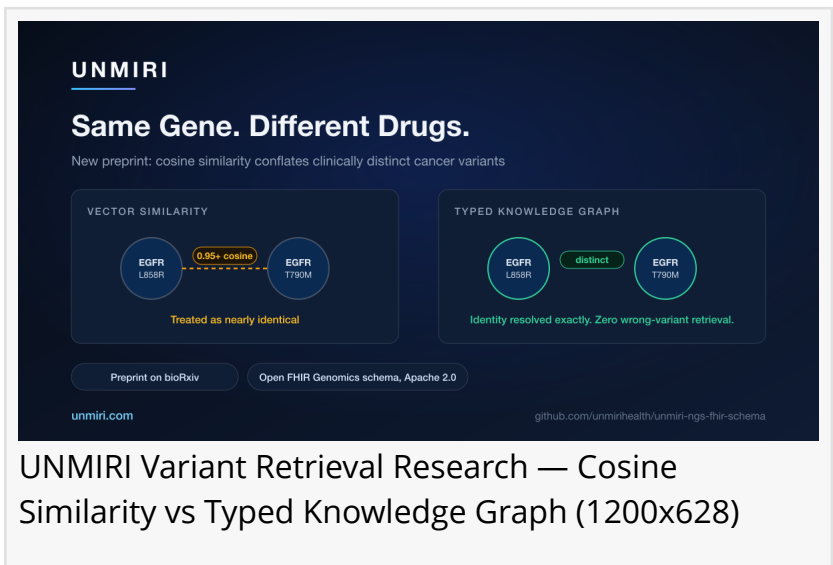


UNMIRI Publishes Open FHIR Genomics Schema and Variant Retrieval Research

A new preprint shows vector-similarity search can treat clinically distinct cancer variants as the same, and a typed knowledge graph keeps them apart

LANGHORNE, PA, UNITED STATES, June 16, 2026 /EINPresswire.com/ -- UNMIRI LLC has published an [open-source FHIR Genomics R4 schema](#) for cross-vendor next-generation sequencing data, along with a [research preprint](#) documenting a failure mode in how many clinical AI tools retrieve genomic evidence. Both are available now to developers and researchers.



UNMIRI
Same Gene. Different Drugs.
 New preprint: cosine similarity conflates clinically distinct cancer variants

VECTOR SIMILARITY

EGFR L858R — 0.95+ cosine — EGFR T790M

Treated as nearly identical

TYPED KNOWLEDGE GRAPH

EGFR L858R — distinct — EGFR T790M

Identity resolved exactly. Zero wrong-variant retrieval.

Preprint on bioRxiv Open FHIR Genomics schema, Apache 2.0

unmiri.com github.com/unmirihealth/unmiri-ngs-fhir-schema

UNMIRI Variant Retrieval Research — Cosine Similarity vs Typed Knowledge Graph (1200x628)

The preprint, "Cosine Similarity Conflates Clinically Distinct Cancer Variants: A Case for Typed-Graph Retrieval in Precision Oncology Decision Support" (bioRxiv, DOI

10.64898/2026.05.05.723102), looks at a method used widely in retrieval-augmented AI: comparing text by the cosine similarity of their embeddings. On a curated benchmark of clinically distinct variant pairs, every pair was treated as nearly identical at a 0.95 similarity threshold under two biomedical encoders. Put simply, the math judged genuinely different variants to be the same. Because two variants in the same gene can call for different drugs, a system that blurs them can surface confident, wrong evidence.

“

Cosine similarity is fine for finding similar paragraphs. It is the wrong tool for deciding whether two genomic alterations are the same thing. Get identity wrong and nothing downstream can save you.”

Umair Khan, co-founder and CTO, UNMIRI

UNMIRI's answer, described in the paper and built into its platform, is to stop treating variant identity as a similarity

problem at all. The platform resolves identity through a typed knowledge graph, matching variants on standardized identifiers exactly, before any retrieval or ranking step. On the same benchmark, that approach brought wrong-variant retrieval to zero.

The open schema puts the same discipline in other developers' hands. It maps the reports that labs such as Foundation Medicine, Tempus, Caris, Guardant, and Natera produce into one structured FHIR R4 Genomics representation, with HGVS variant conventions, biomarker fields, and identifier-based citation patterns. It is published on GitHub under the Apache 2.0 license and archived for citation on Zenodo (DOI 10.5281/zenodo.20042352). UNMIRI does not present the schema as proprietary invention. It is a working implementation of the public FHIR Genomics specification that any team can adopt, fork, or contribute to.

"We published the failure mode because the field is quietly leaning on a method that does not hold up for variants," said Umair Khan, co-founder and CTO of UNMIRI. "Cosine similarity is fine for finding similar paragraphs. It is the wrong tool for deciding whether two genomic alterations are the same thing. Get identity wrong and nothing downstream can save you."

The schema and the preprint sit underneath UNMIRI's broader platform, a set of APIs for cross-vendor NGS interpretation, genomics-aware clinical decision support, variant-grounded trial matching, and prior-authorization decisions. All of those surfaces depend on the same identity-resolution layer the paper examines. UNMIRI is pre-revenue and onboarding design partners through synthetic-data sandboxes.

The preprint is available at <https://doi.org/10.64898/2026.05.05.723102> and the schema at <https://github.com/unmirihealth/unmiri-ngs-fhir-schema>.

About UNMIRI

UNMIRI LLC builds precision oncology infrastructure that healthcare software companies, biotech medical affairs teams, and clinicians can verify. Its cross-vendor APIs normalize next-generation sequencing reports into citation-grounded clinical data using a knowledge-graph architecture and deterministic output. Founded in 2023 and based in Langhorne, Pennsylvania, UNMIRI is led by CEO Nida Uddin and CTO Umair Khan. Learn more at <https://unmiri.com>.

Umair Khan

UNMIRI LLC

press@unmiri.com

Visit us on social media:

[LinkedIn](#)

[Instagram](#)

[Facebook](#)

[X](#)

This press release can be viewed online at: <https://www.einpresswire.com/article/918988763>

EIN Presswire's priority is source transparency. We do not allow opaque clients, and our editors try to be careful about weeding out false and misleading content. As a user, if you see something

we have missed, please do bring it to our attention. Your help is welcome. EIN Presswire, Everyone's Internet News Presswire™, tries to define some of the boundaries that are reasonable in today's world. Please see our Editorial Guidelines for more information.

© 1995-2026 Newsmatics Inc. All Right Reserved.