UCI University of California, Irvine

AI-Driven Therapeutic Target Discovery and Cell-State Reprogramming

SCORCH Fall Meeting 2026

Donald Bren School of
Information & Computer Sciences

Department of Computer Science

Ziheng Duan, Ph.D.
Jing Zhang Lab

September 15th, 2026

The data-to-intervention gap

Substance use burden

1 in 10 people had a drug use disorder.

Mental health burden

1 in 4 adults experienced a mental illness.

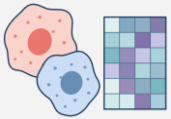
U.S., 2024

Our goal

Discover interventions to treat and cure disease

Biological data provide a starting point

Gene evidence



Cell profiles

Perturbation data

How do we turn data into therapeutic hypotheses?

Which genes should we target?

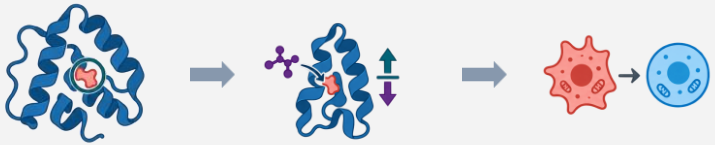
Which genes should we co-perturb?

disease health

Rounded estimates. 2024 U.S. NSDUH. Drug use disorder: 9.8%, ages 12+. Any mental illness: 23.4%, ages 18+.

Target identification: challenges and our approach

What are we looking for?



Identify a target → Modulate activity → Therapeutic benefit

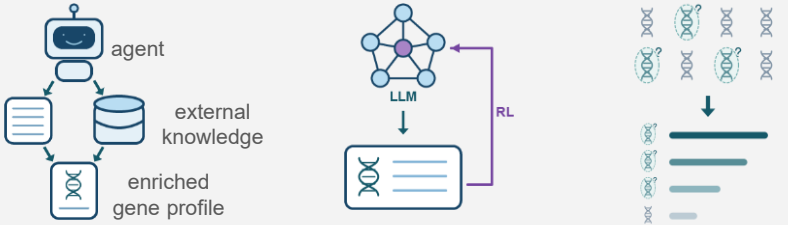
Three challenges

Static features | Limited interpretability | Incomplete labels

How do existing methods rank candidates?

DrugnomeAI: 324 gene-level features from 15 data sources

How does TargetSage address them?



Agentic Profiling | Guided Reasoning | PU Learning

DrugnomeAI: Raies et al., Commun. Biol. 5, 1291 (2022). doi:10.1038/s42003-022-04245-4

TargetSage: Duan et al., NeurIPS 2026 (under review).

TargetSage: integrating evidence to prioritize targets

Biological evidence

19,032 human protein-coding genes

- **Function & networks (353)**
- Chemical & antibody links (71)
- Genomic constraint (44)
- Phenotype & disease (8)
- Expression & sequence (6)

+ Retrieved biomedical evidence
(literature & databases)

Training examples

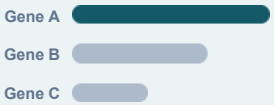
700 known drug targets

18,332 unlabeled genes

PHAROS: a database of drug targets

TargetSage

Target prioritization



Gene-specific rationales

TargetSage: Duan et al., NeurIPS 2026 (under review).

TargetSage leads across 15 benchmark tasks

Adjusted F1 = $R_{soft}^2 / \bar{p}$

R_{soft}: mean score on known targets

$\bar{p}$: mean score across evaluated genes

○ Best baseline per task

◆ TargetSage

Pharos labels

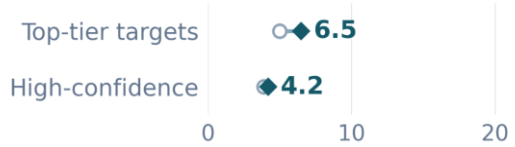
2 tasks

Targets of drugs and active compounds

Druggability tiers

2 tasks

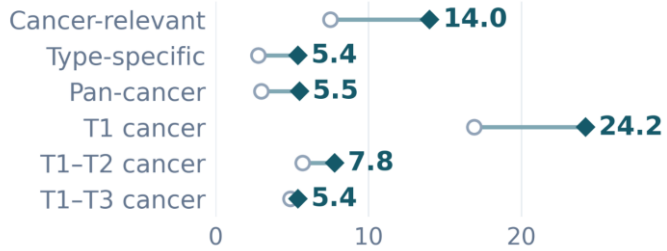
Different levels of supporting evidence



Cancer druggability

6 tasks

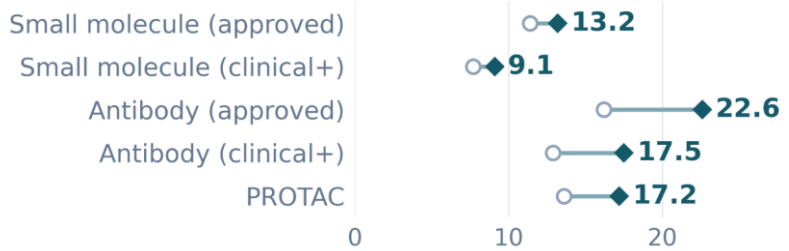
Potential targets for cancer treatment



Drug modality

5 tasks

Small molecules, antibodies and protein degraders



PU learning: Bekker & Davis, Mach. Learn. 109, 719–760 (2020). doi:10.1007/s10994-020-05877-5

TargetSage: Duan et al., NeurIPS 2026 (under review).

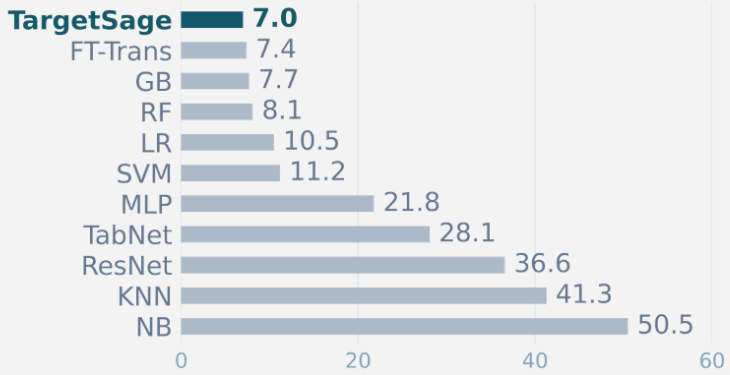
TargetSage predicts future validated targets

Temporal label holdout and independent clinical-target enrichment

Temporal label holdout

2021 Training labels → **2025** 90 newly clinical targets

Median rank percentile (%) ↓



Lowest median rank percentile

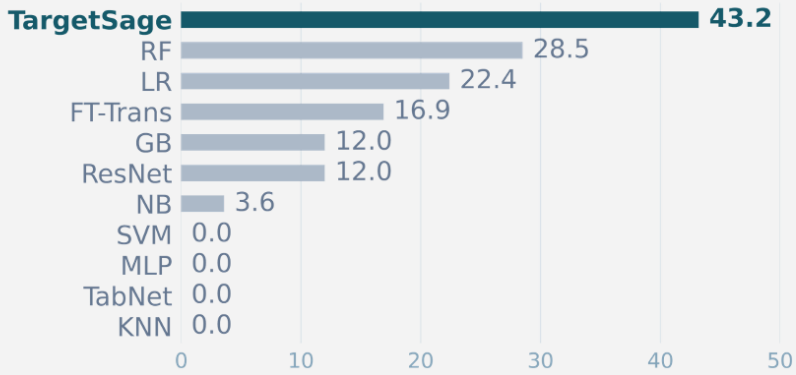
Independent clinical targets

PHAROS 2025 → ChEMBL Phase II/III

136 clinical targets among 11,891 Tbio genes

proteins targeted by drugs in phase II or III clinical trials.

Top-25 enrichment odds ratio ↑



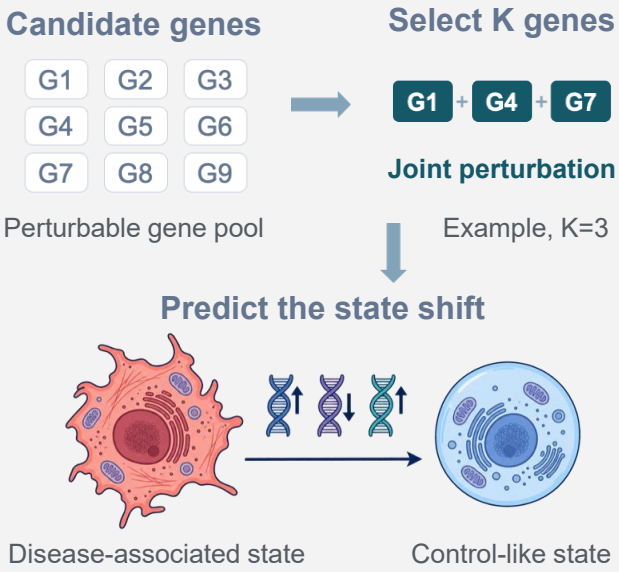
+51% vs. RF (43.2 vs. 28.5)

TargetSage: Duan et al., NeurIPS 2026 (under review).

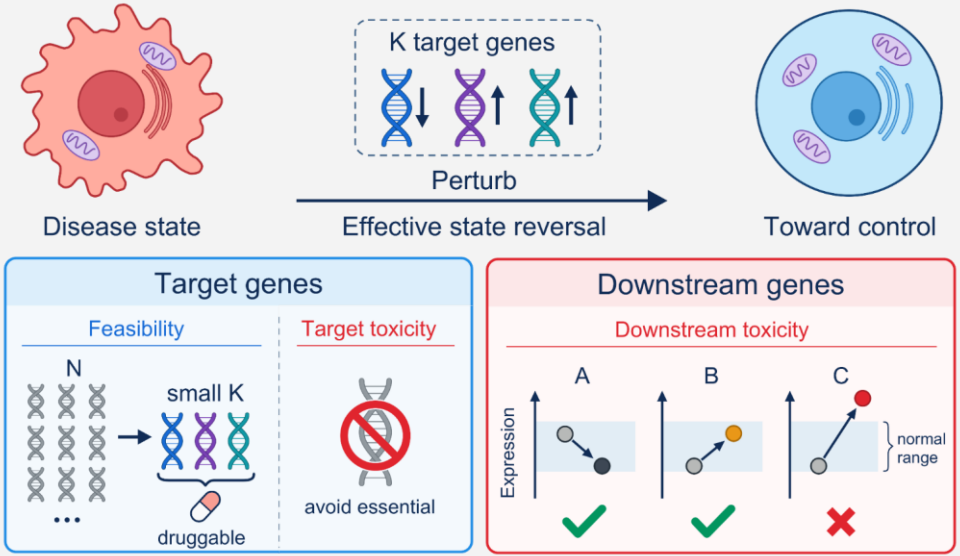
What is cell-state reprogramming?

Directing cells toward a desired identity or functional state.

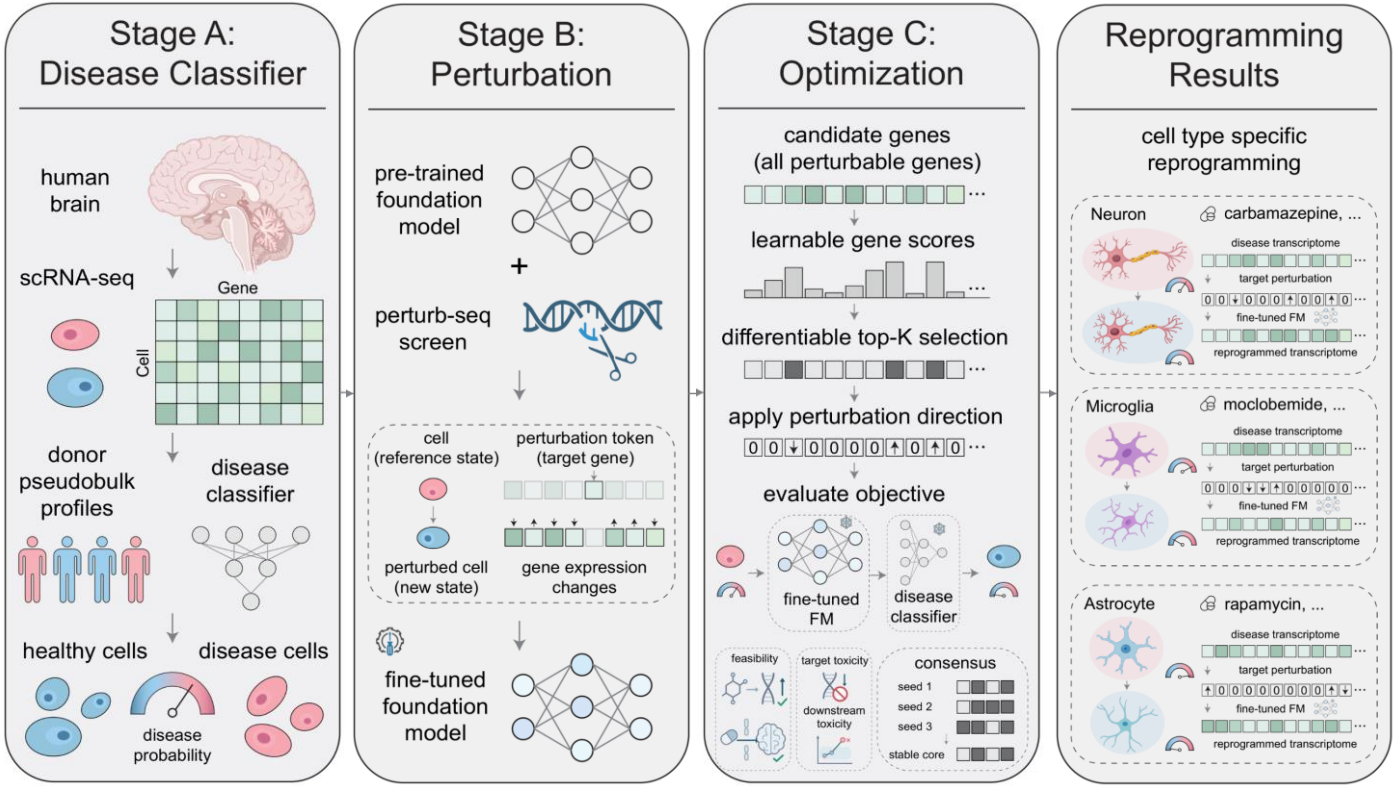
Which K-gene combination best shifts cells toward a desired state?



Therapeutic design: effectiveness, feasibility, and toxicity



CellReprogrammer optimizes gene combinations



therapeutic design implementation

- **Effectiveness**
  - Maximize disease reversal
- **Feasibility**
  - Select K genes
  - Preserve cell identity
- **Toxicity**
  - Essentiality penalty
  - Limit downstream changes

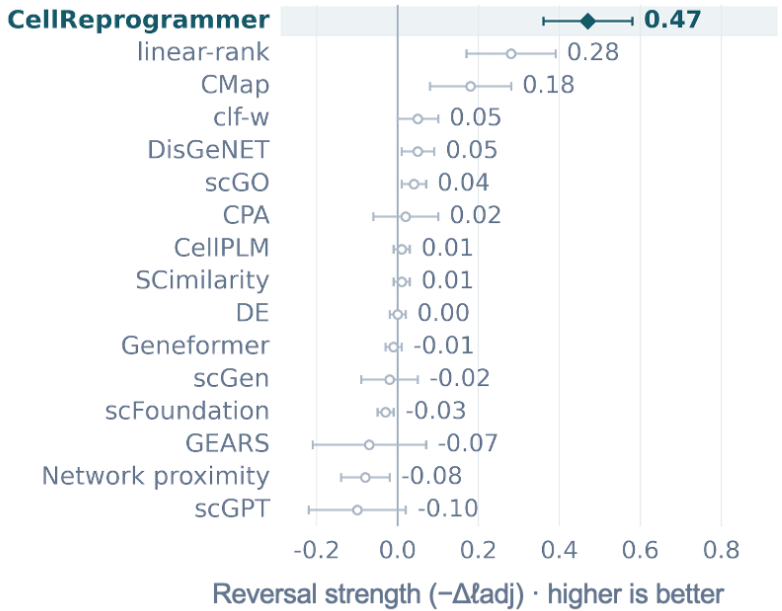
CellReprogrammer: Duan et al., ICLR 2027 (under review).

Most effective disease-state reprogramming

K = 3 genes · matched perturbation budget · 5 disorders × 3 cell types

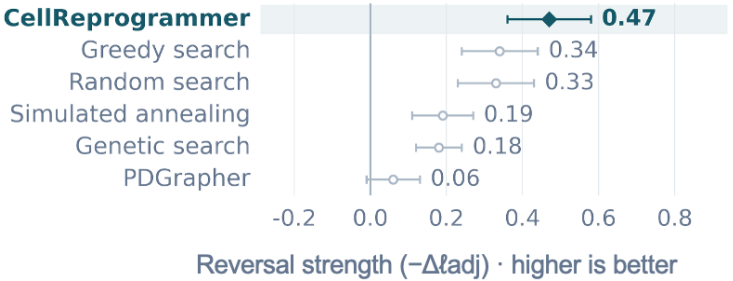
Individual gene ranking

15 baselines



Joint gene selection

5 baselines



0.47 ± 0.11
with feasibility and toxicity penalties

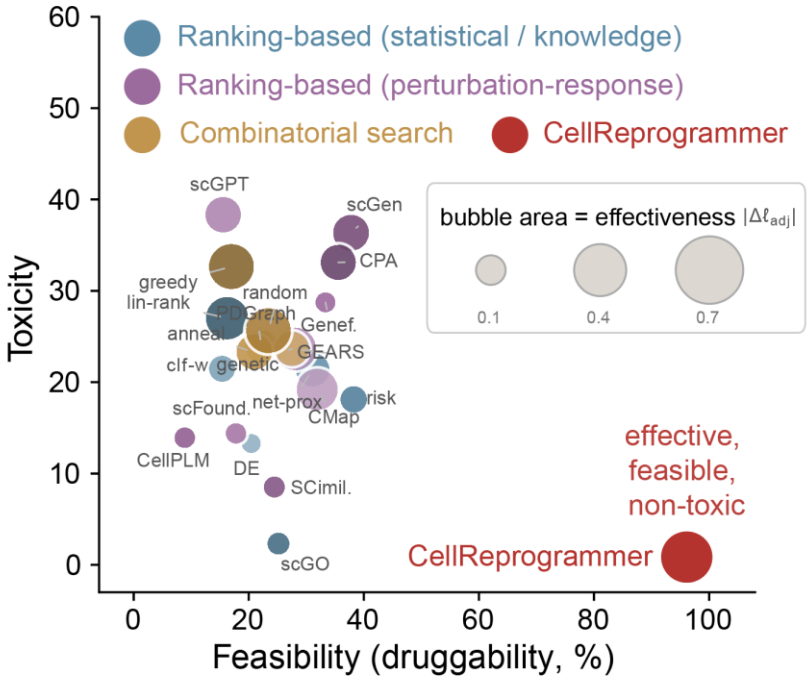
15 / 15
shifted toward control

Mean ± SE · zero = matched random control

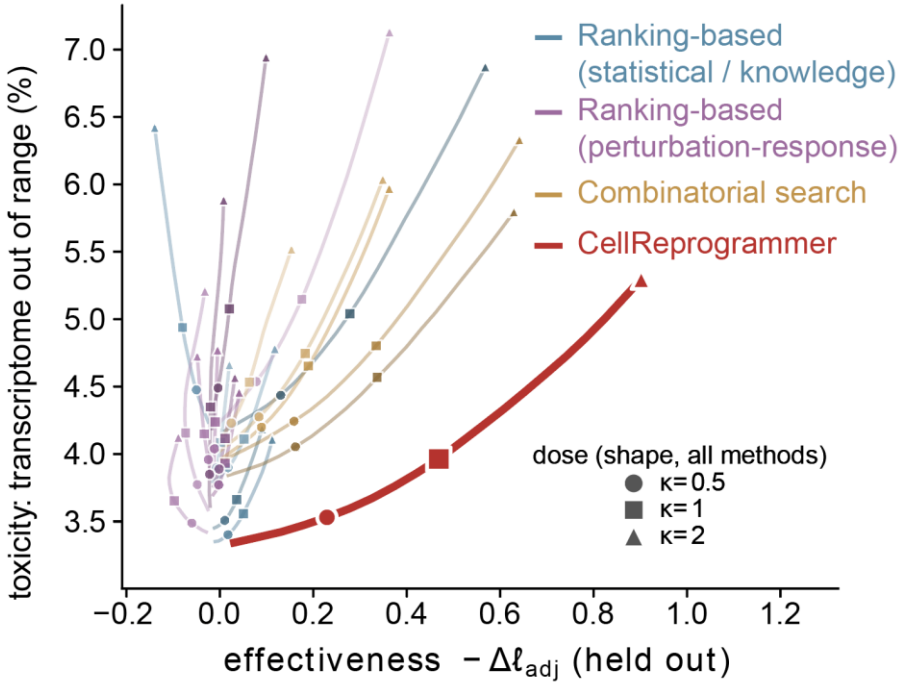
CellReprogrammer: Duan et al., ICLR 2027 (under review).

Co-optimizing effectiveness, feasibility, and toxicity

Effective, feasible and low predicted toxic.



Effect grows with dose at lowest off-manifold cost.



CellReprogrammer: Duan et al., ICLR 2027 (under review).

From biological data to testable hypotheses

Two complementary advances in therapeutic discovery

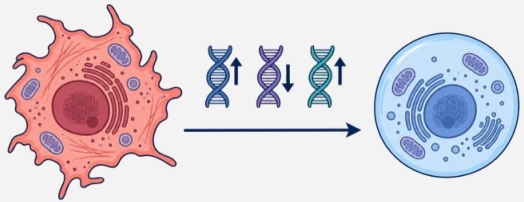
TargetSage

Prioritize therapeutic targets with supporting evidence

Agentic evidence integration
Reward-guided biological reasoning
PU-aware target prioritization

CellReprogrammer

Design gene combinations for a desired cell state



Effectiveness, feasibility, and toxicity at once

AI helps us make scientific discoveries.

Next: validate targets and gene combinations in relevant cellular models.

TargetSage: Duan et al., NeurIPS 2026 (under review).

CellReprogrammer: Duan et al., ICLR 2027 (under review).

Acknowledgments

Special thanks

Hyejung Won
UNC Chapel Hill

Mark Gerstein
Yale University

Jing Zhang & the Zhang Lab

Jing Zhang
UC Irvine

SCORCH

NIH National Institute on Drug Abuse

Thanks for listening!

Detailed feature sources

482 structured features across five evidence categories

Evidence	Source	Dim.	Biological meaning
Function & networks 353 features	CTDbase	239	Pathway / function annotations and counts
	InterPro	97	Protein families and domains
	STRING	10	Network connectivity to known targets
	InWeb	4	Network connectivity to known targets
	Reactome	2	Network connectivity to known targets
	PHAROS	1	Protein interaction count (ppiCount)
Chemical & antibody 71 features	CTDbase	68	Chemical interaction annotations and counts
	PHAROS	2	Antibody and monoclonal counts
	DGIdb	1	Gene–drug interaction types
Genomic constraint 44 features	ExAC	19	Constraint / CNV (18), gene size (1)
	gnomAD	15	Mutation constraint metrics
	Genic Intol.	6	RVIS, MTR, LoF_FDR and geneCov
	Mouse Genes	4	Essentiality, overexpression and GC%
Phenotype & disease 8 features	GWAS	6	GWAS hits, P values, odds ratios, trait flag
	MGI	1	Mouse–human phenotypes
	OMIM	1	Disease association count
Expression & sequence 6 features	PHAROS	4	HPM/HPA tissue specificity
	UniProt	1	Sequence length
	GTEx	1	Tissue expression profile

Six-tool retrieval pipeline

1 Read and verbalize structured features

Example: pLI = 1.00 → strong LoF intolerance

2–5 Retrieve complementary evidence

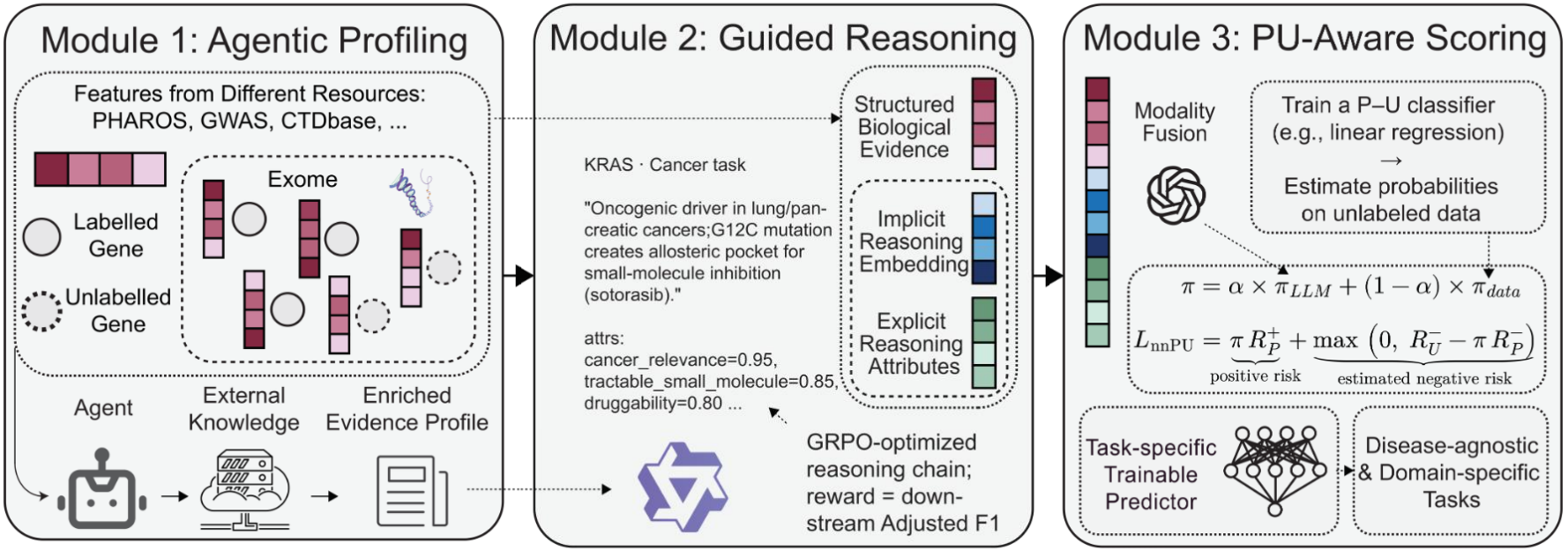
NCBI Gene	Gene function and disease context
UniProt	Function, location and tissue specificity
PubMed	Mechanistic and therapeutic literature
Open Targets	Tractability and safety liabilities

6 Check evidence before reasoning

Remove direct label disclosures and flag conflicting statements.

TargetSage turns evidence into target scores

INPUT: Gene evidence + known targets → OUTPUT: Target scores + rationales



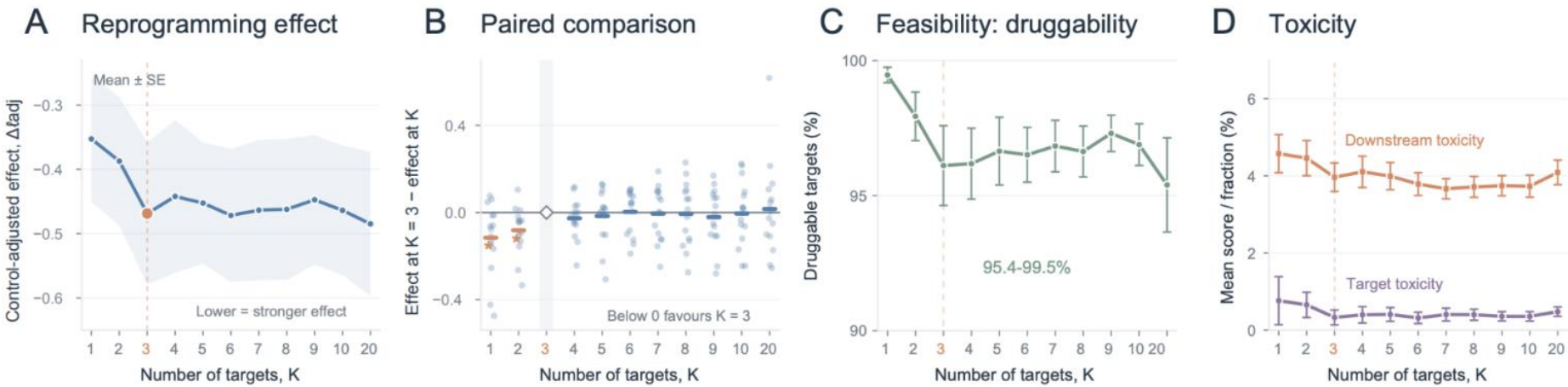
482 features · 15 databases
6-tool evidence-retrieval agent

23 therapeutic attributes
GRPO-guided gene-specific reasoning

Gene-adaptive multimodal fusion
Hybrid prior + non-negative PU risk

Three targets balance reprogramming effect and set size

K = 3 is a compact operating point under the shared evaluation protocol.



Stronger than K = 1 or 2

Limited gains beyond K = 3

Feasibility and toxicity remain stable