Gene Target Intelligence: A Multi-Agent, Multi-Omics Knowledge Graph from Target Identity to Therapeutic Mechanism

Swaraj Basu and Radhakrishna (RK) Bettadapura
Strand Life Sciences, Bangalore, India.

Visit our website

Takeaway

A fault-tolerant multi-agent workflow links independently sourced target evidence into ranked, provenance-aware therapeutic chains **without allowing language-model synthesis to create evidence.**

1 Need and Aim

Target assessment requires simultaneous review of disease genetics, drug development, pathway membership, tissue and cell-type expression, and perturbational response. These data use different identifiers, evidence scales, and access patterns. Manual integration is slow and can obscure provenance. We developed a modular agent system that resolves a target once, retrieves complementary evidence in parallel, and preserves source-level traceability through a common knowledge-graph schema.

Objective

- **Resolve** symbols and identifiers reproducibly.
- **Integrate** clinical, molecular, expression, and perturbation evidence.
- **Explain** gene → pathway → disease → drug connections with source-aware plots and tables.

Design Principles

- Parallel, failure-tolerant agents
- Local-first analytical files
- Explicit provenance and bounded graphs
- Grounded narrative synthesis

Input → Output

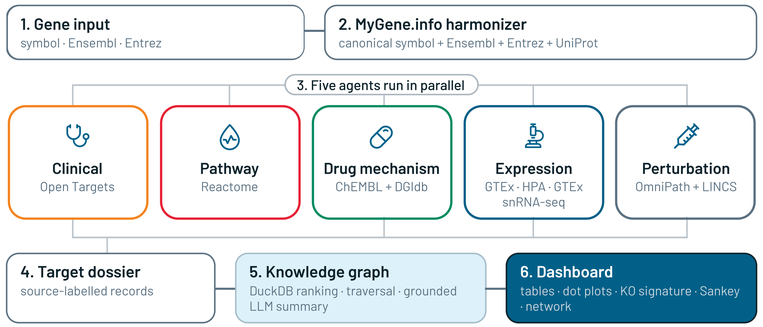
EGFR → Target Profile · Clinical Evidence · Expression Atlas · Biology · Mechanistic Insight · How It Works

One harmonized input; six coordinated evidence views.

GENOME-WIDE GRAPH SCALE

2.03M	66.5k	18,484
edges	nodes	genes with ≥1 disease edge

2 Multi-Agent Method

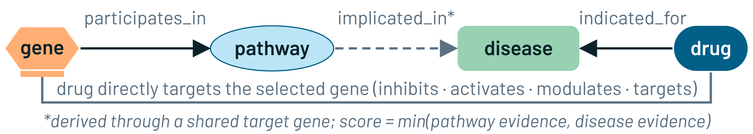


Source-to-Evidence Matrix

SOURCE	EVIDENCE RETAINED	DASHBOARD OUTPUT
MyGene.info	Canonical IDs	Target identity
Open Targets 26.03	Disease score, stage, indication, MoA	Disease / drug landscape
Reactome	Curated memberships	Pathway table and graph
ChEMBL + DGIdb	Drug action and interaction	Mechanism table
GTEx bulk RNA-seq†	Median tissue TPM	RNA dot plot
HPA IHC + MS	Protein level, reliability, intensity	Protein plots
GTEx snRNA-seq	Cell type × tissue mean expression	Single-cell plot
OmniPath	Directed activation / inhibition	Causal neighbours
LINCS L1000 + Enrichr	KO response genes, enriched terms	Perturbation plots

CellxGene is an optional single-cell fallback; PanglaoDB is prepared in setup but not used. †Local GTEx file (74,628 genes, 68 tissues) post-dates v8 although the setup URL points to v8 — confirm release label before submission.

Integration and Scoring



Eligible chains require a direct drug–target edge and a matching drug indication. Chains are ranked by the log-sum of pathway, derived pathway–disease, drug–disease, and phase-normalized drug evidence. Scores closer to zero indicate stronger combined evidence — in plain terms, one weak link pulls the whole chain down. A diversity cap prevents one disease from dominating.

Inference is labelled — primary traversals and fallback cross-joins are retained as distinct evidence types.

3 Integrated Results · EGFR Case Study

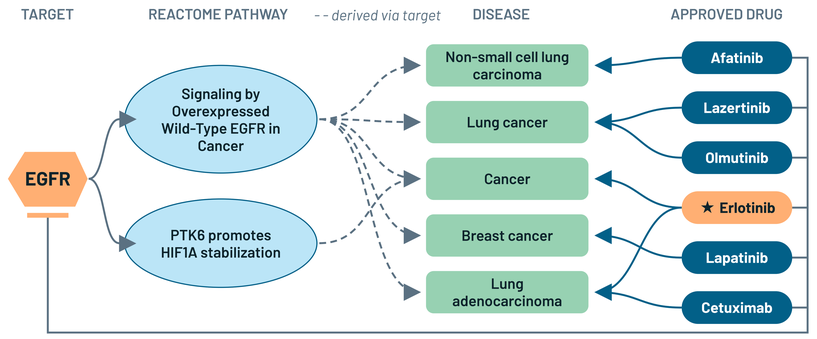


Figure 1. Provenance-aware pathway-to-therapy network for EGFR. Reactome memberships connect the target to pathways; pathway-disease relations are derived through shared target evidence (dashed); Open Targets supplies disease associations, direct target mechanisms, and clinical indications. ★ = drug retained for ≥2 diseases. Top-ranked traversal chains only (K = 5); salt forms merged; pathway ties broken by pathway size.

CLINICAL
Non-small cell lung carcinoma scores highest (Open Targets 0.85). 87 EGFR drug records span preclinical to approved (28 approved).

MECHANISTIC
Overexpressed WT EGFR signalling carries 7 of 8 chains; strongest: → NSCLC → afatinib (−0.16).

TRANSLATIONAL
★ **Erlotinib** reaches 2 retained diseases — a known indication pattern serving as a positive control.

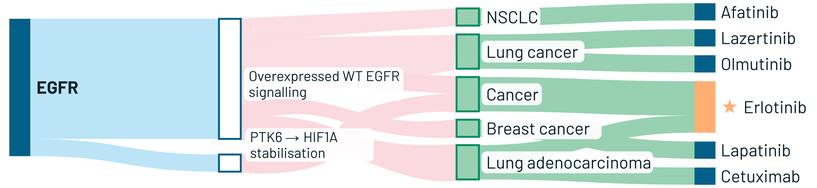


Figure 2. Evidence flow. Link width is the number of retained chains sharing a node pair, not evidence magnitude. Hatched links = derived. The displayed disease → drug direction reverses the canonical drug → disease indication edge to preserve left-to-right biological flow.

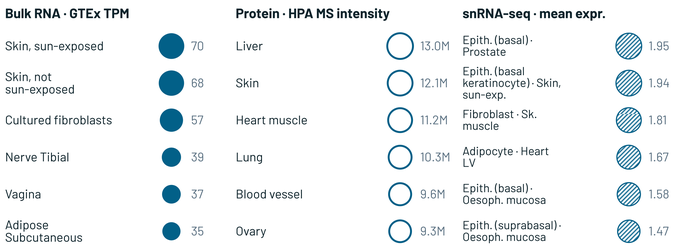


Figure 3. Tissue and cell context. Circle area ∝ value; fill = assay (solid RNA, open protein, hatched snRNA-seq). Original values plotted; graph edges use within-gene normalized scores. HPA IHC: High (Enhanced) in bronchial/nasopharyngeal basal cells and placental trophoblasts.



Enrichr top terms (adj. p)

▲ up: Apoptosis 0.004; Interferon Alpha Response 0.022; AUF1 (hnRNP D0) Binds And Destabilizes mRNA 0.038

▼ down: Aflatoxin Activation And Detoxification 0.017; Metabolism of xenobiotics by cytochrome P450 0.037; TGF-beta signaling pathway 0.037

Figure 4. EGFR loss (LINCS L1000). Top 6 genes per direction; black = up, white = down. CRISPR KO preferred, shRNA if unavailable. Overlaps of 2–3 genes; Y-linked RPS4Y1 likely reflects cell-line sex mix.

Key Findings

- Parallel agents unify heterogeneous evidence into a single target dossier while allowing individual sources to fail gracefully.
- A canonical graph makes each displayed pathway-to-therapy chain traceable to explicit source or derived edges.
- Expression and perturbation panels add tissue context and functional consequences to the clinical network.

Limitation

- Evidence scores from different sources are ranking aids, not calibrated probabilities.
- Pathway → disease is a shared-gene inference, not a direct curated assertion.
- API-backed content changes over time; data accessed 2026-10-06.
- Local GTEx file (74,628 genes) post-dates the v8 setup URL — confirm release label.
- All Reactome memberships score 1.0; live pathway choice is tie-order dependent.

Reproducibility and Provenance

Local Parquet → DuckDB SQL
NetworkX/PyVis → Plotly/Streamlit
Open Targets bulk release: 26.03 · DGIdb export: Jan 2024
Code/data access date: 2026-10-06
Case-study snapshot: python/01_case_study_snapshot.py
Repository / interactive dashboard: [QR code, top right]

Acknowledgments and References

We acknowledge MyGene.info, Open Targets, Reactome, ChEMBL, DGIdb, GTEx, the Human Protein Atlas, OmniPath, LINCS/SigCom and Enrichr. mygene.info · platform.opentargets.org · reactome.org · ebi.ac.uk/chembl · dgidb.org · gtexportal.org · proteinatlas.org · omnipathdb.org · maayanlab.cloud/sigcom-lincs · maayanlab.cloud/Enrichr. [Primary citations to add before submission.]