG	PMID: 31331973 (Steele et al.,2020)
Cytokine signaling in immune system	Reactome	SPP1-high TAM	MIF	Oh et al.,2023 (PMID: 37633924)
Dendritic cell differentiation (BATF3-dependent)	Reactome	cDC1	BATF3	Hogg et al.,2025 (PMID: 40815670)
Dendritic cell differentiation (IRF4-dependent)	Reactome	cDC2	IRF4	Hogg et al.,2025 (PMID: 40815670)
ECM-receptor interaction	KEGG; KEGG (hsa04512)	myCAF; SPP1-high TAM; TAM - Lipid-Associated	COL1A1; SPP1	Öhlund 2017 (PMID: 28232471); Oh et al.,2023 (PMID: 37633924); Daliani et al.,2022
Ephrin signaling	Reactome	Arterial EC	EFNB2	Shiau 2022
ErbB signaling pathway	KEGG; KEGG (hsa04012)	Acinar; ADM-like/ acinar-like meta- plastic; Basal-like/ Squamous Malignant Cells	EGFR	PMID: 22975374; PMID 22387390 (Navas et al., 2012)
Extracellular matrix organization	Reactome	myCAF	COL1A1; FAP	Öhlund 2017 (PMID: 28232471); Feig 2013 (PMID: 24277834)
Extracellular Matrix Remodeling	Gene Ontology	myCAF	FAP; LOXL2	Lo et al.,2017 (PMID: 28622507)
Fc ϵ RI signaling pathway	KEGG (hsa04664)	Mast Cell	TPSAB1; TPSB2; CPA3	Ma et al.,2019
FLT3 Signaling	Reactome	cDC1	FLT3	PMID: 32243813
Focal adhesion	KEGG (hsa04510)	myofibroblastic CAFs (myCAFs)	FAP; ACTA2	Öhlund et al.,2017 (PMID: 28930670)
$\gamma\delta$ T Cell TCR Signaling	KEGG (hsa04660); Reactome (R-HSA-198933)	$\gamma\delta$ T Cells ($\gamma\delta$ T Cells)	TRGC1; TRGC2; LAG3; KLRB1; CD161	PMID: 26808574 (Daley et al.,2016); PMID: 31562229 (Daley et al.,2019)
Glucagon signaling /endocrine pancreas markers	Reactome	Endocrine islet cells	GCG; SST; CHGA	PMID: 35995947
Granzyme-Mediated Apoptosis	Reactome (R-HSA-75158)	CD8+ Cytotoxic/ Effector	GZMB; PRF1	PMID: 32472114 (Sade-Feldman et al.,2019)
Hallmark Epithelial Mesenchymal Transition	MSigDB Hallmark; MSigDB Hallmark (HALLMARK_EMT); HALLMARK_Malignant_Basal_Epithelial-like; Stress-Induced/ MESENCHYMAL-like Malignant Cells	Partial EMT malignant; EMT-high malignant; Basal-like/ squamous malignant;	ZEB1; AXL; VIM; TP63; EGFR	PMID: 28414315; PMID: 34235075; PMID: 29502946; PMID 26322394 (Zheng et al.,2015)
Hallmark G2M Checkpoint	MSigDB Hallmark	Cycling malignant	PLK1; AURKA; CDK1	PMID: 40188436; PMID: 25632225
Hallmark Glycolysis	MSigDB Hallmark; MSigDB Hallmark (HALLMARK_GLYCOLYSIS)	Hypoxia/ glycolysis-high malignant; Stress-Induced/ EMT-like Malignant Cells	SLC2A1; HIF1A	PMID: 31100311; PMID 32795408 (Chand et al., 2020)
Hallmark Hypoxia	MSigDB Hallmark	Hypoxia/ glycolysis-high malignant	HIF1A; CA9	PMID: 31100311
Hallmark MYC Targets V1	MSigDB Hallmark	Malignant_Basal-like	MYC	PMID: 29942284
Hallmark: Unfolded Protein Response	MSigDB Hallmark	Plasma cell	XPB1; MZB1	PMID: 36099917
Hematopoietic cell lineage	KEGG (hsa04640)	pDC	TCF4	Kenkel et al.,2017

Continued on next page

Pathway Name	Source	Associated Cell Type(s)	Key Targets Involved	Citation
HGF/MET signaling pathway	Reactome (R-HSA-8875878)	Classical/Maintainer-like Malignant Cells	MET	PMID 16533754 (Di Renzo et al.,2006)
HIF-1 signaling pathway	KEGG	Hypoxia/glycolysis-high malignant	HIF1A;CA9; SLC2A1	PMID: 31100311;PMID: 27535970
Hyaluronan metabolic process	Reactome	Quiescent stellate/resident fibroblast	HYAL1;HYAL2	Chiorean 2026 (PMID: 41681979)
IL-1 signaling pathway	KEGG (hsa04061)	TAM - Inflammatory	IL1B	Biffi et al.,2019
IL-10 signaling	Reactome (R-HSA-6783783)	TAM - Immunosuppressive	IL10	Ruffell et al.,2014
IL-17 signaling pathway	KEGG	$\gamma\delta$ T	IL17A;RORC	PMID: 27569912
IL-6/JAK/STAT3 Signaling	Hallmark	iCAF	IL6R;JAK2	Erez et al.,2010 (PMID: 20592298)
IL6/JAK/STAT3 signaling	MSigDB Hallmark	iCAF	IL6;STAT3	Dosch 2021 (PMID: 34518296);de Castro Silva 2022 (PMID: 36107485)
Immune checkpoint signaling (LAG-3 axis)	MSigDB (C7 immunologic signatures)	CD8 exhausted T	LAG3	PMID: 33803936
Immune checkpoint signaling (TIGIT axis)	Reactome	CD8 exhausted T; CD8 Trm-like;Treg; NK	TIGIT;PVR	PMID: 34358448
Immune checkpoint signaling (TIM-3 axis)	MSigDB (C7 immunologic signatures)	CD8 exhausted T	HAVCR2	PMID: 34072549
Immunoregulatory signaling	Reactome (R-HSA-198933)	pDC	LILRA4	Kenkel et al.,2017
Inhibitory signaling by PD-1	Reactome; KEGG (hsa04514);Reactome (R-HSA-389357)	CD8 cytotoxic effector T;CD8 exhausted T; CD8 Trm-like;CD4 helper T;CD8+ T Cells - Exhausted (Tex)	PDCD1; HAVCR2	PMID: 36689623;PMID: 33649582 (Sivakumar et al.,2023);PMID: 32472114 (Sade-Feldman et al.,2019)
Insulin secretion	KEGG	Endocrine islet cells	INS;CHGA	PMID: 35995947
Integrin cell surface interactions	Reactome	Treg	ITGAV;ITGB5	PMID: 37292693
Integrin Signaling Pathway	KEGG	TAM_SPP1+	SPP1	PMID: 34114569
Interferon α response	MSigDB Hallmark (HALLMARK_INTERFERON_ α -RESPONSE)	IFN-stimulated TAM; pDC	STAT1;ISG15; IRF7	Peng et al.,2019 (PMID: 31273297)
Interferon α/β signaling	Reactome	Interferon-stimulated malignant;NKT/MAIT-like	JAK1;JAK2; STAT1;ISG15; IFNAR1	PMID: 26351344;PMID: 24658320;PMID: 31709456; PMID: 38980903
Interferon signaling	KEGG (hsa04630)	cDC1	IRF8	Meyer et al.,2018
Interleukin-1 signaling	Reactome	iCAF;apCAF; Mesothelial-like stromal;Schwann/peripheral glia; Inflammatory TAM; TREM2-APOE-like TAM	IL1R1;IL1B; IL1A	Dosch 2021 (PMID: 34518296);Huang 2022 (PMID: 35523176);Xue 2023 (PMID: 37524695); Yang et al.,2025 (PMID: 39956331)
JAK-STAT signaling pathway	KEGG; KEGG (hsa04630)	Interferon-stimulated malignant;iCAF;B cell (naïve/memory);CD4- T.conv;inflammatory CAFs (iCAFs)	JAK1;JAK2; STAT1;IL6; STAT3;IL12A; EBI3	PMID: 26351344;PMID: 24658320;Dosch 2021 (PMID: 34518296); de Castro Silva 2022 (PMID: 36107485);PMID: 36099917;PMID: 23293072; Elyada et al.,2019 (PMID: 31197017)
Keratinization	Reactome (R-HSA-6805567)	Basal-like/Squamous Malignant Cells; Ductal Epithelial Cells (Non-malignant)	KRT5;KRT19	PMID 31748746 (Maurer et al.,2019);PMID 35683463 (Buscail et al., 2022)

Continued on next page

Pathway Name	Source	Associated Cell Type(s)	Key Targets Involved	Citation
LAG3 Immune Checkpoint Pathway	Custom/ Literature	CD8_T_ex_terminal	LAG3	PMID: 34509489
Leukocyte transendothelial migration	KEGG	Venous EC	VCAM1;SELE; ACKR1	Craven 2015;Shiau 2022
Lipid metabolism	MSigDB Hallmark	PMN-MDSC	OLR1	Condamine et al.,2015
MAPK signaling pathway	KEGG	Antigen-presentation-competent malignant; ADM-like/acinar-like metaplastic;Acinar; Schwann/peripheral glia;Malignant_Classical;Malignant_Basal-like	MAP2K1; MAP2K2; EGFR;REG3A; REG3B; EXTL3;MDK; KRAS	PMID: 41278823; PMID: 22975374;PMID: 34006485;Xue 2023 (PMID: 37524695);PMID: 25482143
Mesothelin-mediated tumor antigen programs	Reactome	Classical/progenitor malignant ductal	MSLN	PMID: 18281514
MHC class II antigen presentation	Reactome	Antigen-presentation-competent malignant; apCAF	HLA-DRA; CD74	PMID: 41278823;Huang 2022 (PMID: 35523176); Elyada et al.,2019 (PMID: 31123163)
Mucin-type O-glycan biosynthesis /mucin biology	KEGG	Classical/progenitor malignant ductal	MUC1	PMID: 15372205
NAD+ Metabolic Pathway	KEGG	Plasma_Cell	CD38	PMID: 33144837
Natural killer cell mediated cytotoxicity	KEGG	CD8 cytotoxic effector T;NK	GZMB; FCGR3A	PMID: 34296197
Natural killer cell mediated cytotoxicity (inhibitory receptor module)	KEGG	NK	KLRC1;HLAE	PMID: 36787695
Neuroactive ligand-receptor interaction	KEGG (hsa04080)	Adrenergic/Neural-like Progenitor Cells	ADRB2	PMID 27624061 (Partecke et al.,2016)
Neurotrophin Signaling	KEGG; KEGG (hsa04722)	Schwann_Cell; Adrenergic/Neural-like Progenitor Cells	NTRK1;RET; NTRK2	Cavel et al.,2012 (PMID: 22116068);PMID 31748746 (Maurer et al.,2019)
NF- κ B signaling pathway	KEGG; MSigDB Hallmark (HALLMARK_TNFA_SIGNALING_VIA_NFKB)	iCAF;Inflammatory TAM	IL1R1;IL1B	de Castro Silva 2022 (PMID: 36107485);Yang et al.,2025 (PMID: 39956331)
NK Cell Cytotoxicity	KEGG (hsa04650);Reactome (R-HSA-202411)	Natural Killer (NK) Cells	NCR1;KLRK1; TIGIT	PMID: 34670907 (Van Audenaerde et al.,2021)
Notch Signaling	KEGG	Endothelial.Blood	DLL4	Noguera-Troise et al.,2006 (PMID: 17189185)
Nuclear receptor transcription pathway	Reactome	Acinar	NR5A2	PMID: 23645620
p53 signaling pathway	KEGG (hsa04115)	Basal-like/Squamous Malignant Cells	TP63	PMID 30100244 (Somerville et al.,2018)
Pancreatic Ductal Development	Developmental Biology	Ductal_Normal	SOX9	PMID: 21852445
Pancreatic Endoderm Lineage Specification	Developmental Biology	Malignant_Classical	GATA6	PMID: 28246101
Pancreatic secretion	KEGG	Normal ductal;Acinar	CFTR;SLC4A4; PRSS1	PMID: 36522548;PMID: 35995947
PD-1 Checkpoint Pathway in Cancer	KEGG	CD8_T_ex.progenitor	PDCD1	PMID: 20453085
PD-L1 expression and PD-1 checkpoint pathway in cancer	KEGG	Interferon-stimulated malignant;CD8 cytotoxic effector T;CD8 exhausted T;CD8 Trm-like;CD4 helper T	CD274;ISG15; PDCD1	PMID: 31709456;PMID: 35480107

Continued on next page

Pathway Name	Source	Associated Cell Type(s)	Key Targets Involved	Citation
PDGF signaling	Reactome; Reactome (R-HSA-186763)	Pericyte;vSMC; myCAF;inflammatory CAFs (iCAF)s& Vascular Leiomyocytes (Pericytes)	PDGFRB; PDGFRA	Yuzawa 2012 (PMID: 22403002);Natarajan 2022; Hwang et al.,2008 (PMID: 18424622);Dominguez et al.,2020 (PMID: 33577702) ;Hosein et al.,2019 (PMID: 31291569)
Phagosome	KEGG (hsa04145)	TAM - Immunosup- pressive	MRC1;CD163	Kemp et al.,2021
PI3K-Akt signaling pathway	KEGG (hsa04151)	Mast Cell	KIT	Wang et al.,2021
Plasma Cell Differentiation	Reactome (R-HSA-983697)	Plasma Cells	SDC1;MZB1; PRDM1	PMID: 32341562 (Kandalaft et al.,2022)
Plasma cell differ- entiation program (BLIMP1 axis)	MSigDB (C3 TF targets)	Plasma cell	PRDM1	PMID: 36099917
PPAR signaling pathway	KEGG (hsa03320)	TAM - Lipid- Associated	LPL	Lobba et al.,2024
Regulation of CFTR activity	Reactome (R-HSA-5619125)	Ductal Epithelial Cells (Non-malignant)	CFTR	PMID 37931504 (Liu et al.,2023)
Regulatory T Cell Maintenance	Reactome (R-HSA-8877330) ;KEGG (hsa04659)	CD4+ T Cells - Regu- latory T Cells (Tregs)	FOXP3;IL2RA; CTLA4; TNFRSF18	PMID: 28783719 (Jang et al.,2017);PMID: 29769723 (Wing et al.,2019)
Retinoic acid signaling	Reactome	Quiescent stellate/ resident fibroblast	RARA	Kocher 2020 (PMID: 32973176)
Retinol metabolism	KEGG (hsa00830)	retroperitoneal/ Stellate-like Fibroblasts	LRAT	Helms et al.,2020 (PMID: 32939096)
SCF-KIT signaling	Reactome	Mast cell	KIT	Ammendola et al.,2021 (PMID: 33669751); Deplanque et al.2015 (PMID: 25858497)
Signaling by nuclear receptors	Reactome	Quiescent stellate/ resident fibroblast	VDR;RARA	Sherman 2014 (PMID: 25259922);Kocher 2020 (PMID: 32973176)
Signaling by TAM receptors	Reactome	Basal-like/squamous malignant;Partial EMT malignant; EMT-high malignant	AXL	PMID: 34235075;PMID: 30938713
Solute carrier- mediated anion transport (bicarbonate transport)	Reactome	Normal ductal	SLC4A4	PMID: 36522548
SOX9 transcrip- tional regulation (ductal lineage program)	Reactome	Normal ductal;Acinar	SOX9	PMID: 23201164
Squamous Dif- ferentiation Program	DevelopmentalMalignant_Biology	Malignant_Basal-like	KRT17;TP63	PMID: 21460848
Syndecan- 1 mediated interactions	Reactome	Plasma cell	SDC1	PMID: 36099917
T cell receptor signaling pathway	KEGG; KEGG (hsa04660) ;Reactome (R-HSA-202403)	CD8 cytotoxic effector T;CD8 Trm-like; CD4 helper T;NKT/ MAIT-like;γδ T; CD8+ Naive/Memory- like;CD8+ Cytotoxic/ Effector;CD4+ Naive/ Central Memory; CD4+ Th1;γδ T Cells	CD8A;CD4; TRAV10; TRAV11; TRDC;TCF7; LEF1;TBX21; IFNG;TNF; TRGC1; TRGC2	PMID: 33803936;PMID: 31331973 (Steele et al., 2020);PMID: 26808574 (Daley et al.,2016)
TBK1/IKKε signal- ing (innate kinase axis)	Reactome	Partial EMT malig- nant;EMT-high malignant	TBK1;AXL	PMID: 30938713
TGF-β signaling pathway	KEGG; Reactome; Hallmark; KEGG (hsa04350)	Partial EMT malig- nant;myCAF;apCAF; Mesothelial-like stromal;NK_Cell; Stress-Induced/ EMT-like Malignant Cells	TGFBR1; TGFB1; TGFBR2; LRRRC15	PMID: 30318515; Dominguez 2020 (PMID: 31699795);Huang 2022 (PMID: 35523176); Öhlund et al.,2017 (PMID: 28249877);PMID: 21822295;PMID 27466318 (David et al.,2016)

Continued on next page

Pathway Name	Source	Associated Cell Type(s)	Key Targets Involved	Citation
Th1 Differentiation	KEGG (hsa04658) ;Reactome (R-HSA-8877330)	CD4+ T Cells - T helper 1 (Th1)	TBX21;IFNG; TNF	PMID: 26200351 (Stromnes et al.,2015)
Tie2 signaling	Reactome	Venous EC;Activated/ angiogenic EC	TEK;ANGPT2	Roos-Mattila 2023 (PMID: 37907568)
Tight Junction Assembly	Reactome	Malignant.Classical		PMID: 32320671
Tight junction interactions / cell junction organization	Reactome	Classical/progenitor malignant ductal	CLDN18	PMID: 36051096
TNF signaling pathway	KEGG	CD8 cytotoxic effector T	TNFRSF9	PMID: 35480107
TNFR Superfamily Co-stimulation	Custom/ Literature	CD8.T_effector	TNFRSF9	PMID: 29990506
Toll-like receptor signaling pathway	KEGG; KEGG (hsa04620)	IFN-stimulated TAM; pDC;M-MDSC	TLR9;S100A8; S100A9	Carbone et al.,2021 (PMID: 34479922); Porembka et al.,2012
TP63 regulates transcription of cell cycle and squamous differentiation genes	Reactome	Basal-like/squamous malignant	TP63;KRT5	PMID: 30428345
Transcriptional regulation by FOXP3	MSigDB (C3 TF targets)	Treg	FOXP3	PMID: 36968070
Transcriptional Regulation of Exhaustion	MSigDB Hall-mark (not directly) ;Custom (TOX network)	CD8+ T Cells - Exhausted (Tex)	TOX	PMID: 33649582 (Sivakumar et al.,2023)
Transcriptional regulation of pancreas development	Reactome (R-HSA-1266738)	Classical/Maintainer-like Malignant Cells	GATA6	PMID 30442762 (Martinelli et al.,2017)
Transcriptional repression by BCL6 (GC fate control)	Reactome	B cell (naïve/memory)	BCL6	PMID: 36099917
TREM2 signaling / Immunoregulatory macrophage program	Reactome; Custom/ Literature; Reactome (R-HSA-6798695)	TREM2-APOE-like TAM;TAM- TREM2+;TAM - Lipid-Associated	TREM2;APOE	Kemp et al.,2021 (PMID: 33782087);PMID: 36720108;Katz et al.,2023
TROP2 (TACSTD2)-linked epithelial tumor antigen axis	Reactome	Basal-like/squamous malignant	TACSTD2	PMID: 36878461
Tryptophan Catabolism	KEGG	apCAF	IDO1	Li et al.,2021 (PMID: 34127627)
Tryptophan metabolism	KEGG (hsa00380)	M-MDSC	IL4I1	Sadik et al.,2020
Vascular smooth muscle contraction	KEGG (hsa04270)	Vascular Leiomyocytes (Pericytes)	RGS5	Dominguez et al.,2020 (PMID: 33577702)
VEGF signaling pathway	KEGG; Reactome	Capillary EC;Arterial EC;Activated/ angiogenic EC; Lymphatic EC; Endothelial. Blood;Endothelial. Lymphatic	KDR;FLT1; VEGFA;NRP1; FLT4;VEGFC	Doi 2012 (PMID: 22207048);Zhou 2016 (PMID: 26804176); Schneider 2006 (PMID: 16525637);Tang 2001 (PMID: 11291931);Clark et al.,2008 (PMID: 18838551) ;Kurahara et al.,2011 (PMID: 21118944)
VISTA-mediated immune suppression (VSIR axis)	MSigDB (C7 immunologic signatures)	CD8 exhausted T	VSIR	PMID: 34072549

Table S3: PDAC LLM-driven mechanism hypotheses

Sequence	Upstream	Core Nodes	Downstream	Biological Rationale
CFTR → EGFR → EGR1 → MAP2K1 → MYC	CFTR, EGFR	EGR1, MAP2K1	MYC	Immediate-early transcriptional relay links ductal receptor input to proliferative amplification.
EGFR → ELK1 → VIM	EGFR	ELK1	VIM	MAPK-phosphorylated ELK1 converts receptor signaling into EMT-associated cytoskeletal remodeling.
EGFR → MAZ → VIM	EGFR	MAZ	VIM	Indirect transcriptional bridging channels EGFR activity into mesenchymal invasive state.
AXL → ERBB2 → EGFR → EGR1 → MYC	AXL, ERBB2	EGFR, EGR1	MYC	RTK heterodimer crosstalk and immediate-early transcription expand oncogenic growth signaling.
EGFR → HSP90AA1 → APEX1 → MYC	EGFR	HSP90AA1, MYC, APEX1		Chaperone-assisted proteostasis relays receptor stress signaling into redox-adaptive proliferation.
IL6 → STAT3 → MAX → VIM	IL6, STAT3	MAX	VIM	STAT3-to-MAX transcriptional coupling propagates inflammatory signaling into EMT output.
KIT → MAX → STAT3 → IL6	KIT	MAX, STAT3	IL6	Indirect nuclear coupling links mast-cell receptor input to fibroinflammatory cytokine reinforcement.
MYC → MAX → RARA → ESR2 → STAT3	MYC	MAX, RARA, ESR2	STAT3	MYC/MAX transcriptional output couples malignant growth to fibroblast nuclear receptor reprogramming.
PDGFRB → STAT5B → IFNG	PDGFRB	STAT5B	IFNG	Near-direct JAK/STAT bridging connects vascular stromal signaling to effector cytokine output.
PDGFRB → STAT5B → STAT3 → IL6	PDGFRB	STAT5B, STAT3	IL6	Perivascular signaling may route through STAT modules to sustain inflammatory stroma.

Continued on next page

Sequence	Upstream	Core Nodes	Downstream	Biological Rationale
FLT3 → TCF7 → CXCR6 → EGR1 → IFNG	FLT3, TCF7, CXCR6	EGR1	IFNG	Resident-memory differentiation gains an immediate-early relay before terminal effector cytokine production.
CD4 → PPARG → MAP2K1 → MYC	CD4	PPARG, MAP2K1	MYC	Helper-T transcriptional cues may indirectly reinforce malignant mitogenic transcription.
AXL → NR3C1 → STAT3 → MAZ → VIM	AXL	NR3C1, STAT3, MAZ	VIM	Stress-response and inflammatory transcription cooperate to intensify mesenchymal plasticity.
NGFR → KLF11 → MAP2K1 → MYC	NGFR	KLF11, MAP2K1	MYC	Neural-stromal input engages transcriptional buffering before re-entering proliferative MAPK circuitry.
EGFR → CTCF → STAT3 → IL6	EGFR	CTCF, STAT3	IL6	Chromatin-organization relay links epithelial RTK activity to inflammatory transcriptional programs.
EGFR → NFIL3 → PLK1 → AURKA → CDK1 → MKI67	EGFR	NFIL3, PLK1, AURKA, CDK1	MKI67	Transcriptional rewiring bridges receptor signaling to mitotic progression and cycling persistence.
NCR1 → MAX → STAT3 → IL6	NCR1	MAX, STAT3	IL6	Immune-contact signals may be transcriptionally assimilated into inflammatory tumor-promoting stroma.
EGFR →... → EGR1 → MYC	EGFR	EGR1	MYC	EGFR signaling induces early response TF EGR1, amplifying MYC-driven proliferation.
MYC ↔ MAX... RARA	MYC	MAX	RARA-pathway genes	The oncogenic MYC-MAX complex transcriptionally cross-regulates the RARA differentiation pathway.
IL6 → STAT3 →... → MAX →... → VIM	IL6, STAT3	MAX	VIM	Inflammatory STAT3 signaling co-opts MYC/MAX transcriptional machinery to promote epithelial-mesenchymal transition.
EGFR/STAT3... (Chromatin)↔ CTCF	EGFR, STAT3 signaling	CTCF	Co-regulated target genes	CTCF-mediated chromatin architecture integrates signals from disparate oncogenic pathways like EGFR and STAT3.

Continued on next page

Sequence	Upstream	Core Nodes	Downstream	Biological Rationale
AXL $\leftrightarrow$ ERBB2 $\rightarrow$ (Signaling Cascade) $\rightarrow$ MYC	AXL, ERBB2	AXL-ERBB2 complex	MYC	AXL and ERBB2 receptor crosstalk amplifies downstream signaling to drive MYC expression.
AXL $\rightarrow$... $\leftrightarrow$ NR3C1 $\leftrightarrow$... $\rightarrow$ MYC	AXL	NR3C1	MYC	Glucocorticoid receptor (NR3C1) signaling intersects with AXL pathway to modulate MYC expression.
(Hypoxia) $\rightarrow$ APEX1 $\rightarrow$ (Redox signals) $\leftrightarrow$ HSP90AA1	Hypoxia, APEX1	HSP90AA1	Stabilized client proteins	Hypoxic redox stress increases demand for HSP90 chaperone activity to maintain proteostasis.
(PDGF) $\rightarrow$ PDGFRB $\rightarrow$ STAT5B $\nrightarrow$ IFNG	PDGFRB (on CAF/Pericyte)	STAT5B (in T-cell)	IFNG (in T-cell)	Stromal PDGFRB signaling activates STAT5B to suppress T-cell anti-tumor IFN- γ production.
KIT (Mast Cell) $\rightarrow$... $\rightarrow$ MAZ $\rightarrow$... $\rightarrow$ NCR1 (NK Cell)	KIT (on Mast Cell)	MAZ (in NK Cell)	NCR1 (on NK Cell)	Mast cell signaling activates MAZ-dependent transcription in NK cells, modulating their cytotoxicity.
(Antigen) $\rightarrow$ CD4 $\rightarrow$ PPARG $\rightarrow$ (Metabolites) $\rightarrow$ MYC	CD4 T-Cell	PPARG	MYC (in Malignant Cell)	T-cell metabolic reprogramming via PPARG alters the TME to support malignant proliferation.
EGFR $\rightarrow$ (MAPK Cascade) $\rightarrow$ ELK1 $\rightarrow$ VIM	EGFR	ELK1	VIM	EGFR-MAPK signaling activates transcription factor ELK1 to directly drive the EMT program.
EGFR $\rightarrow$ EGR1 $\rightarrow$ MYC	EGFR	EGR1	MYC	Acinar EGFR directly induces EGR1 to transcriptionally amplify MYC in malignant cells.
EGFR $\rightarrow$ ELK1 $\rightarrow$ NCR1	EGFR	ELK1	NCR1	EGFR-driven MAPK cascade activates ELK1 to induce NK-activating receptor NCR1.
NCR1 $\leftrightarrow$ MAX $\leftrightarrow$ STAT3	NCR1	MAX	STAT3	NK-derived signal converges on MAX to modulate STAT3-driven iCAF activation.
KIT $\leftrightarrow$ MAZ $\leftrightarrow$ NCR1	KIT	MAZ	NCR1	Mast cell KIT signaling co-opts MAZ to alter NK cell surveillance capacity.
AXL $\leftrightarrow$ NR3C1 $\leftrightarrow$ MYC	AXL	NR3C1	MYC	Basal-like AXL engages glucocorticoid receptor to reinforce MYC-driven proliferation.

Continued on next page

Sequence	Upstream Core Nodes		Downstream	Biological Rationale
CD4 → PPARG → MYC	CD4	PPARG	MYC	Helper T cell cues activate PPARG, linking lipid metabolism to MYC in malignant cells.
EGFR ↔ CTCF ↔ STAT3	EGFR	CTCF	STAT3	EGFR-STAT3 communication relies on CTCF-mediated chromatin looping for signal fidelity.
MYC ↔ ESR2 ↔ RARA	MYC	ESR2	RARA	MYC intersects with estrogen receptor β to modulate retinoic acid signaling in fibroblasts.

Table S4: PDAC upgraded therapy combinations

Targets	Therapy Design	Intervention Rationale
ERBB2 + EGFR + MYC	Dual EGFR/ERBB2 blockade with neratinib or lapatinib + OMO-103 (MYC inhibitor /MYC-MAX disruptor,investigational)	Shuts receptor redundancy while extinguishing terminal MYC dependency.
EGFR + HSP90AA1 + APEX1	EGFR TKI (for example osimertinib)+ HSP90 inhibitor ganetespib or PU-H71 + APE1/Ref-1 inhibitor APX3330 (investigational)	Destabilizes receptor signaling and redox/chaperone buffering simultaneously.
MYC + MAX + RARA + STAT3	OMO-103 (MYC-MAX disruptor,investigational)+ all-trans retinoic acid/tretinoin + napabucasin (STAT3-pathway inhibitor, investigational)	Breaks malignant MYC/MAX output and stromal inflammatory transcriptional coupling.
KIT + STAT3 + IL6	KIT inhibitor imatinib or avapritinib + napabucasin + anti-IL6 antibody siltuximab	Collapses receptor-driven fibroinflammatory feed-forward signaling.
PDGFRB + STAT3 + IL6	PDGFR-directed kinase inhibition with imatinib + napabucasin + siltuximab	Disables perivascular signaling before stromal STAT3-IL6 reinforcement.
AXL + NR3C1 + STAT3	Bemcentinib (investigational AXL inhibitor)+ mifepristone (glucocorticoid receptor antagonist) + napabucasin	Disrupts EMT persistence by removing stress-response and inflammatory co-signaling.
EGFR + PLK1 + AURKA	EGFR TKI + onvansertib (PLK1 inhibitor,investigational)+ alisertib (AURKA inhibitor,investigational)	Blocks receptor-driven cell-cycle entry and mitotic completion.
PPARG + MAP2K1 + MYC	PPAR γ modulation with pioglitazone (repurposed)+ MEK inhibitor trametinib or cobimetinib + OMO-103	Reprograms bridge-state signaling while suppressing downstream mitogenic addition.
MAX + STAT3 + IL6	OMO-103 + napabucasin + siltuximab	Collapses immune-contact transcription-cytokine relay sustaining inflammatory stroma.
EGFR,MYC	EGFR inhibitor (e.g.,Erlotinib)+ BET inhibitor (e.g.,OTX015)	Dual blockade of upstream oncogenic driver and its downstream transcriptional amplifier.
AXL,ERBB2	AXL inhibitor (e.g.,Bemcentinib) + HER2/ERBB2 inhibitor (e.g., Trastuzumab)	Prevent compensatory signaling by co-targeting receptor tyrosine kinase (RTK)crosstalk.
STAT3, MAX/MYC	STAT3 inhibitor (e.g.,STATTIC)+ BET inhibitor (e.g.,JQ1)	Inhibit inflammatory signaling and its convergence on the core MYC-MAX proliferation machinery.
APEX1, HSP90AA1	APEX1 inhibitor (e.g.,E3330)+ HSP90 inhibitor (e.g.,Ganetespib)	Induce redox stress/DNA damage while blocking the essential chaperone-mediated survival response.

Continued on next page

Targets	Therapy Design	Intervention Rationale
PDGFRB, STAT5B	PDGFR inhibitor (e.g., Imatinib) + STAT5 inhibitor	Remodel the tumor stroma while directly reversing T-cell immunosuppression.
AXL, NR3C1	AXL inhibitor (e.g., Bemcentinib) + GR Antagonist (repurposed Mifepristone)	Block both RTK-driven proliferation and hormone receptor-mediated transcriptional bypass mechanisms.
EGFR, MAP2K1	EGFR inhibitor (e.g., Gefitinib) + MEK inhibitor (e.g., Trametinib)	Vertical pathway inhibition to block both proliferation and ELK1-mediated EMT.
EGFR, EGR1, MYC	EGFR: Tyrosine Kinase Inhibitor (e.g., Erlotinib); EGR1: HDAC inhibitor (e.g., Vorinostat); MYC: BET Bromodomain Inhibitor (e.g., JQ1 analog)	Disrupt acinar-derived EGFR→EGR1→MYC axis; block proliferation and lineage plasticity.
NCR1, STAT3	NCR1: IL-15 superagonist (e.g., ALT-803); STAT3: Small molecule STAT3 dimerization inhibitor (e.g., Napabucasin analog)	Activate NK NCR1 while simultaneously blocking iCAF STAT3 inflammatory feed-forward loop.
KIT, MAZ	KIT: Tyrosine Kinase Inhibitor (e.g., Imatinib); MAZ: G-quadruplex stabilizer (e.g., Quarfloxin analog)	Disarm mast cell KIT signaling and abrogate MAZ-mediated transcriptional support for immune evasion.
AXL, MYC	AXL: Tyrosine Kinase Inhibitor (e.g., Bemcentinib); MYC: BET Bromodomain Inhibitor (e.g., JQ1 analog)	Combine AXL blockade with MYC transcriptional repression to collapse basal-like squamous program.
CD4, PPARG	CD4: Immune modulator (e.g., low-dose IL-2); PPARG: Thiazolidinedione agonist (e.g., Pioglitazone)	Enhance Treg CD4 activation of PPARG to enforce quiescent fibroblast differentiation over malignancy.
EGFR, CTCF	EGFR: Tyrosine Kinase Inhibitor (e.g., Erlotinib); CTCF: PARP inhibitor (e.g., Olaparib)	Force synthetic lethality by disrupting EGFR signaling and CTCF-mediated chromatin insulation repair.
MYC, RARA	MYC: BET Bromodomain Inhibitor (e.g., JQ1 analog); RARA: Retinoid X receptor agonist (e.g., Bexarotene)	Co-target MYC addiction and RARA-mediated stromal reprogramming in quiescent fibroblasts.

Table S5: PDAC target-drug evidence ledger (deduplicated)

Target	Candidate Drug	Drug Class	Evidence Citation (PMID/DOI/Author-Year)	Direction	Status
APEX1	APX3330	APE1/Ref-1 inhibitor	PMID 18645011 (Fishel 2008)	support	investigational
APEX1	E3330	APEX1 redox inhibitor	Fishel et al.,2011 (PMID:21768412)	support	preclinical
AURKA	Alisertib	Aurora kinase A inhibitor	PMID 25632225 (Wang 2015)	support	investigational
AXL	Bemcentinib	AXL tyrosine kinase inhibitor	PMID 29180468 /DOI 10.1158/0008-5472.CAN-17-1973 (Ludwig 2018)	support	investigational
AXL	Bemcentinib (R428)	Tyrosine Kinase Inhibitor (TKI)	Holland et al. 2010,PMID 20103617	support	investigational
CD4	Aldesleukin (low-dose IL-2)	Cytokine (Immunomodulator)	Koreth et al. 2011,PMID 21998206	support	approved
CD4	No validated drug	None validated	No direct evidence found	mixed	unknown
CFTR	Ivacaftor	CFTR potentiator	No direct evidence found	mixed	approved
CTCF	No validated drug	None validated	No direct evidence found	mixed	unknown
CTCF	Olaparib	PARP Inhibitor	Ong et al. 2013,PMID 23355391	support	approved
CXCR6	No validated drug	None validated	No direct evidence found	mixed	unknown
EGFR	Erlotinib	EGFR tyrosine kinase inhibitor	PMID 17452677 (Moore 2007)	support	approved
EGR1	No validated drug	None validated	No direct evidence found	mixed	unknown
EGR1	Vorinostat (SAHA)	HDAC inhibitor	Kang et al. 2006,PMID 16908059	support	preclinical
ELK1	No validated drug	None validated	No direct evidence found	mixed	unknown
ERBB2	Trastuzumab	HER2/ERBB2 antibody	Safran et al.,2013 (PMID:23959918)	support	approved
ESR2	No validated drug	None validated	No direct evidence found	mixed	unknown
FLT3	Gilteritinib	FLT3 tyrosine kinase inhibitor	DOI 10.1126/sciimmunol.adp3978 (2025)	contradict	approved
HSP90AA1	Ganetespib	HSP90 inhibitor	PMID 28301350 /DOI 10.1097/COC.0000000000000377 (Cardin 2018)	contradict	investigational
IFNG	Emapalumab	anti-IFN- γ monoclonal antibody	PMID 28405527 /DOI 10.1080/2162402X.2017.1291106 (Lu 2017)	contradict	approved
IL6	Siltuximab	anti-IL-6 monoclonal antibody	DOI 10.1158/1535-7163.MCT-16-0899 (Long 2017)	support	approved
KIT	Imatinib	KIT/PDGFR tyrosine kinase inhibitor	PMID 18437079 (Gharibo 2008)	contradict	approved
KLF11	No validated drug	None validated	No direct evidence found	mixed	unknown
MAP2K1	Trametinib	MEK1/2 inhibitor	PMID 24915778 (Infante 2014)	contradict	approved
MAX	No validated drug	None validated	No direct evidence found	mixed	unknown
MAX	OMO-103	MYC-MAX disruptor	PMID 38321218 (Garraalda 2024)	support	investigational
MAZ	Quarflloxin (CX-3543)	G-Quadruplex Stabilizer	Drygin et al. 2009,PMID 19276382	mixed	preclinical
MK167	No validated drug	None validated	No direct evidence found	mixed	unknown
MYC	Birabresib (OTX015)	BET Bromodomain Inhibitor	Sahai et al. 2016,PMID 27355368	support	investigational
MYC	JQ1 /OTX015	BET inhibitor	Sahai et al.,2014 (PMID:25119024)	support	investigational
MYC	OMO-103	MYC inhibitor /MYC-MAX disruptor	PMID 38321218 (Garraalda 2024)	support	investigational
NCR1	N-803 (ALT-803)	IL-15 Superagonist	Rosario et al. 2016,PMID 26527552	support	investigational
NCR1	No validated drug	None validated	No direct evidence found	mixed	unknown
NFIL3	No validated drug	None validated	No direct evidence found	mixed	unknown
NGFR	No validated drug	None validated	No direct evidence found	mixed	unknown
NR3C1	Mifepristone	Glucocorticoid receptor antagonist	PMID 34873175 (Deng 2021)	mixed	approved
PDGFRB	Imatinib	PDGFR/KIT tyrosine kinase inhibitor	PMID 36384875;PMID 15893416 (Shi 2023;Chen 2006)	mixed	approved
PLK1	Onvansertib	PLK1 inhibitor	PMID 35950917 /DOI 10.1158/0008-5472.CAN-22-0018 (Zhang 2022)	support	investigational
PPARG	Pioglitazone	PPAR γ agonist	PMID 25364454 (Ninomiya 2014)	support	approved
RARA	All-trans retinoic acid (tretinoin)	RAR agonist	PMID 32973176 /DOI 10.1038/s41467-020-18636-w (Kocher 2020)	support	approved
RARA	Bexarotene	Rexinoid (RXR Agonist)	Boehm et al. 1994,PMID 8150206	support	approved
STAT3	Napabucasin	STAT3-pathway inhibitor	PMID 36969338 (Bekaii-Saab 2023)	mixed	investigational

Continued on next page

Target	Candidate Drug	Drug Class	Evidence Citation (PMID/DOI/Author-Year)	Direction	Status
STAT3	Napabucasin (BBI608)	STAT3 Dimerization Inhibitor	Li et al. 2015,PMID 26211891	support	investigational
STAT5B	No validated drug	None validated	No direct evidence found	mixed	unknown
STAT5B	Pimozide	STAT5 inhibitor (repurposed)	Redell et al.,2011 (PMID:21415250)	support	preclinical
TCF7	No validated drug	None validated	No direct evidence found	mixed	unknown
VIM	No validated drug	None validated	No direct evidence found	mixed	unknown

Table S6: AD cell-specific targets

Cell Type	Target Symbol	Target Class	Evidence Level & Citation
Antigen-Presenting Microglia (MHC-II+ MG)	CIITA	Key Driver Gene (Coactivator)	Mathys et al.,2019
Antigen-Presenting Microglia (MHC-II+ MG)	HLA-DRA	Surface Marker (MHC Complex)	Gate et al.,2020
Arterial endothelial cell	EFNB2	Identity driver	Human brain vascular atlas in AD context (Yang 2022)(jdc.jefferson.edu)
Arterial endothelial cell	GJA5	Surface marker	Human brain vascular atlas in AD context (Yang 2022)(jdc.jefferson.edu)
Arterial endothelial cell	TNFRSF21	Receptor	AD mouse + brain endothelial validation (Huang 2024;PMID 38609388)
Astrocyte.DAA	GFAP	Intermediate Filament	PMID: 29086448 (Mathys et al., 2019)
Astrocyte.Trophic	BDNF	Neurotrophin	PMID: 28671697 (Choi et al., 2018)
Astrocyte.Trophic	SLC7A10	Amino Acid Transporter	PMID: 23575632 (Sakimura et al.,2013)
Capillary endothelial cell	LRP1	Endocytic receptor	Mechanistic BBB clearance study (Storck 2016;PMID 26619118)
Capillary endothelial cell	MFSD2A	Surface transporter marker	Human brain vascular atlas in AD context (Yang 2022)
Capillary endothelial cell	TEK	Receptor tyrosine kinase	Human multiregion AD vascular snRNA-seq (Sun 2023;PMID 37264161)
Capillary Endothelial Cells (CAP-ENDO)	PECAM1	Adhesion Molecule (Surface Marker)	Clinical Pathology /Sweeney et al.,2018
Capillary Endothelial Cells (CAP-ENDO)	SLC2A1	Transporter (Driver Gene)	Clinical /Winkler et al.,2015
Choroid plexus epithelial cell	AQP1	Surface water-channel marker	AD choroid-plexus literature / CP biology (Kant 2018;Wostyn 2011)
Choroid plexus epithelial cell	TTR	Secreted chaperone / driver	AD choroid-plexus functional literature (Wostyn 2011;Delvenne 2024)
Corticothalamic /L6b glutamatergic neuron	CTGF	Matricellular protein (extracellular marker)	Single-soma AD transcriptomics; Otero-Garcia 2022 (PMID: 35882228)

Continued on next page

Cell Type	Target Symbol	Target Class	Evidence Level & Citation
Corticothalamic /L6b glutamatergic neuron	NPTX2	Secreted synaptic organizer (therapeutic target)	AD-model rescue plus human AD biomarker evidence; Papanikolaou 2025 (PMID: 40345184);Xiao 2017 (PMID: 28440221)
Deep-Layer Glutamatergic Neurons (L5/L6-EXC)	MAPT	Intracellular Protein (Therapeutic Target)	Goedert et al.,2006
Dentate gyrus granule neuron	NTRK2	Receptor tyrosine kinase / TrkB (surface marker)	Human/animal AD evidence for TrkB impairment and rescue; Fonseca-Gomes 2024 (PMID: 39205389);Wang 2025 (PMID: 41163643)
Dentate gyrus granule neuron	PPARG	Nuclear receptor (therapeutic target)	Mouse dentate-granule rescue study;Nenov 2015 (PMID: 25540218)
Dentate gyrus granule neuron	PROX1	Homeobox transcription factor (driver)	Human AD hippocampus neurogenesis study;Cao 2024 (PMID: 38012054)
Differentiating oligodendroglial cell	GPR17	Surface marker - GPCR	AD remyelination lineage context;Tong 2022,PMID 35865954.
Disease-Associated Microglia (DAM)	CLEC7A	Surface Marker (Receptor)	Keren-Shaul et al.,2017
Disease-Associated Oligodendrocytes (DAO)	C4B	Complement Protein (Surface Marker)	Clinical scRNA-seq /Kenigsbuch et al.,2022
ECM / structural pericyte	COL4A1	ECM driver gene	Human multiregion AD vascular snRNA-seq + tissue validation (Sun 2023;PMID 37264161)
ECM / structural pericyte	SMAD3	Transcriptional effector	Human AD gliovascular transcriptomics (İş 2024)(Nature)
Endothelial_Arterial	CLDN5	Tight Junction Protein	PMID: 32327568
Endothelial_Capillary	ABCB1	Efflux Transporter	PMID: 19763178
Endothelial_Venous	ICAM1	Adhesion Molecule	PMID: 31391558
Endothelial_Venous	SELP	Adhesion Molecule	PMID: 31806693
Entorhinal/ Hippocampal Glutamatergic Neurons (HC-EXC)	EPHA4	Receptor Tyrosine Kinase (Therapeutic Target)	Fu et al.,2014

Continued on next page

Cell Type	Target Symbol	Target Class	Evidence Level & Citation
Ependymal cell	CD24	Surface marker	AD border-cell atlas + canonical human ependymal marker (Sun 2023;HPA/Siletti 2023)
Ependymal cell	FOXJ1	Transcription factor	Canonical human ependymal driver;relevant cell class captured in AD atlas (MacDonald 2021; Sun 2023)(Frontiers)
Ependymal cell	INSR	Receptor	Human multiregion AD vascular snRNA-seq (Sun 2023;PMID 37264161)
Ependymal_Cell	DNAH5	Motor Protein	PMID: 31235610
Ex_Neuron_ Hippocampal_ CA1_CA3	NRGN	Surface Marker /Calmodulin-Binding	Clinical Biomarker;PMID: 25611512
Ex_Neuron_ Hippocampal_ CA1_CA3	PRNP	Surface Receptor / Driver	Functional Driver;PMID: 19165176
Ex_Neuron_ Hippocampal_ CA1_CA3	SIGMAR1	Receptor / Chaperone	Emerging Target;PMID: 29077933
Ex_Neuron_L2_ 3.IT	CADM1	Surface Marker /Adhesion	snRNA-seq Marker;Mathys et al.,2019 (PMID: 31042697)
Ex_Neuron_L5_ 6_CT_PT	KCNQ2	Ion Channel	Preclinical AD Model;PMID: 30337467
Ex_Neuron_L5_ 6_CT_PT	PCDH11X	Surface Marker /Adhesion	Genetics/Linkage;PMID: 21494639
Extratelencephalic projecting glutamatergic neuron (ET)	FEZF2	Transcription factor (driver)	AD-model Fezf2+ ET neuron study;Sun 2022 (PMID: 35194025)
Extratelencephalic projecting glutamatergic neuron (ET)	KCNG3	Voltage-gated K+ channel (therapeutic target)	Mouse AD interventional evidence;Zhang 2022 (PMID: 36323265)
Extratelencephalic projecting glutamatergic neuron (ET)	FHY1	GPI-anchored glycoprotein (surface marker)	AD-model circuit study using Thy1-labeled ET neurons;Sun 2022 (PMID: 35194025)
Fibroblast_ Perivascular	COL1A1	Extracellular Matrix	PMID: 29899444
Fibroblast_ Perivascular	DCN	Proteoglycan	PMID: 35318340
Fibroblasts /Vascular Smooth Muscle Cells (FIB/ vSMC)	ACTA2	Cytoskeletal Protein (Surface Marker)	Clinical scRNA-seq /Yang et al., 2022
Hippocampal- entorhinal pyramidal neuron	ATP8B1	P4-ATPase membrane flip-pase (membrane marker)	Human resilience atlas;Castanho 2025 (PMID: 41035073)

Continued on next page

Cell Type	Target Symbol	Target Class	Evidence Level & Citation
Hippocampal-entorhinal pyramidal neuron	MEF2C	Transcription factor (driver)	Human resilience atlas;Castanho 2025 (PMID: 41035073)
Hippocampal-entorhinal pyramidal neuron	RELN	Secreted glycoprotein (therapeutic pathway target)	Human multiregion atlas plus mechanistic AD studies; Mathys 2024 (PMID: 39048816) ;Kobro-Flatmoen 2023 (PMID: 37091586)
Homeostatic astrocyte	AQP4	Therapeutic target - Water channel	Preclinical AD mouse;Yang 2012, PMID 22365952;Smith 2019, PMID 31068220.
Homeostatic astrocyte	GLUL	Driver gene - Metabolic enzyme	Human AD astrocyte snRNA-seq;Serrano-Pozo 2024,PMID 39528672. (Nature)
Homeostatic astrocyte	SLC1A2	Surface marker - Glutamate transporter	Human AD astrocyte snRNA-seq;Serrano-Pozo 2024,PMID 39528672. (Nature)
Homeostatic Astrocytes (HOM-AST)	SOX9	Key Driver Gene (Transcription Factor)	Sun et al.,2017
Homeostatic microglia	P2RY12	Surface marker - Purinergic receptor	Human AD brain;Walker 2020, PMID 31968618.
Homeostatic microglia	TMEM119	Therapeutic target - Surface transmembrane protein	Human AD brain + 5xFAD;Liu 2025,PMID 40373772.
Homeostatic Microglia (HOM-MG)	TGFBR1	Key Driver Gene (Receptor)	Butovsky et al.,2014
Homeostatic OPC	CHRM1	Therapeutic target - Muscarinic receptor	Preclinical AD mouse;Chen 2021,PMID 34102111.
Homeostatic OPC	OLIG2	Driver gene - Transcription factor	AD remyelination studies;Tong 2022,PMID 35865954;Chen 2021, PMID 34102111.
Homeostatic OPC	PDGFRA	Surface marker - Receptor tyrosine kinase	AD mouse + lineage analysis;Chacón-De-La-Rocha 2020, PMID 33343301.
Inflammatory /cytokine-response microglia	CD33	Surface marker - Siglec receptor	Human genetics + AD models; Griciuc 2013;Griciuc 2020.
Inflammatory /cytokine-response microglia	INPP5D	Driver gene - Phosphatase	Human AD brain and AD models;Tsai 2021,PMID 33631273; Chou 2023,PMID 38016942.

Continued on next page

Cell Type	Target Symbol	Target Class	Evidence Level & Citation
Inflammatory /cytokine-response microglia	NLRP3	Therapeutic target - Inflammasome sensor	Human/genetic/preclinical AD evidence;Tan 2013,PMID 23686772;Hanslik 2020.
Inh_Neuron_PVALB	KCNC1	Ion Channel	Functional Target;PMID: 20147564
Inh_Neuron_SST	LHX6	Transcription Factor /Driver	Developmental Driver;PMID: 30413645
Inh_Neuron_SST	NPY	Neuropeptide / Co-Transmitter	Functional Driver;PMID: 22015920
Inh_Neuron_SST	SSTR2	GPCR	Preclinical Validation;PMID: 31611707
Inh_Neuron_VIP	CALB2	Calcium-Binding Protein /Marker	Cell Identity Marker;PMID: 28892079
Inh_Neuron_VIP	VIPR1	GPCR	Functional Target;PMID: 32424049
Interferon-response microglia	IFITM3	Surface marker - Interferon-induced transmembrane protein	Human late-onset AD + mouse models;Hur 2020,PMID 32879487.
Interferon-response microglia	IFNAR1	Therapeutic target - Type I interferon receptor	Preclinical AD models;Roy 2020; Roy 2022,PMID 35443157.
Interferon-Responsive Microglia (IRM)	IRF7	Key Driver Gene (Transcription Factor)	Mathys et al.,2019
Intermediate astrocyte	ETNPPL	Driver gene - Metabolic enzyme	Human AD astrocyte snRNA-seq;Serrano-Pozo 2024,PMID 39528672. (Nature)
Intermediate astrocyte	STAT3	Therapeutic target - Transcription factor	Preclinical AD mouse; Haim 2015,PMID 25673868; Reichenbach 2019.
Intratellencephalic projecting glutamatergic neuron (IT)	KCNH7	Ion channel (therapeutic candidate)	Human multi-omic nomination in AD-vulnerable IT neurons;Rexach 2024 (PMID: 39265576);Dharshini 2026 (PMID: 41620473)
Intratellencephalic projecting glutamatergic neuron (IT)	OPCML	Cell-adhesion GPI protein (surface marker)	Human snRNA/snATAC marker of vulnerable AD IT neurons; Rexach 2024 (PMID: 39265576)
Intratellencephalic projecting glutamatergic neuron (IT)	RORB	Transcription factor (driver)	Human EC/cortex snRNA-seq plus neuropathology;Leng 2021 (PMID: 33432193);Dharshini 2026 (PMID: 41620473)

Continued on next page

Cell Type	Target Symbol	Target Class	Evidence Level & Citation
LAMP5/ PAX6/RELN superficial interneuron	LAMP5	Lysosome- associated membrane pro- tein (key driver gene)	Preclinical mechanistic AD evidence;Deng 2022,PMID: 35780436.
LAMP5/ PAX6/RELN superficial interneuron	NMBR	GPCR (surface marker)	Human cortical taxonomy;Hodge 2019.
Lipid-processing /DAM-like microglia	GPNMB	Surface marker - Glycoprotein	Human microglial state atlases; Sun 2023,PMID 37774678; Gerrits 2021,PMID 33609158.
Lipid-processing /DAM-like microglia	TREM2	Therapeutic target - Immune receptor	Human AD genetics + models; Nguyen 2020,PMID 32840654; Mancuso 2024.
Mature myelinating oligodendrocyte	BACE1	Therapeutic tar- get - Aspartyl protease	Emerging human/mouse AD evidence;Ishii 2024,PMID 39548583.
Mature myelinating oligodendrocyte	BIN1	Driver gene - AD risk gene / adaptor protein	Human AD brain;De Rossi 2016, PMID 27488240.
Mature Oligo- dendrocytes (MAT-OLIG)	MOG	Surface Glyco- protein (Surface Marker)	Clinical scRNA-seq /Mathys et al.,2019
Meningeal fibroblast	SLC26A2	Solute-carrier driver	Human AD leptomeningeal snRNA-seq (Kearns 2023)(Nature)
Meningeal fibroblast	SLC4A4	Surface transporter marker	Human brain vascular atlas in AD context (Yang 2022)(ResearchGate)
Meningeal fibroblast	TGFB1	Cytokine pathway target	Human AD leptomeningeal snRNA-seq + CellChat (Kearns 2023)(Nature)
Microglia_DAM	LPL	Lipase	PMID: 34038742 (Marschallinger et al.,2020)
Microglia_ Homeostatic	CX3CR1	GPCR	PMID: 20392936 (Lee et al., 2010)
Oligodendrocyte Progenitor Cells (OPC)	SOX10	Transcription Factor (Driver Gene)	Clinical scRNA-seq /Grubman et al.,2019
Oligodendrocyte_ DOL	B2M	MHC Class I Component	PMID: 34648354 (Zhou et al., 2020)
Oligodendrocyte_ DOL	FOS	Transcription Factor	PMID: 35314834 (Pandey et al., 2022)
Oligodendrocyte_ Mature	MBP	Myelin Structural Protein	PMID: 33116353 (Kenigsbuch et al.,2020)
Parvalbumin- Positive GABAergic Interneurons (PV-INH)	PPARGC1A	Transcriptional Coactivator (Key Driver)	Katsouri et al.,2011

Continued on next page

Cell Type	Target Symbol	Target Class	Evidence Level & Citation
Perivascular fibroblast / VLMC	ABCA8	Surface transporter marker	Human brain vascular atlas in AD context (Yang 2022)(ResearchGate)
Perivascular fibroblast / VLMC	CEMIP	ECM-remodeling driver	Human AD gliovascular / vascular atlas data (İş 2024;Yang 2022)(Nature)
Perivascular fibroblast / VLMC	SPP1	Secreted signaling target	AD mouse + human perivascular study (De Schepper 2023)(Nature)
Perivascular macrophage /border-associated macrophage	CD163	Surface marker	Human leptomeningeal/BAM validation (Kearns 2023;De Schepper 2023)(Nature)
Perivascular macrophage /border-associated macrophage	CD36	Scavenger receptor	AD neurovascular mechanism study (Park 2017;PMID 28515043)
Proliferating /cycling microglia	CSF1R	Therapeutic target - Receptor tyrosine kinase	Preclinical AD/tauopathy models;Olmos-Alonso 2016,PMID 26747862;Mancuso 2019,PMID 31504240.
Proliferating /cycling microglia	MKI67	Driver gene - Cell-cycle protein	Human microglial state atlas;Sun 2023,PMID 37774678.
PVALB interneuron	ERBB4	Receptor tyrosine kinase (surface marker)	PV-IN proteomics + AD vulnerability mapping;Kumar 2024; Gabitto 2024. (Nature)
PVALB interneuron	PVALB	Calcium-binding protein (key driver gene)	Human AD atlas;Gabbito 2024. (Vizgen)
PVALB interneuron	SCN1A	Voltage-gated sodium channel	Preclinical mechanistic AD evidence;Verret 2012;Martinez-Losa 2018,PMID: 29551491.
Rare adaptive immune T cell infiltrate	CD8A	Surface marker	Human AD CSF/leptomeningeal studies (Gate 2020;PMID 31915375;Hobson 2026;PMID 41593242)
Rare adaptive immune T cell infiltrate	IFNG	Cytokine effector gene	Human AD leptomeningeal immune profiling (Hobson 2026; Kearns 2023)(ResearchGate)
Rare adaptive immune T cell infiltrate	IL17A	Cytokine target	AD meningeal T-cell study (Brigas 2021;PMID 34469732)
Reactive / remyelinating OPC	CSPG4	Surface marker - Proteoglycan (NG2)	AD mouse OPC analyses; Chacón-De-La-Rocha 2020, PMID 33343301.
Reactive / remyelinating OPC	LINGO1	Therapeutic target - Trans-membrane receptor	Preclinical AD mouse;Wu 2018, PMID 29427384;Yang 2022, PMID 35014704.

Continued on next page

Cell Type	Target Symbol	Target Class	Evidence Level & Citation
Reactive / remyelinating OPC	MYRF	Driver gene - Transcription factor	AD remyelination studies;Tong 2022,PMID 35865954.
Reactive / stress-associated oligodendrocyte	APOE	Driver gene - Lipid transport protein	Mouse + human AD oligo-dendrocytes;Park 2023,PMID 36781874.
Reactive / stress-associated oligodendrocyte	CD74	Surface marker - MHC-II chaperone	Mouse + human AD oligo-dendrocytes;Park 2023,PMID 36781874.
Reactive / stress-associated oligodendrocyte	MAPK1	Therapeutic target - ERK1/2 kinases	Mouse + human AD oligo-dendrocytes;Park 2023,PMID 36781874.
Reactive / stress-associated oligodendrocyte	MAPK3	Therapeutic target - ERK1/2 kinases	Mouse + human AD oligo-dendrocytes;Park 2023,PMID 36781874.
Reactive astrocyte R0	COL21A1	Surface marker - ECM protein	Human AD astrocyte snRNA-seq;Serrano-Pozo 2024,PMID 39528672. (Nature)
Reactive astrocyte R0	LAMA2	Driver gene - ECM glycoprotein	Human AD astrocyte snRNA-seq;Serrano-Pozo 2024,PMID 39528672. (Nature)
Reactive astrocyte R1	C3	Therapeutic target - Complement component	Human AD brain + AD models; Lian 2015,PMID 25533482;Wu 2019,PMID 31433986.
Reactive astrocyte R1	CD44	Surface marker - Adhesion receptor	Human AD astrocyte snRNA-seq;Serrano-Pozo 2024,PMID 39528672. (Nature)
Reactive astrocyte R2	MAOB	Therapeutic target - Enzyme	Preclinical AD mouse;Jo 2014; Nam 2024.
Reactive astrocyte R2	SERPINA3	Driver gene - Secreted serpin	Human spatial AD data + AD mouse intervention;Karaahmet 2024;Liu 2025,PMID 41345651.
SNCG interneuron	RXRG	Nuclear receptor	Emerging/preclinical AD evidence;Mandrekar-Colucci 2011; Biyong 2021;SNCG marker-set linkage.
SNCG interneuron	SNCG	Synuclein family protein (key driver gene)	Human AD atlas;Gabbitto 2024.
SNCG interneuron	THSD7B	Membrane glycoprotein (surface marker)	Human neocortical marker set. (EMBL-EBI)
Somatostatin-Positive GABAergic Interneurons (SST-INH)	GABRA5	GABA-A Receptor Sub-unit (Surface Marker)	Ali et al.,2019

Continued on next page

Cell Type	Target Symbol	Target Class	Evidence Level & Citation
Somatostatin-Positive GABAergic Interneurons (SST-INH)	MME	Metalloprotease (Key Driver)	Saito et al.,2005
SST interneuron	SST	Neuropeptide (key driver gene)	Human AD atlas;Gabbitto 2024. (Vizgen)
SST interneuron	SSTR4	GPCR	Preclinical AD evidence;Sandoval 2012/2013,PMID: 22449380, 23669069.
Subcortical_Neuron_Cholinergic_BF	CHAT	Enzyme / Marker	Diagnostic Marker/Driver;PMID: 9832156
Subcortical_Neuron_Cholinergic_BF	NGF	Growth Factor	Validated Target (Gene Therapy);CERE-110 Phase II;PMID: 25136048
Subcortical_Neuron_Cholinergic_BF	NTRK1	Receptor Tyrosine Kinase	Driver of Vulnerability;PMID: 16965557
Subcortical_Neuron_Noradrenergic_LC	ADRA2A	GPCR /Surface Marker	Target/Modulator;PMID: 23637676
Subcortical_Neuron_Noradrenergic_LC	SLC6A2	Transporter	Validated Target (Phase II/III); Atomoxetine;PMID: 31531234
Subcortical_Neuron_Noradrenergic_LC	TH	Enzyme /Driver	Functional Driver;PMID: 29632247
Transport pericyte	ABCC9	Transporter / identity driver	Human vascular subtype atlas in AD context (Yang 2022)(Cell)
Transport pericyte	PDGFRB	Surface receptor marker	Human AD pericyte study + multiregion vascular snRNA-seq (Smyth 2022;Sun 2023)
Transport pericyte	PPIA	Chaperone / inflammatory target	AD mechanistic study (Bell 2012;PMID 22622580)
Trophic-factor / inflammatory-signaling astrocyte	CHI3L1	Therapeutic target - Secreted glycoprotein	Astrocyte-specific AD models; Lananna 2020;Zeng 2023,PMID 38042775.
Trophic-factor / inflammatory-signaling astrocyte	IL1R1	Surface marker - Cytokine receptor	Human AD astrocyte snRNA-seq;Serrano-Pozo 2024,PMID 39528672. (Nature)
Trophic-factor / inflammatory-signaling astrocyte	SOCS3	Driver gene - Signaling regulator	Human AD astrocyte snRNA-seq;Serrano-Pozo 2024,PMID 39528672. (Nature)

Continued on next page

Cell Type	Target Symbol	Target Class	Evidence Level & Citation
Upper-Layer Glutamatergic Neurons (L2/L3-EXC)	GRIN2B	NMDA Receptor Subunit (Therapeutic Target)	Li et al.,2011
Upper-Layer Glutamatergic Neurons (L2/L3-EXC)	REST	Transcription Factor (Key Driver)	Lu et al.,2014
Upper-Layer Glutamatergic Neurons (L2/L3-EXC)	SLC17A7	Vesicular Transporter (Surface Marker)	Kashani et al.,2008
Vascular smooth muscle cell	MYOCD	Transcription factor	Human vascular cell-state atlas in AD context (Yang 2022)(Nature)
Vascular smooth muscle cell	NOTCH3	Surface receptor marker	Human vascular cell-state atlas in AD context (Yang 2022)(Cell)
Vascular_Smooth_Muscle	SRF	Transcription Factor	PMID: 23083741
Vasoactive Intestinal Peptide-Positive GABAergic Interneurons (VIP-INH)	GAD1	Decarboxylase Enzyme (Key Driver)	Schwab et al.,2013
Vasoactive Intestinal Peptide-Positive GABAergic Interneurons (VIP-INH)	VIPR2	GPCR (Therapeutic Target)	Song et al.,2012
Venous endothelial cell	NR2F2	Transcription factor	Human brain vascular atlas in AD context (Yang 2022)
Venous endothelial cell	VCAM1	Adhesion receptor	A β 42 BBB-endothelium mechanism study (Salian 2024;PMID 39455283)
VIP interneuron	HTR3A	Ligand-gated ion channel (surface marker)	Human cortical interneuron literature + AD atlas;Poorthuis 2018;Gabitto 2024.
VIP interneuron	VIP	Neuropeptide	Preclinical AD evidence;Song 2012,PMID: 22328918;Korkmaz 2019,PMID: 30498985.

Table S7: AD cell-specific pathways

Pathway Name	Source	Associated Cell Type(s)	Key Targets Involved	Citation
ABC Transporter Efflux at BBB	KEGG hsa02010	Endothelial.Capillary	ABCB1	PMID: 19763178
Acute-Phase Response Signaling	MSigDB HALL-MARK-INFLAMMATORY-RESPONSE	Astrocyte.DAA	SERPINA3; GFAP	PMID: 16541075 (Abraham et al.,1990)
Alzheimer disease	KEGG; KEGG hsa05010	Mature myelinating oligodendrocyte;Deep-Layer Glutamatergic Neurons (L5/L6-EXC);Entorhinal/Hippocampal Glutamatergic Neurons (HC-EXC);Ex_Neuron_L2_3.IT;Ex_Neuron_Hippocampal.CA1_CA3	BACE1;MAPT; GRIN2B	Ishii 2024,PMID 39548583. (KEGG);Goedert et al., 2006;Vassar et al.,1999; PMID: 21709374
Amyloid- β clearance	Reactome	Somatostatin-Positive GABAergic Interneurons (SST-INH)	MME	Saito et al.,2005
Antigen processing and presentation	KEGG	Antigen-Presenting Microglia (MHC-II+ MG)	CD74;HLA-DRA;CIITA	Mathys et al.,2019
AP-1 Transcription Factor Network	KEGG hsa04010	Oligodendrocyte.DOL	FOS	PMID: 35314834 (Pandey et al.,2022)
APOE Lipid Transport and Amyloid Deposition	Reactome R-HSA-8964043	Pericyte;Vascular_Smooth_Muscle	APOE	PMID: 34139134
APOE-Lipoprotein Clearance	KEGG hsa05010	Microglia.DAM; Microglia.IRM	APOE;LPL	PMID: 31932797 (Krasemann et al.,2017)
Aquaporin-mediated transport	Reactome	Homeostatic Astrocytes (HOM-AST)	AQP4	Iliff et al.,2012
Astrocytic Glutamate-Glutamine Uptake And Metabolism	Reactome	Homeostatic astrocyte	SLC1A2;GLUL	Dai 2023;Scott 2011,PMID 20416976. (Reactome)
Bicarbonate Transport and CSF pH Homeostasis	Reactome R-HSA-425381	Choroid.Plexus_Epithelial	SLC4A4	PMID: 30635560
Blood-Brain Barrier Tight Junction Dynamics	Reactome R-HSA-8939350	Endothelial.Arterial; Endothelial.Capillary; Endothelial.Venous; Choroid.Plexus_Epithelial	CLDN5; MFSD2A	PMID: 32327568
Calcium Signaling Pathway	KEGG hsa04020	Inh_Neuron_PVALB; Inh_Neuron_VIP	PVALB;CALB2	PMID: 25818302
Cell adhesion molecules (CAMs)	KEGG; KEGG hsa04514	Intratolencephalic-projecting glutamatergic neuron (IT);Venous endothelial cell;Capillary Endothelial Cells (CAP-ENDO);Ex_Neuron_L2_3.IT;Ex_Neuron_L5_6.CT_PT	OPCML; VCAM1; PECAM1; CADM1; PCDH11X	Rexach 2024 (PMID: 39265576);Liu 2012 (PMID: 22017384);Salian 2024;PMID 39455283;Lau 2023;PMID 37735240.;Yousef et al.,2019;PMID: 31042697
Cholinergic synapse	KEGG	Homeostatic OPC	CHRM1	Chen 2021,PMID 34102111. (KEGG)
Circadian Entrainment	KEGG hsa04713	Subcortical_Neuron_Noradrenergic.LC	ADRA2A	PMID: 31531234
Clathrin-mediated endocytosis	Reactome	Mature Oligodendrocytes (MAT-OLIG)	BIN1	Crotti et al.,2019
Complement cascade	Reactome; KEGG hsa04610	Disease-Associated Oligodendrocytes (DAO);Astrocyte.DAA	C4B;C3	Kenigsbuch et al.,2022; PMID: 28166218 (Liddelow et al.,2017)
CSF1R-PI3K Survival Signaling	Reactome R-HSA-9008059	Microglia-Prolif	CSF1R	PMID: 28090338 (Spangenberg et al.,2016)
CX3CR1-CX3CL1 Axis	KEGG hsa04060	Microglia_Homeostatic	CX3CR1	PMID: 20392936 (Lee et al.,2010)
D-Serine Glycine Transport	Reactome R-HSA-351202	Astrocyte.Trophic	SLC7A10	PMID: 23575632 (Sakimura et al.,2013)

Continued on next page

Pathway Name	Source	Associated Cell Type(s)	Key Targets Involved	Citation
Endocytosis	KEGG	Capillary endothelial cell;Vascular smooth muscle cell; Fibroblasts /Vascular Smooth Muscle Cells (FIB/vSMC)	LRP1	Storck 2016;PMID 26619118;Kanekiyo 2012; PMID 23152628.;Kanekiyo et al.,2012
Ephrin signaling	Reactome	Entorhinal/ Hippocampal Glutamatergic Neurons (HC-EXC)	EPHA4	Fu et al.,2014
ErbB signaling pathway	KEGG	Parvalbumin-Positive GABAergic Interneurons (PV-INH)	ERBB4	Yin et al.,2013
Extracellular Matrix Organization in Perivascular Space	Reactome 1474244	Fibroblast_ Perivascular	COL1A1;DCN	PMID: 29899444
GABAergic synapse	KEGG; KEGG hsa04727	Somatostatin-Positive GABAergic Interneurons (SST-INH); Vasoactive Intestinal Peptide-Positive GABAergic Interneurons (VIP-INH);Inh_ Neuron_SST;Inh_ Neuron_PVALB	GABRA5; GAD1;LHX6; KCNC1	Ali et al.,2019;Schwab et al.,2013;PMID: 30413645
Glucose Uptake and Cerebral Hypometabolism	Reactome R-HSA-189200	Endothelial_Arterial	SLC2A1	PMID: 25614620
Glutamatergic Clearance	Reactome R-HSA-210500	Astrocyte_ Homeostatic	SLC1A2	PMID: 26834613 (Takahashi et al.,2015)
Glutamatergic synapse	KEGG; KEGG hsa04724	Homeostatic Astrocytes (HOM-AST) ;Upper-Layer Glutamatergic Neurons (L2/L3-EXC);Ex_ Neuron_L2_3.IT; Ex_Neuron_L5_6_ CT_PT;Ex_Neuron_ Hippocampal_CA1_ CA3	SLC1A2; GRIN2B; SLC17A7; NRGN	Takahashi et al.,2015;Li et al.,2011;PMID: 12604975
Glymphatic Aquaporin-4 Transport	Custom/ Reactome R-HSA-432047	Astrocyte_ Homeostatic	AQP4	PMID: 22896675 (Iliff et al.,2012)
GOBP_ OLIGODENDROCYTES: GO DIFFERENTIATION Biological Process	MSigDB	Reactive / remyelinating OPC; Differentiating oligodendroglial cell	LINGO1; GPR17;MYRF; OLIG2	Tong 2022,PMID 35865954;Wu 2018,PMID 29427384. (GSEA)
GPCR Downstream Signaling	Reactome R-HSA-388396	Inh_Neuron_SST; Subcortical_Neuron_ Noradrenergic_LC	SSTR2; ADRA2A	PMID: 23637676
Hallmark TNFalpha Signaling via NF-kappaB	MSigDB Hallmark	Endothelial_Arterial; Endothelial_Venous	VCAM1; ICAM1;SELP	PMID: 31391558
HALLMARK_ COMPLEMENT	MSigDB Hallmark	Reactive astrocyte R1;Perivascular fibroblast /VLMC; Perivascular macrophage / border-associated macrophage	C3;SPP1	Lian 2015,PMID 25533482; Wu 2019,PMID 31433986. (GSEA);De Schepper 2023; PMID 36747024. (Nature)
HALLMARK_ IL6_JAK_STAT3_ SIGNALING	MSigDB Hallmark	Intermediate astrocyte;Reactive astrocyte R0	STAT3	Haim 2015,PMID 25673868;Ceyzériat 2018. (GSEA)
HALLMARK_ INFLAMMATORY_ RESPONSE	MSigDB Hallmark	Reactive astrocyte R2;Trophic-factor /inflammatory- signaling astrocyte	SERPINA3; CHI3L1	Lananna 2020,PMID 33328329;Liu 2025,PMID 41345651. (GSEA)
HALLMARK_ INTERFERON_α_ RESPONSE	MSigDB	Interferon-Responsive Microglia (IRM)	IFNAR1; IFITM3;IRF7	Roy et al.,2020
IL-17 signaling pathway	KEGG	Rare adaptive immune T cell infiltrate	IL17A	Brigas 2021;PMID 34469732.
Insulin signaling pathway	KEGG	Ependymal cell	INSR	Sun 2023;PMID 37264161.

Continued on next page

Pathway Name	Source	Associated Cell Type(s)	Key Targets Involved	Citation
Interaction of DAP12 and TREM2	Reactome; Reactome R-HSA-2395430	Lipid-processing / DAM-like microglia; Disease-Associated Microglia (DAM); Microglia_DAM	TREM2;APOE; CLEC7A	Nguyen 2020,PMID 32840654. (Reactome);Keren-Shaul et al., 2017;PMID: 28802038 (Keren-Shaul et al.,2017)
Interferon α/β signaling	Reactome; MSigDB HALL-MARK- INTERFERON_ α - RESPONSE	Interferon-response microglia;Choroid plexus epithelial cell; Microglia_IRM	IFNAR1; IFITM3	Roy 2022,PMID 35443157; Hur 2020,PMID 32879487. (GSEA);Baruch 2014; PMID 25147279;Dá Mesquita 2015;PMID 26092102.;PMID: 32499645 (Sala Frigerio et al.,2019)
Interferon γ signaling	Reactome	Rare adaptive immune T cell infiltrate	IFNG;CD8A	Gate 2020;PMID 31915375;Hobson 2026; PMID 41593242.
Interleukin-1 signaling	Reactome	Trophic-factor / inflammatory- signaling astrocyte	IL1R1;SOCS3	Serrano-Pozo 2024, PMID 39528672;Lananna 2020,PMID 33328329. (Reactome)
JAK-STAT signaling pathway	KEGG	Disease-Associated Astrocytes (DAA); Disease-Associated Oligodendrocytes (DAO)	STAT3; SERPINA3	Ceyzeriat et al.,2018;Chen et al.,2020
Long-Term Potentiation (LTP)	KEGG hsa04720	Ex_Neuron_ Hippocampal_CA1_ CA3	GRIN2B;PRNP	PMID: 19165176
LRP1-Mediated Amyloid- β Clearance	Reactome R-HSA-9613829	Endothelial_Capillary	LRP1	PMID: 18568035
MAPK1/MAPK3 signaling	Reactome; KEGG hsa04010	Reactive / stress-associated oligodendrocyte; Subcortical_Neuron_ Cholinergic_BF	MAPK1; MAPK3; NTRK1	Park 2023,PMID 36781874. (Reactome); PMID: 16965557
MECP2 Regulated Neuronal Genes	MSigDB Hallmark: M5936	Ex_Neuron_L5_6_CT_ PT	RORB	PMID: 30382198
MHC class II antigen presentation	Reactome; Reactome R-HSA-2132295	Reactive / stress-associated oligodendrocyte; Antigen-presenting microglia;Microglia_IRM;Oligodendrocyte_ DOL	CD74;B2M	Park 2023,PMID 36781874. (Reactome);Bryan 2008,PMID 18786268;Mancuso 2024. (Reactome);PMID: 34480022 (Chen et al., 2021)
Motile Cilia Formation and CSF Flow	Reactome R-HSA-5617833	Ependymal_Cell	FOXJ1;DNAH5	PMID: 31235610
Myelin Sheath Maintenance	KEGG hsa05022	Oligodendrocyte_ Mature	MBP	PMID: 33116353 (Kenigsbuch et al.,2020)
Neuroactive ligand-receptor interaction	KEGG; KEGG hsa04080	SST interneuron; VIP interneuron; Somatostatin-Positive GABAergic Interneurons (SST-INH); Vasoactive Intestinal Peptide-Positive GABAergic Interneurons (VIP-INH); Inh_Neuron_SST; Inh_Neuron_VIP; Subcortical_Neuron_ Noradrenergic_LC	SST;SSTR4; VIP;VIPR2; SSTR2;VIPR1; ADRA2A	PMID: 33116353 (Watamura 2022,PMID: 34737456;Sandoval 2012, PMID: 22449380. (genome.jp);Song 2012, PMID: 22328918;Korkmaz 2019,PMID: 30498985. (genome.jp);Sandoval et al.,2013;Song et al.,2012; PMID: 31611707
Neurotrophin signaling pathway	KEGG; Reactome R-HSA-9026527; KEGG hsa04722	Dentate gyrus granule neuron;Deep-Layer Glutamatergic Neurons (L5/L6-EXC); Astrocyte_Trophic; Subcortical_Neuron_ Cholinergic_BF	NTRK2;BDNF; NGF;NTRK1	Fonseca-Gomes 2024 (PMID: 39205389);Chen 2008 (PMID: 17918233);Ginsberg et al.,2019; PMID: 28671697 (Choi et al.,2018);PMID: 25136048
NF- κ B signaling pathway	KEGG	Transport pericyte; Pericytes (PERI)	PPIA;PDGFRB	Bell 2012;PMID 22622580; Halliday 2016;PMID 25757756.;Bell et al.,2012
NOD-like receptor signaling pathway	KEGG	Inflammatory / cytokine-response microglia	NLRP3	Tan 2013,PMID 23686772. (KEGG)
NOTCH3 Signaling and Small Vessel Integrity	KEGG hsa04330	Vascular_Smooth_ Muscle	NOTCH3	PMID: 34542266

Continued on next page

Pathway Name	Source	Associated Cell Type(s)	Key Targets Involved	Citation
NR1H3 & NR1H2 regulate gene expression linked to cholesterol transport and efflux	Reactome	Reactive / stress-associated oligodendrocyte;Lipid-processing /DAM-like microglia	APOE	Blanchard 2022;Mok 2023. (Reactome);Krasemann 2017,PMID 28930663; Nguyen 2020,PMID 32840654. (Reactome)
P2Y receptors	Reactome	Homeostatic microglia	P2RY12	Walker 2020,PMID 31968618;Sun 2023,PMID 37774678. (KEGG)
PDGFR-α Signaling in Gliogenesis	Reactome R-HSA-186797	Oligodendrocyte_Precursor.OPC	PDGFRA	PMID: 37353786 (Auguste et al.,2023)
PDGFRB Pericyte Survival Signaling	Reactome R-HSA-186797	Pericyte	PDGFRB	PMID: 27708414
Phase 0 - rapid depolarisation	MSigDB CP:REACTOME	PVALB interneuron	SCN1A	Verret 2012,PMID: 22541439;Martinez-Losa 2018,PMID: 29551491. (GSEA)
Plasma lipoprotein clearance	Reactome	Border-Associated Macrophages (BAM)	APOE;CD36	Mrdjen et al.,2018
Potassium Channels	Reactome; Reactome R-HSA-1296072	Intratelencephalic-projecting glutamatergic neuron (IT); Extratelencephalic-projecting glutamatergic neuron (ET);Ex_Neuron_L5_6_CT_PT;Inh_Neuron_PVALB	KCNH7; KCNG3; KCNQ2; KCNC1	Rexach 2024 (PMID: 39265576);Zhang 2022 (PMID: 36323265);PMID: 30337467
PPAR signaling pathway	KEGG	Dentate gyrus granule neuron	PPARG	Nenov 2015 (PMID: 25540218)
Prion Disease	KEGG hsa05020	Ex_Neuron_Hippocampal_CA1_CA3	PRNP	PMID: 19165176
Protein Processing in Endoplasmic Reticulum	KEGG hsa04141	Ex_Neuron_Hippocampal_CA1_CA3	SIGMAR1	PMID: 29077933
Purinergic GPCR Signaling	Reactome R-HSA-418594	Microglia_Homeostatic	P2RY12	PMID: 30617255 (Cserep et al.,2020)
Reelin signaling pathway	PID/ MSigDB C2; Reactome	Hippocampal-entorhinal pyramidal neuron;LAMP5/ PAX6/RELN superficial interneuron	RELN	Mathys 2024 (PMID: 39048816);Kobro-Flatmoen 2023 (PMID: 37091586) ;Mathys 2024,PMID: 39048816;Lopera 2023, PMID: 37188781. (Reactome)
REST-dependent transcriptional repression	Reactome	Upper-Layer Glutamatergic Neurons (L2/L3-EXC)	REST	Lu et al.,2014
RHO GTPases Activate NADPH Oxidases	Reactome	Perivascular macrophage / border-associated macrophage	CD36	Park 2017;PMID 28515043.
Serotonergic synapse	KEGG; KEGG hsa04726	VIP interneuron; Subcortical_Neuron_Noradrenergic.LC	HTR3A; SLC6A2	Liu 2023,PMID: 36926728; Férézou 2002,PMID: 12196560. (genome.jp); PMID: 31531234
Signaling by CSF1 (M-CSF)in myeloid cells	Reactome	Inflammatory / cytokine-response microglia;Proliferating /cycling microglia	INPP5D; CSF1R	Tsai 2021,PMID 33631273; Chou 2023,PMID 38016942. (Reactome) ;Olmos-Alonso 2016, PMID 26747862;Mancuso 2019,PMID 31504240. (Reactome)
Signaling by ERBB4	Reactome	PVALB interneuron	ERBB4	Zhang 2017,PMID: 28684271;Deng 2024, PMID: 38982150. (Reactome)
Signaling by PDGF Signaling by Rho GTPases	Reactome Reactome	Pericytes (PERI) Oligodendrocyte Progenitor Cells (OPC)	PDGFRB LINGO1	Montagne et al.,2015 Wu et al.,2018
Synaptic Vesicle Cycle	KEGG hsa04721	Subcortical_Neuron_Cholinergic_BF	CHAT	PMID: 9832156
TGF-β signaling pathway	KEGG; Reactome	ECM /structural pericyte;Meningeal fibroblast;Homeostatic Microglia (HOM-MG)	SMAD3; TGFB1; TGFB1	İş 2024;Kearns 2023; PMID 37923721. (Nature); Butovsky et al.,2014

Continued on next page

Pathway Name	Source	Associated Cell Type(s)	Key Targets Involved	Citation
Tie2 Signaling	Reactome	Capillary endothelial cell	TEK	Flotho 2025;PMID 40909593;Sun 2023;PMID 37264161.
Transcriptional activation of mitochondrial biogenesis	Reactome	Parvalbumin-Positive GABAergic Interneurons (PV-INH)	PPARGC1A	Katsouri et al.,2011
Transmission across Chemical Synapses	Reactome	Corticothalamic /L6b glutamatergic neuron	NPTX2	Papanikolaou 2025 (PMID: 40345184);Xiao 2017 (PMID: 28440221)
Transport of glucose	Reactome	Capillary Endothelial Cells (CAP-ENDO)	SLC2A1	Winkler et al.,2015
Transthyretin-Mediated Amyloid Sequestration	Reactome	Choroid_Plexus_Epithelial	TTR	PMID: 17693656
Tyrosine Metabolism	KEGG	Subcortical_Neuron_Noradrenergic_LC	TH	PMID: 29632247
Vascular Smooth Muscle Contractility and Phenotype	KEGG	Vascular_Smooth_Muscle	ACTA2; MYOCD;SRF	PMID: 23083741
Vascular smooth muscle contraction	KEGG	Fibroblasts /Vascular Smooth Muscle Cells (FIB/vSMC)	ACTA2; MYOCD	Chow et al.,2007
Wnt signaling pathway	KEGG	Arterial endothelial cell	TNFRSF21	Huang 2024;PMID 38609388.

Table S8: AD LLM-driven mechanism hypotheses

Sequence	Upstream	Core Nodes	Downstream	Biological Rationale
BIN1 → MAX → [HES1-like transcriptional gap] → STAT3 → C3	BIN1	MAX,STAT3	C3	BIN1-linked trafficking stress engages astrocytic transcriptional reactivity and complement output.
BIN1 → MAX → [PEX6 trafficking gap] → ELK1 → EGR1 → C3	BIN1	MAX,ELK1, EGR1	C3	Endosomal-trafficking disruption propagates immediate-early transcription that culminates in complement induction.
C3 → EGR1 → [RAB36 vesicular gap] → MYC → [ARL8B lysosomal gap] → STAT3 → CHI3L1	C3	EGR1,MYC, STAT3	CHI3L1	Complement signaling feeds transcriptional amplification loops that stabilize reactive astroglial inflammatory states.
C3 → EGR1 → CTCFL → [ADRM1 proteostasis gap] → FKBP8 → CAT → [VWAS mitochondrial redox gap] → TP53 → CD44	C3	EGR1, CTCFL, FKBP8,CAT, TP53	CD44	Complement stress couples proteostasis-redox surveillance to adhesive astrocyte remodeling.
C3 → EGR1 → CTCFL → DYM → MAX → [FUCA2/RAB21 endosomal gap] → EPHA4 → MAPT	C3	EGR1, CTCFL,DYM, MAX,EPHA4	MAPT	Inflammatory complement cues traverse nuclear-endosomal bridges to excitatory receptor stress and tauopathy.
C3 → EGR1 → CTCFL → DYM → MAX → POLDIP3 → [TERF2 nuclear-metabolic gap] → PPARG → MAPT	C3	EGR1, CTCFL, DYM,MAX, POLDIP3, PPARG	MAPT	Complement-linked nuclear bridging reprograms neuronal metabolism toward tau-vulnerable states.
CD44 → TP53 → [VWAS mitochondrial gap] → CAT → FKBP8 → [ADRM1 proteostasis gap] → CTCFL → DYM → MYC → [ARL8B vesicular gap] → STAT3 → C3	CD44	TP53,CAT, FKBP8, CTCFL,DYM, MYC,STAT3	C3	Adhesion stress propagates redox-proteostasis-transcription coupling into sustained astrocytic complement activation.
BIN1 → MAX → DYM → CTCFL → [ADRM1 proteostasis gap] → FKBP8 → CAT → TP53 → CD44	BIN1	MAX,DYM, CTCFL, FKBP8,CAT, TP53	CD44	BIN1-to-CD44 bridging links glial trafficking defects to adhesive inflammatory phenotype switching.
EPHA4 → [RAB21/FUCA2 endosomal gap] → MAX → [HES1-like transcriptional gap] → STAT3 → C3	EPHA4	MAX,STAT3	C3	Synaptic receptor stress couples through endosomal-transcriptional bridges to astrocytic inflammatory output.
EPHA4 → [RAB21/FUCA2 endosomal gap] → MAX → POLDIP3 → [TERF2 nuclear-metabolic gap] → PPARG → MAPT	EPHA4	MAX, POLDIP3, PPARG	MAPT	EPHA4 signaling converges on metabolic nuclear control through endosomal and RNA-processing bridges.
PPARG → [TERF2 RNA-processing gap] → POLDIP3 → MAX → [HES1-like transcriptional gap] → STAT3 → C3	PPARG	POLDIP3, MAX,STAT3	C3	Metabolic transcriptional imbalance relays into STAT3-driven inflammatory glial programming.
CD44 → TP53 → [VWAS redox gap] → CAT → FKBP8 → [ADRM1 proteostasis gap] → CTCFL → DYM → MAX → POLDIP3 → [TERF2 nuclear-metabolic gap] → PPARG → MAPT	CD44	TP53,CAT, FKBP8, CTCFL, DYM,MAX, POLDIP3, PPARG	MAPT	CD44-associated stress reaches neuronal metabolic regulators through redox-proteostatic bridge architecture.
STAT3 →... → MYC →... → EGR1 → C3	STAT3	MYC,EGR1	C3	Inflammatory signaling engages MYC/EGR1 transcriptional hubs to drive astrocyte complement expression.
BIN1 → MAX → HES1 → STAT3	BIN1	MAX,HES1	STAT3	Oligodendrocyte risk gene BIN1 dysregulates MAX/HES1 axis,triggering chronic glial inflammation.
CD44 → TP53 →... → MYC → STAT3	CD44	TP53,MYC	STAT3	Astrocyte stress sensing (CD44) activates a TP53/MYC cascade sustaining neuroinflammation.
EPHA4 →... → MAX →... → POLDIP3 → PPARG	EPHA4	MAX, POLDIP3	PPARG	Neuronal receptor dysfunction alters metabolic resilience via a MAX-dependent transcriptional program.
BIN1 → MAX →... → ELK1 → EGR1 → C3	BIN1	MAX,ELK1, EGR1	C3	Oligodendrocyte stress signals via a transcriptional cascade to drive astrocyte reactivity.
C3 → EGR1 →... → MAX →... → EPHA4	C3	EGR1,MAX	EPHA4	Astrocyte reactivity initiates a transcriptional cascade altering neuronal receptor (EPHA4) signaling.

Continued on next page

Sequence	Upstream	Core Nodes	Downstream	Biological Rationale
CD44 → TP53 → ... → MAX → ... → PPARG	CD44	TP53,MAX	PPARG	Chronic astrocyte stress reprograms neuronal metabolism via a TP53/MAX-dependent signaling cascade.
C3 → EGR1 → ... → MAX → ... → PPARG	C3	EGR1,MAX	PPARG	Complement activation signals via EGR1/MAX to suppress neuronal pro-survival PPARG pathways.
BIN1 → MAX → HES1 → STAT3 → SERPINA3	BIN1	MAX,STAT3	SERPINA3	MAX couples endocytic adaptor loss to astrocyte reactive transcriptional output.
BIN1 → MAX → PEX6 → ELK1 → EGR1 → C3	BIN1	MAX,EGR1	C3	BIN1 perturbation activates ELK1-EGR1 axis driving complement C3 upregulation.
EPHA4 → RAB21 → FUCA2 → MAX → HES1 → STAT3	EPHA4	MAX,STAT3	(Astrocyte Reactivity)	EphA4 crosstalk with MAX-HES1 module sustains STAT3-mediated glial inflammation.
PPARG → TERF2 → POLDIP3 → MAX → HES1 → STAT3	PPARG	MAX,STAT3	(Inflammatory Secretome)	PPARG nuclear signaling diverted via MAX-HES1 to amplify STAT3 inflammatory program.
C3 → EGR1 → CTCFL → DYM → MAX → FUCA2 → RAB21 → EPHA4	C3	EGR1,MAX	EPHA4	Complement EGR1-CTCFL relay remodels synaptic EphA4 signaling via MAX hub.
C3 → EGR1 → RAB36 → MYC → ARL8B → STAT3	C3	EGR1,MYC	STAT3	C3-EGR1-MYC axis licenses STAT3 nuclear entry and reactive astrogliosis.
BIN1 → MAX → POLDIP3 → TERF2 → PPARG	BIN1	MAX,PPARG	(Metabolic Gene Program)	BIN1-MAX connection modulates PPARG via TERF2/POLDIP3,altering neuronal lipid metabolism.
CD44 → TP53 → VWA8 → CAT → FKBP8 → ADRM1 → CTCFL → DYM → MYC → ARL8B → STAT3	CD44	TP53,MYC, STAT3	(Reactive Astrocyte State)	CD44-TP53 stress pathway funnels through MYC to enforce STAT3-driven reactivity.
C3 → EGR1 → CTCFL → ADRM1 → FKBP8 → CAT → VWA8 → TP53 → CD44	C3	EGR1,TP53	CD44	Complement EGR1-CTCFL cascade stabilizes TP53-CD44 adhesion and inflammatory feed-forward.
EPHA4 → RAB21 → FUCA2 → MAX → POLDIP3 → TERF2 → PPARG	EPHA4	MAX	PPARG	EphA4-MAX hub redirects signal to PPARG-TERF2,disrupting neuronal nuclear receptor homeostasis.

Table S9: AD upgraded therapy combinations

Targets	Therapy Design	Intervention Rationale
ELK1,EGR1, C3	ELK1/EGR1-axis MEK inhibitor trametinib or selumetinib,plus C3 inhibitor pegcetacoplan.	Cuts transcriptional relay before complement amplification stabilizes inflammatory signaling.
EGR1,MYC, STAT3,C3	MEK inhibitor trametinib or selumetinib for the EGR1 arm, MYC inhibitor OMO-103, STAT3 inhibitor TTI-101,plus pegcetacoplan.	Serial blockade collapses the complement-to-transcription feed-forward inflammatory loop.
TP53,CAT, STAT3,C3	p53 reactivator eprenetapopt, catalase mimetic EUK-134, STAT3 inhibitor TTI-101,plus pegcetacoplan.	Redox-checkpoint and inflammatory-effector co-blockade should blunt astrocytic adhesive conversion.
EGR1, PPARG, MAPT	MEK inhibitor trametinib or selumetinib for the EGR1 arm, PPARG agonist pioglitazone,plus MAPT-lowering ASO BIIB080.	Pairs inflammatory transcription control with metabolic reset and tau suppression.
TP53,CAT, PPARG, MAPT	eprenetapopt,catalase mimetic EUK-134,pioglitazone,plus MAPT-lowering ASO BIIB080.	Intercepts redox-p53 bridge, restores metabolic tone,and lowers terminal tau burden.
STAT3,MYC, C3	JAK/STAT inhibitor (e.g., Tofacitinib),BET inhibitor (e.g., JQ1),Complement C3 inhibitor (e.g.,Pegcetacoplan)	Dismantle astrocyte inflammatory induction,transcription,and downstream complement-mediated damage.
CD44,TP53, PPARG	CD44 antagonist (e.g.,Monoclonal Antibody),MDM2 inhibitor (p53 stabilization),PPAR- γ agonist (e.g., Pioglitazone)	Block astrocyte stress sensing while restoring neuronal metabolic resilience and survival pathways.
EPHA4, MAX,PPARG	EPHA4 receptor tyrosine kinase inhibitor,BET inhibitor (indirectly targets MYC/MAX),PPAR- γ agonist	Normalize pathogenic neuronal receptor signaling, transcriptional state,and metabolic function.
BIN1,MAX, STAT3	(No direct BIN1 drug),BET inhibitor,JAK/STAT inhibitor	Intervene at the central transcriptional bridge to decouple oligodendrocyte risk from glial inflammation.
STAT3,MYC	STAT3 inhibitor (e.g.,Napabucasin) + MYC inhibitor (e.g.,Omomyc mini-protein /BET bromodomain inhibitor JQ1)	Dual blockade collapses the reactive astrocyte transcriptional core driven by MAX-EGR1 relay.
C3,EGR1	Complement C3 inhibitor (e.g., Pegcetacoplan)+ EGR1 transcriptional repressor (e.g.,Mithramycin A)	Suppress complement-initiated feed-forward loop and downstream EGR1-mediated glial activation.
PPARG, STAT3	PPAR γ agonist (e.g.,Pioglitazone) + STAT3 inhibitor (e.g.,Stattic)	Restore neuronal metabolic homeostasis while preventing MAX-HES1 diversion to inflammatory STAT3.

Continued on next page

Targets	Therapy Design	Intervention Rationale
CD44,TP53	CD44 antagonist (e.g.,Anti-CD44 mAb /HA oligosaccharides)+ p53 modulator (e.g.,APR-246)	Disrupt CD44-TP53 stress synapse while normalizing MYC-STAT3 reactive astrocyte transition.
EPHA4, STAT3	EphA4 inhibitor (e.g.,Peptide KYL)+ STAT3 inhibitor (e.g., Napabucasin)	Block synaptic maladaptation signal and downstream glial inflammation via MAX-HES1 hub.

Table S10: AD target-drug evidence ledger (deduplicated)

Target	Candidate Drug	Drug Class	Evidence Citation (PMID/DOI/Author-Year)	Direction	Status
ADRM1	No validated drug	None validated	No direct evidence found	mixed	unknown
ARL8B	No validated drug	None validated	No direct evidence found	mixed	unknown
BIN1	No validated drug	None validated	No direct evidence found	mixed	unknown
C3	Pegcetacoplan	Complement C3 inhibitor	Hillmen 2021;PMID 33730455	support	approved
CAT	EUK-134	SOD/catalase mimetic	Sánchez-Carbente 2005;PMID 15637642;DOI 10.1038/sj.cdd.4401560	support	preclinical
CAT	No validated drug	None validated	No direct evidence found	mixed	unknown
CD44	Anti-CD44 mAb	CD44 antagonist	PMID:29545084	support	investigational
CD44	Anti-CD44 mAb (e.g.,RG7356)	CD44 antagonist	PMID:27098676 (Matsumoto et al., 2016)	mixed	investigational
CD44	Bivatuzumab mertansine	Anti-CD44v6 antibody-drug conjugate	Tijink 2006;PMID 17062682	contradict	unknown
CHI3L1	Anti-CHI3L1 monoclonal antibody	Monoclonal antibody	Ham 2026;PMID 41573517	support	preclinical
CTCF	No validated drug	None validated	No direct evidence found	mixed	unknown
DYM	No validated drug	None validated	No direct evidence found	mixed	unknown
EGR1	Mithramycin A	Sp1/EGR1 transcriptional inhibitor	PMID:15126370 (Blume et al.,2004)	mixed	preclinical
EGR1	Trametinib	MEK inhibitor (indirect EGR1-pathway suppressor)	Cesana 2023;PMID 36888606	support	approved
ELK1	No validated drug	None validated	No direct evidence found	mixed	unknown
ELK1	Trametinib	MEK inhibitor (indirect ELK1-pathway suppressor)	Elbadawy 2021;PMID 34034619	support	approved
EPHA4	Dasatinib	Receptor Tyrosine Kinase inhibitor	PMID:24681604	support	approved
EPHA4	KYL peptide	EphA4 inhibitory peptide	PMID:23426964 (Vargas et al.,2013)	support	preclinical
EPHA4	Rhynchophylline	Small-molecule EphA4 inhibitor	Fu 2021;PMID 34738783	support	preclinical
FKBP8	No validated drug	None validated	No direct evidence found	mixed	unknown
FUCA2	No validated drug	None validated	No direct evidence found	mixed	unknown
HES1	No validated drug	None validated	No direct evidence found	mixed	unknown
MAPT	BIIB080 (MAPTRx)	Tau antisense oligonucleotide	Mummery 2023;PMID 37095250	support	investigational
MAX	10058-F4	MYC-MAX dimerization inhibitor	PMID:25442568	support	preclinical
MAX	No validated drug	None validated	No direct evidence found	mixed	unknown
MAX	OMO-103	MYC/MAX interaction inhibitor	Garralda 2024;PMID 38321218	mixed	investigational
MYC	JQ1	BET bromodomain inhibitor	PMID:12917343 (Ferrer et al.,2003)	mixed	investigational
MYC	OMO-103	MYC inhibitor	Garralda 2024;PMID 38321218	support	investigational
MYC	Omomyc	MYC dominant-negative mini-protein	PMID:12917343 (Ferrer et al.,2003)	mixed	investigational
PEX6	No validated drug	None validated	No direct evidence found	mixed	unknown
POLDIP3	No validated drug	None validated	No direct evidence found	mixed	unknown
PPARG	Pioglitazone	PPAR γ agonist	Sato 2011;PMID 19923038;Burns 2021;PMID 34146512	mixed	approved
RAB21	No validated drug	None validated	No direct evidence found	mixed	unknown
RAB36	No validated drug	None validated	No direct evidence found	mixed	unknown
SERPINA3	No validated drug	None validated	No direct evidence found	mixed	unknown
STAT3	Napabucasin	STAT3 inhibitor	PMID:28099414 (Liddelow et al., 2017)	mixed	investigational
STAT3	TTI-101	STAT3 inhibitor	Tsimberidou 2025;PMID 39792482	support	investigational
STAT3	Tofacitinib	JAK/STAT inhibitor	PMID:30654817	support	approved
TERF2	No validated drug	None validated	No direct evidence found	mixed	unknown
TP53	APR-246	Mutant p53 reactivator	PMID:12588893 (de la Monte et al., 2003)	mixed	investigational
TP53	Eprentapopt (APR-246)	p53 reactivator	Cluzeau 2021;PMID 33600210;Santini 2024;PMID 38768424	mixed	investigational
TP53	Nutlin-3a	MDM2 inhibitor	PMID:20600295	contradict	preclinical
VWA8	No validated drug	None validated	No direct evidence found	mixed	unknown