

UCSF

UC San Francisco Previously Published Works

Title

Announcing the Biomedical Data Translator: Initial Public Release

Permalink

<https://escholarship.org/uc/item/9c43g8c7>

Journal

Clinical and Translational Science, 18(7)

ISSN

1752-8054

Authors

Fecho, Karamarie
Glusman, Gwênlyn
Baranzini, Sergio E
[et al.](#)

Publication Date

2025-07-01

DOI

10.1111/cts.70284

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Peer reviewed

MINI REVIEW **OPEN ACCESS**

Announcing the Biomedical Data Translator: Initial Public Release

Karamarie Fecho^{1,2}  | Gwênlyn Glusman³  | Sergio E. Baranzini⁴  | Chris Bizon¹  | Matthew Brush⁵  | William Byrd⁶  | Lawrence Chung^{7,8}  | Andrew Crouse⁶  | Eric Deutsch³  | Michel Dumontier⁹  | Aleksandra Foksinska⁶  | Jennifer Hadlock³  | Kaiwen He⁶  | Sui Huang³  | Robert Hubal¹  | Gregory M. Hyde¹⁰  | Sharat Israni⁴  | Kelyne Kenmogne¹  | David Koslicki¹¹  | Jana Dorfman Marcette¹²  | Ewy A. Mathe¹³  | Abrar Mesbah¹⁴  | Sierra A. T. Moxon¹⁵  | Christopher J. Mungall¹⁵  | John Osborne⁶  | Carrie Pasfield¹  | Guangrong Qin³  | Stephen A. Ramsey¹⁶  | Justin Reese¹⁵  | Jared C. Roach³  | Reese Rose¹⁷  | Karthik Soman⁴  | Andrew I. Su¹⁸  | Casey Ta⁷  | Gaurav Vaidya¹  | Rosina Weber¹⁹  | Qi Wei³  | Mark Williams¹³  | Chunlei Wu¹⁸  | Colleen Xu¹⁸  | Chase Yakaboski¹⁰  | The Biomedical Data Translator Consortium

¹Renaissance Computing Institute, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA | ²Copperline Professional Solutions, LLC, Pittsboro, North Carolina, USA | ³Institute for Systems Biology, Seattle, Washington, USA | ⁴University of California – San Francisco, San Francisco, California, USA | ⁵University of North Carolina – Chapel Hill, Chapel Hill, North Carolina, USA | ⁶University of Alabama – Birmingham, Birmingham, Alabama, USA | ⁷Columbia University, New York, New York, USA | ⁸Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA | ⁹Institute of Data Science, Maastricht University, Maastricht, the Netherlands | ¹⁰Dartmouth College, Hanover, New Hampshire, USA | ¹¹Pennsylvania State University, University Park, Pennsylvania, USA | ¹²Montana State University – Billings, Billings, Montana, USA | ¹³National Center for Advancing Translational Sciences, Bethesda, Maryland, USA | ¹⁴CoVar, LLC, Durham, North Carolina, USA | ¹⁵Lawrence Berkeley National Laboratory, Berkeley, California, USA | ¹⁶Oregon State University, Corvallis, Oregon, USA | ¹⁷Beshenich Muir & Associates, LLC, Leavenworth, Kansas, USA | ¹⁸The Scripps Research Institute, San Diego, California, USA | ¹⁹Drexel University, Philadelphia, Pennsylvania, USA

Correspondence: Karamarie Fecho (kfecho@copperlineprofessionalsolutions.com)

Received: 26 November 2024 | **Revised:** 23 April 2025 | **Accepted:** 6 June 2025

Funding: This work was supported by the NCATS Biomedical Data Translator program (other transaction awards OT2TR003434, OT2TR003436, OT2TR003428, OT2TR003448, OT2TR003427, OT2TR003430, OT2TR003433, OT2TR003450, OT2TR003437, OT2TR003443, OT2TR003441, OT2TR003449, OT2TR003445, OT2TR003422, and OT2TR003435; contract number 75N95021P00636). Additional funding was provided by the NCATS Intramural Research Program (ZIA TR000276).

ABSTRACT

The growing availability of biomedical data offers vast potential to improve human health, but the complexity and lack of integration of these datasets often limit their utility. To address this, the Biomedical Data Translator Consortium has developed an open-source knowledge graph-based system—Translator—designed to integrate, harmonize, and make inferences over diverse biomedical data sources. We announce here Translator's initial public release and provide an overview of its architecture, standards, user interface, and core features. Translator employs a scalable, federated, knowledge graph framework for the integration of clinical, genomic, pharmacological, and other biomedical knowledge sources, enabling query retrieval, inference, and hypothesis generation. Translator's user interface is designed to support the exploration of knowledge relationships and the generation of insights, without requiring deep technical expertise and gradually revealing more detailed evidence, provenance, and confidence information, as needed by a given user. To demonstrate Translator's application and impact, we highlight features of the user interface in the context of three real-world use cases: suggesting

Karamarie Fecho and Gwênlyn Glusman contributed equally as primary/lead authors; all other authors are listed in alphabetical order.

See collaborating/consortial authors in Appendix S1.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2025 The Author(s). *Clinical and Translational Science* published by Wiley Periodicals LLC on behalf of American Society for Clinical Pharmacology and Therapeutics.

potential therapeutics for patients with rare disease; explaining the mechanism of action of a pipeline drug; and screening and validating drug candidates in a model organism. We discuss strengths and limitations of reasoning within a largely federated system and the need for rich concept modeling and deep provenance tracking. Finally, we outline future directions for enhancing Translator's functionality and expanding its data sources. Translator represents a significant step forward in making complex biomedical knowledge more accessible and actionable, aiming to accelerate translational research and improve patient care.

1 | Introduction

Biomedical research has become increasingly specialized and siloed. Much research focuses on the treatment of symptoms, rather than on the elucidation of the underlying etiology of disease. Furthermore, each discipline may use distinct vocabularies, ontologies, data sources, and data representations to describe and report their research, making the resulting knowledge not interoperable. For example, an immunologist may focus on interactions between immune cells or molecules, without considering the anatomical context, the modulatory role of non-immune biological systems such as the central nervous system or the autonomic nervous system, or knowledge derived from other disciplines.

The Biomedical Data Translator (“Translator”) program was conceptualized and funded in 2016 by the National Center for Advancing Translational Sciences (NCATS). The program aims to radically change the concept of “disease” from one that is symptom-based and narrowly-focused to one that is mechanism-based and considers, for example, the interplay between biomedical entities and biological systems, the relationship between rare and common disease, and the impact of social and environmental determinants of health and disease [1].

The Translator program aims to achieve this ambitious goal through the integration, harmonization, and exploration across hundreds of open clinical and biomedical data sources, using a knowledge graph (KG)-based approach to discovery [1–5], where the answers are fully qualified with evidence and provenance.

Here, we announce the initial public release of the Translator system, describe its user interface (UI) and supported features, and demonstrate the system's application across several high-impact, real-world clinical and translational use cases. We close with a discussion on potential future directions and long-term sustainability plans.

2 | Translator Overview

2.1 | Knowledge Graphs

In a KG, the basic unit of knowledge is the semantic “triple,” or the core “subject-predicate-object” relationship [6]. The “subject” and “object” are represented as “nodes” and map to fundamental domain concepts (e.g., gene, chemical, phenotype, disease) that are findable in an existing biological or biomedical ontology. The “predicate” is represented as an “edge”

and describes the relationship between nodes. For example, a core assertion may state that “prednisone-treats-asthma” [7]. The Translator system maintains the specificity of core assertions by capturing the nuance and context of a given assertion in subject-, object-, and statement-level qualifiers, or node and edge properties, depending on how the KG is technically modeled and implemented [8]. Finally, a “schema” is typically used to specify syntactic and semantic rules that constrain the KG by specifying how knowledge can be represented. The Translator Consortium has developed Biolink Model [4] as the program's preferred data model and high-level schema, since adopted by several other large-scale consortia such as the Monarch Initiative [9] and the Illuminating the Druggable Genome initiative [10].

A KG thus contains interconnected core assertions that together provide a rich source of information upon which reasoning algorithms can be applied, ranging in complexity from simply looking up existing or direct assertions to applying more complex chains of reasoning that infer new assertions that may or may not directly exist within the graph. As such, KGs provide a powerful architecture for knowledge representation and discovery, upon which many reasoning tools can be built to tap into and exploit the structured knowledge. For instance, KGs can be used to explain disease pathophysiology, suggest drug candidates for testing in vitro or by way of predictive computational models, propose existing drugs for repurposing, explore potential mechanistic relationships between biomedical entities such as genes and chemicals or biological systems such as the immune system and the central nervous system, and explain the relationships between rare and common diseases.

2.2 | Translator Architecture

The Translator system is primarily federated and hierarchical in design, with numerous contributing components and services, each serving a distinct role:

- The UI (see next section) provides access to the Translator system.
- Knowledge Providers (KPs) curate and integrate hundreds of openly accessible, domain-specific data and knowledge sources. These include specialized biomedical databases, text-mined resources, clinical knowledge repositories, and structured ontologies, all made interoperable to facilitate seamless data integration, sharing, and utilization.
- Autonomous Relay Agents (ARAs) integrate the knowledge provided by the KPs and apply reasoning and

inference algorithms to derive new insights in response to user queries.

- The Autonomous Relay System (ARS) serves as a central relay station: it receives and broadcasts user queries from the UI to the ARAs; merges, scores, and sorts the results that are returned from the ARAs; provides node annotations (e.g., US Food and Drug Administration (FDA) approval status on chemicals); and returns the merged-and-scored results to the UI.
- The Standards and Reference Implementation (SRI) component coordinates the integration, harmonization, normalization, and interoperability across the many diverse Translator components by implementing the Translator Reasoner Application Programming Interface (TRAPI) standard [11], Biolink Model, and tools to support identifier normalization, node annotation, and ordering and organizing of results.

We provide more details in the [Supporting Information](#) (“Translator Supporting Infrastructure and Operations”, pp. 2–8) and within a prior publication [3].

2.3 | Translator User Interface

The Translator system is accessible via a UI that supports multiple templated queries and provides for deep exploration of results and all supporting evidence and provenance (see Video 1 for a user tutorial).

The current version of the UI provides users with five distinct templated question or search options, accessible via a dropdown menu (Figure 1A). These options were introduced to streamline the design process and ensure that the site remains user-friendly. Each search option is accompanied by an icon representing the category of results the search will return, and the expected input is indicated in parentheses. These “visual cues” help users understand what to search for and what results to expect, enhancing the overall usability of the site.

After a user selects a specific query, they then can input a search term (Figure 1B). The UI helps users identify their desired search terms by providing available matching terms relevant to the query type. The UI then submits the query to the ARS for processing, and the retrieved results are presented to the user as a list of answers sorted by score (see [Supporting Information](#): “Translator Supporting Infrastructure and Operations – ARS”, pp. 6–7). A filtering system was developed to assist the user in reducing the number of returned query results (Figure 1C).

The relationship between the queried concept and each of the results (“paths”) may have been retrieved directly from a knowledge source, or indirectly by inference across multiple direct connections (Figure 1D). Each result may therefore include a set of intermediary nodes and a network of edges relating them to the query. A user can view attributes like the description, the categories that a drug fits into, a description of returned chemicals or drugs, and other properties. The UI also offers users an option to display the relationships within a network (“graph”) view that provides more detailed information about node connectivity than the path view. The most granular view displays the full evidence and provenance supporting the connection between two concepts.

3 | Application Use Cases

3.1 | Suggesting Potential Therapeutics for Patients With Rare Disease

This use case originated with a patient who presented with a rare condition determined to be caused by a loss-of-function variant in the *MET* gene [12], clinically manifesting as non-alcoholic fatty liver disease (NAFLD). A typical strategy for identifying treatment options is to search for drugs that increase *MET* activity; however, overexpression of *MET* is associated with multiple forms of cancer [13]. As such, we queried Translator to identify drugs that increase *MET* and are associated with NAFLD pathways but are *not* associated with cancer



VIDEO 1 | Short tutorial: Overview of the Biomedical Data Translator (“Translator”) system and example use-case applications. Translator was designed to accelerate translational science through a powerful resource to aid in the navigation and interpretation of complex biomedical data. By integrating and semantically harmonizing hundreds of diverse data sources and then applying complex reasoning algorithms to the integrated and harmonized data, Translator enables users to explore both direct and indirect relationships between genes, diseases, drugs, and other biomedical entities, as well as the provenance and evidence supporting all assertions. Learn more at ui.transltr.io. Video content can be viewed at Appendix S2.

A. Select a template

Smart Query Pathfinder Query

Select a question and enter a search term to find paths between biomedical entities

What drugs may treat conditions related to...

What drugs may treat conditions related to a disease?

What chemicals may increase the activity of a gene?

What chemicals may decrease the activity of a gene?

What genes' activity may be increased by a chemical?

What genes' activity may be decreased by a chemical?

Enter a Disease or Phenotype

Use terms for the most up-to-date results.

Bethlem Myopathy Breast Cancer

B. Select a disease

Smart Query Pathfinder Query

Select a question and enter a search term to find paths between biomedical entities

What drugs may treat conditions related to...

adreno

Adrenocortical Insufficiency Disease

Adrenogenital Syndrome Disease

Adrenoleukodystrophy Disease

Adrenocortical Carcinoma, Hereditary Disease

Adrenomyodystrophy Disease

C. Filter results by chemical category and role

Filters Clear All

Text Filter

Results

Development Stage

Drug

Phase 2 Drug

Phase 3 Drug

Other

Clinical Trial Indications

Chemical Classification

Adjuvant

Antineoplastic A...

Results

Showing 1-10 of 352 (out of 1119)

Development Stage Drug Chemical Classification Diagnostic Agent Chemical Classification Antineoplastic Role Chemical Classification Adjuvant

Hydrocortisone Cypionate

Cortisone Acetate

Prednisolone

Prednisone

Dexamethasone Sodium Phosphate

D. Explore direct and indirect evidence in support of an answer

Filters Clear All

Text Filter

Results

Development Stage

Drug

Phase 2 Drug

Phase 3 Drug

Other

Clinical Trial Indications

Chemical Classification

Adjuvant

Antineoplastic A...

Results

Showing 1-10 of 352 (out of 1119)

Development Stage Drug Chemical Classification Diagnostic Agent Chemical Classification Antineoplastic Role Chemical Classification Adjuvant

Hydrocortisone Cypionate

Paths

Graphs

Direct

Indirect

Supporting Path 4.a Evidence

Direct Path 2 Evidence

FIGURE 1 | Legend on next page.

FIGURE 1 | Translator UI and select features. (A) Users select a templated question or search option as an initial step when using Translator. (B) After selecting a template, users then select a search term, “Adrenocortical insufficiency” in this example, facilitated by the autocomplete feature. (C) After running a query, users may filter results by a number of facets, including “chemical category” and “chemical role classification”. (D) Translator evidence and provenance are returned by the UI to support each answer. In this example, the user asked: “what drugs may treat conditions related to adrenocortical insufficiency?” After filtering for “drug” and excluding “adjuvant”, “aetiopathogenetic role”, and “diagnostic agent”, Translator returned “hydrocortisone cypionate” as the top result. Three direct paths were returned. One direct path was supported by two different primary knowledge sources, DrugCentral and SIDER, with URLs to Translator Wiki pages that include a brief description of the source and example edges or records. The indirect evidence included a path that asserted “hydrocortisone cypionate is in clinical trials for chronic primary adrenal insufficiency, which is a subclass of adrenocortical insufficiency,” with URLs provided for the relevant clinical trials. Results can be found at: <https://ui.transltr.io/results?l=Adrenocortical%20Insufficiency&i=MONDO:0000004&t=0&r=0&q=453bdb5c-e076-4471-ad17-90185d364b18> (generated 02:44 PM EST, October 28, 2024).

(Figure 2A). When asked “what chemicals may increase the activity of MET [human]?”, Translator returned 900 results. We then filtered the results down to 283 by requiring the chemical category to be “Drug” and then further filtered to six by also requiring the chemical role classification to be “Pathway Inhibitor”. Three of these results (etoposide, hydroxyurea, and methotrexate) are antineoplastic agents, identified as indirect assertions (inferences), with multiple supporting paths each (Figure 2B). Requiring instead the chemical role classification to be “Anti-inflammatory Agent”, given the role of inflammation in the pathology of NAFLD [14], we identified 10 results, two of which (tretinoin and prednisolone) also are used as antineoplastic agents among other indications. In combination, we were able to shortlist five potential drug treatments that were deemed by our clinical experts as being worthy of further analysis: etoposide; hydroxyurea; methotrexate; tretinoin; and prednisolone. For each of these drug candidates, Translator provided detailed annotation and reasoning provenance that can now be evaluated, together with external or proprietary information not provided by Translator, by our clinical colleagues for viability, in terms of toxicity profile, potential side-effects, etc. For instance, LiverTox indicates that etoposide has a relatively low potential for irreversible hepatic toxicity [15], and Drug Bank indicates that the drug has relatively few side-effects [16].

3.2 | Exploring and Explaining the Mechanism of Action of a Pipeline Drug

Vascular Ehlers Danlos Syndrome (vEDS) is a rare disease caused by mutations in the *COL3A1* gene [17] for which there are no approved treatments. Celiprolol, a drug that is indicated for the treatment of mild to moderate hypertension and effort-induced angina pectoris [18, 19], is in development for repurposing in the treatment of vEDS (Acer Therapeutics, since acquired by Zevra Therapeutics) [20]. However, the drug development effort has been hampered by a lack of an acceptable animal model and uncertainty regarding celiprolol’s potential mechanism of action in the treatment of vEDS (Figure 3A). As such, we turned to Translator as a non-animal methodology [21, 22] for deeper exploration of the potential therapeutic efficacy of celiprolol, including its mechanism of action. Translator’s top answer to “what drugs may treat conditions related to vEDS?” was (s)-celiprolol (an enantiomer of celiprolol), with a direct assertion sourced from Drug Repurposing Hub [23] (Figure 3B). We then asked “what gene’s activity may be decreased/increased

by celiprolol?” The top result was a direct assertion from DrugBank [24], supported by a publication [25], that celiprolol decreases the activity of *ADRB1*, reinforcing celiprolol’s established mechanism of action as a selective β_1 receptor antagonist [18, 19] (Figure 3C). Among the indirect or inferred results was *COL1A1*, which Translator evidence stated “encodes the pro- α_1 chains of type I collagen”. The reasoning chain, and the supporting evidence and provenance, showed that celiprolol increases the activity of *ADRB2*, which then decreases the activity of *COL1A1*. The assertions were provided by DrugCentral [26] and Text Mining KP [27], respectively. Thus, the indirect inferred result both supports celiprolol’s established mechanism of action as a β_2 receptor partial agonist [18, 19] and suggests that the efficacy of celiprolol in the treatment of vEDS may be indirect and reflect a decrease in the expression or activity of *COL1A1* [28]. Indeed, mutations in *COL1A1* result in an alteration of the collagen pro- α_1 [1] chain that disrupts the structure of type 1 collagen fibrils, trapping collagen inside cells and increasing the risk of blood vessel rupture [28] As such, decreases in the altered protein chain would reduce the impact of a structural disruption in collagen.

3.3 | Screening and Validating Drug Candidates Using a Model Organism

Bi-allelic, loss-of-function variants in the *CAMSAP1* gene cause a rare syndrome that is associated with a distinct craniofacial appearance, primary microcephaly, severe neurodevelopmental delay, cortical visual impairment, and seizures [29]. Mutations in the CAMSAP homolog *ptrn-1* also cause increased convulsions and paralysis in (*ptrn-1*) mutant *C. elegans* worms [30, 31]. We used Translator to identify drug candidates targeting molecular signaling pathways known to involve CAMSAPs for subsequent screening in the *C. elegans* model (Figure 4B). The selected targets included the genes *DAPK1* [32], *MAP3K12* [30], *MAP3K5* (ASK1) [33], *EZR* [34], *RDX* [34], *MSN* [34], and *KIF2A* [35]. For each gene, we asked “what chemicals may decrease the activity of Gene X?” After filtering results by US FDA approval status, the evidence supporting each drug was manually evaluated by our colleagues, prioritizing those results with easily verifiable publications. For example, one result was for acetylsalicylic acid and included an indirect assertion that “acetylsalicylic acid decreases the activity of KIF2A”, with 10 supporting paths involving nine unique intermediary genes. One supporting path asserted that “acetylsalicylic acid causes increased activity or abundance of BDNF, which causes reduced

activity or abundance of KIF2A” [36, 37] (Figure 4B). After this curation process, our colleagues identified several candidates for further testing: acetylsalicylic acid; amantadine; quercetin; spermidine; epigallocatechin gallate; and resveratrol. Our

colleagues then conducted a low-throughput repurposed drug screen to determine the impact of the candidates in a convulsion/paralysis assay with wildtype versus CAMSAP (*ptrn-1*) mutant *C. elegans* worms [38]. Four of the six drugs impacted

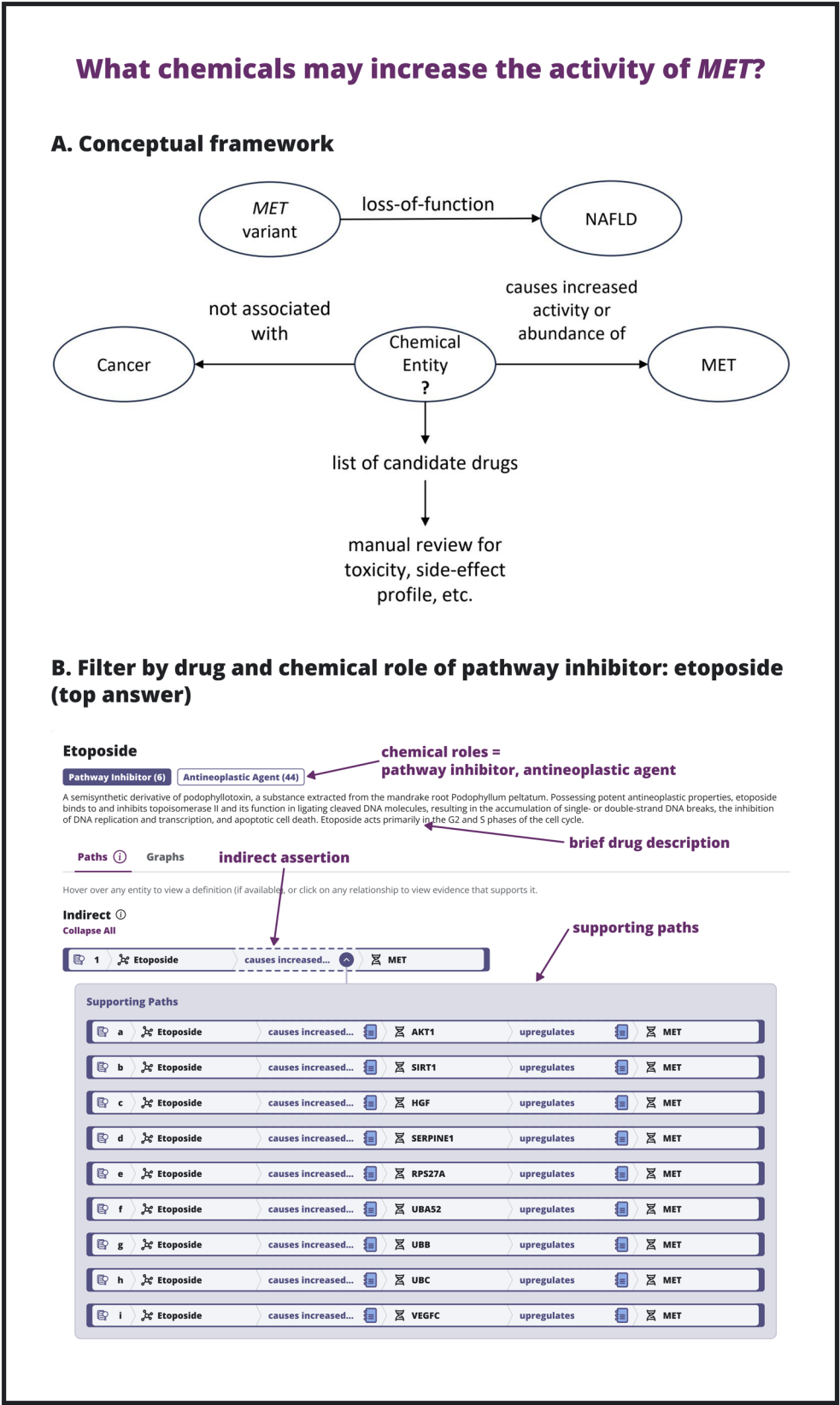


FIGURE 2 | Legend on next page.

FIGURE 2 | Use case: Suggesting potential therapeutics for a patient with rare disease. (A) Conceptual framework. This use case was driven by a patient with a loss-of-function genetic variant in the *MET* gene that presented clinically as non-alcoholic fatty liver disease (NAFLD). Translator was used to seek chemical entities that cause an increase in the activity or abundance of *MET* but are *not* associated with cancer, a common occurrence with overexpression of *MET*. (B) After filtering by “drug” and chemical role of “pathway inhibitor”, Translator returned etoposide as the top answer, along with a brief description of the drug and a flag noting that it has an additional chemical role of “antineoplastic agent”. The indirect assertion was accompanied by nine supporting paths. For example, one path asserted that “etoposide causes increased activity or abundance of AKT1, which up-regulates *MET*”. Results can be found at: [https://ui.transltr.io/results?l=MET%20\(Human\)&i=NCBIGene:4233&t=1&r=0&q=1e513a61-9d3a-4aee-90c4-8e42c004933f](https://ui.transltr.io/results?l=MET%20(Human)&i=NCBIGene:4233&t=1&r=0&q=1e513a61-9d3a-4aee-90c4-8e42c004933f) (generated 2:32 PM EST, October 25, 2024).

the mutant phenotype, with salicylic acid (derivative of acetyl-salicylic acid), amantadine, and quercetin causing a reduction in convulsions/paralysis, and epigallocatechin gallate causing an increase in convulsions/paralysis (data not shown). For spermidine, the phenotype in the CAMSAP mutant was too variable to be conclusive. Resveratrol had no effect (data not shown). Further testing is warranted to fully assess the selected candidates' therapeutic potential and resolve observed inconsistencies. Although Translator can provide suggestions and hypotheses for further investigation, along with full evidence and provenance to support results, the system does not suggest specific study protocols. Nonetheless, these findings underscore the power of Translator to propose drug candidates for further testing, using user-identified methodologies and protocols, and the importance of targeted testing for viable therapeutic candidates for rare genetic conditions like *CAMSAP1* deficiency.

4 | Discussion and Next Steps

The current Translator UI [39] embodies a philosophy of minimalist design and progressive disclosure, carefully crafted to manage the complexity of vast biomedical data and the Translator infrastructure. This approach prioritizes user experience by presenting a clean, intuitive interface that the user can customize, gradually revealing more detailed information as needed. This design strategy has proven effective in handling single-query interactions, allowing users to navigate from broad search terms to specific, evidence-backed relationships derived from the underlying Translator system. Herein, we present three application use cases, each one highlighting key features of the UI and the application of Translator to drive translational innovation by suggesting therapeutics for rare disease, explaining the mechanism of action of a pipeline drug, and providing candidate drug targets for subsequent screening and validation in a model organism.

Translator is not the only biomedical KG-based system available today; others include Causaly [40], Elsevier's EmBiology [41], BIOTEQUE [42, 43], Open Targets [44], and STRING [45] (also see Haas [46]). We note, however, that Translator differs from these systems in several distinct ways (see Table 1 for a qualitative comparison of Translator versus select biomedical KG-based systems). First, both the Translator UI and the underlying infrastructure are open source and publicly available, while most other systems are proprietary. Second, Translator is not simply a biomedical question-answer system; rather, Translator provides a comprehensive reasoning system that is largely federated and operates on numerous KGs to derive answers to user queries. Indeed, the federated nature of the Translator system allows for

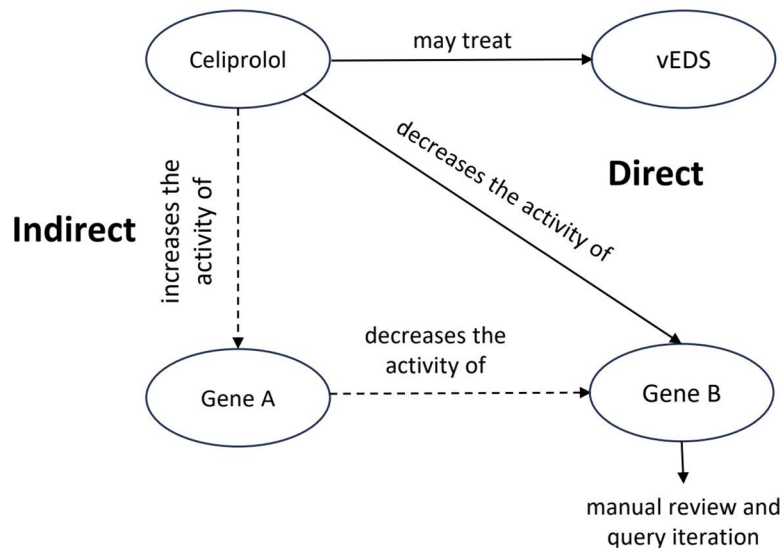
a hybrid approach to reasoning, leveraging the strengths of distinct reasoning systems and algorithms and operating on both static and dynamic underlying knowledge sources in near-real time. Other KG-based biomedical systems likewise invoke advanced reasoning algorithms, but these systems are designed to operate on single KGs and, in some cases, highly specialized data sources, such as pathways (BIOTEQUE) [42, 43] or proteins (STRING) [45]. Finally, a major strength and distinguishing feature of Translator is that the system is fully transparent and provides users with the underlying evidence, provenance, and confidence supporting its answers. The provenance includes the “primary knowledge source”, which is the source that provided a specific edge assertion (e.g., CTD), and the “knowledge level”, which is the type of knowledge that is being reported (e.g., a general statement of knowledge, a statistical association, computational predictions). Other examples of evidence and provenance for edges and results, include (when available) supporting publications, *p* values, and experimental data such as binding affinities. Translator provides confidence in results by way of a “Confidence Score”, which is a composite of individual scores returned from each of the ARAs, with each ARA developing its score based on its choice of evidence and logic. The UI provides a facet to filter out direct results for users who may be more interested in the novelty or originality of indirect results. (See “Result Scoring and Annotation” and “Evidence and Provenance” sections of the [Supporting Information](#), pp. 9–12.)

Another strength of Translator is the breadth of knowledge sources that are captured and integrated into the Translator system. This diversity also is likely to be a differentiator of Translator when comparing other biomedical KG systems, although it is difficult to fully ascertain the knowledge available within the proprietary KG systems. The Translator Consortium aims to include as many publicly available datasets as possible [1], regardless of country of origin, organism, or focus area. The only “restriction” placed on Translator knowledge sources is that they must be publicly available. The Consortium continually ingests new knowledge sources, with the goal of providing a broad range of data sources and types. Moreover, because Translator's knowledge sources are integrated and semantically harmonized, the reasoning algorithms that are applied to the data to derive new insights consider ALL of the available knowledge, with filters in place or under development to allow users to hone in on subsets of the available data (e.g., humans only) (See “Knowledge Sources and Knowledge Providers” section of the [Supporting Information](#), pp. 2–3, for additional information.)

Other considerations worth mentioning are the theoretical constructs of “dark data” and “blind spots” [47] and the related constructs of “known unknowns” and “unknown unknowns” [48].

What gene's activity may be decreased by celiprolol?

A. Conceptual framework



B. Direct result: *ADRB1*

Direct Path 1 Evidence

ADRB1 has decreased activity caused by Celiprolol · Mark as Unseen

The screenshot shows a path visualization interface. At the top, a path is displayed: **ADRB1** → "has decreased activity" → **Celiprolol**. Below this, a table lists supporting publications and knowledge sources.

TITLE	DATE(S)	JOURNAL	SNIPPET	SOURCE & RATIONALE
Effects of the antioxidative beta-blocker celiprolol on endothelial progenitor cells in hypertensive rats	2008	American Journal Of Hypertension	Endothelial progenitor cells (EPCs) derived from bone marrow migrate to areas of endothelial damage and repair them. EPC... Read More	DrugBank

Annotations on the screenshot:

- two direct edges**: Points to the path visualization.
- supporting publication**: Points to the first row of the table.
- primary knowledge source**: Points to the DrugBank source in the table.

C. Indirect result: *COL1A1*

The screenshot shows the interface for the indirect result *COL1A1*. It includes a description of the gene, a table of supporting paths, and annotations.

COL1A1
This gene encodes the pro- $\alpha 1$ chains of type I collagen whose triple helix comprises two $\alpha 1$ chains and one $\alpha 2$ chain. Type I is a fibril-forming collagen found in most connective tissues and is abundant in bone, cornea, dermis and tendon. Mutations in this gene are associated with osteogenesis imperfecta types I-IV, Ehlers-Danlos syndrome type VIA, Ehlers-Danlos syndrome Classical type, Caffey Disease and idiopathic osteoporosis. Reciprocal translocations between chromosomes 17 and 22, where this gene and the gene for platelet-derived growth factor beta are located, are associated with a particular type of skin tumor called dermatofibrosarcoma protuberans, resulting from unregulated expression of the growth factor. Two transcripts, resulting from the use of alternate polyadenylation signals, have been identified for this gene. [provided by R. Dalgleish, Feb 2008]

NAME	EVIDENCE	PATHS	SCORE
COL1A1	Publications (1) Sources (3)	2 Paths	1.06

Supporting Paths

NAME	EVIDENCE	PATHS	SCORE
COL1A1	Publications (1) Sources (3)	2 Paths	1.06

Annotations on the screenshot:

- indirect assertion**: Points to the path visualization.
- supporting path**: Points to the first row of the Supporting Paths table.

FIGURE 3 | Legend on next page.

FIGURE 3 | Use case: Exploring and explaining the mechanism of action of a drug. (A) Conceptual framework. This use case focused on celiprolol, a pipeline drug in clinical trials for the treatment of vascular Ehlers Danlos Syndrome (vEDS). The lack of an acceptable animal model of vEDS prompted the use of Translator to suggest a mechanism of action to support celiprolol's efficacy in the treatment of vEDS. Translator sought both direct and indirect evidence for celiprolol decreasing the activity of an unspecified Gene B. (B) Among the direct evidence was an assertion that "ADRB1 has decreased activity caused by celiprolol," with two direct edges, one of which was from DrugBank and included a supporting publication. (C) Among the indirect evidence was an assertion that "COL1A1 is downregulated by ADRB2, which has increased activity or abundance caused by celiprolol". Results can be found at: <https://ui.transltr.io/results?l=Celiprolol&i=CHEBI:94461&t=4&r=0&q=fd0628dd-54c5-490d-98b6-89483d35ad00> (generated 03:47 PM EST, October 18, 2024).

The Translator Consortium currently is exploring these topics in the context of a "Novelty Score" that is under development (see "Results Scoring and Annotation" section of the [Supporting Information](#), pp. 9–11). For example, consider a Novelty Score based on the recency of publications. Blind spots resulting from dark data will decrease over time as new publications arise and new edge predictions are made, but new blind spots will inevitably emerge to replace the known blind spots. The Translator Consortium is considering several approaches to account for these important considerations, including the implementation of a variety of filters to allow users to prioritize answers based on their interests. For example, a user may be interested in edges with little to no publication support as a way to capture "novelty".

The advent of large language models (LLMs) has sparked interest among many biomedical KG communities, including those highlighted in Table 1, in how best to leverage the strengths of LLMs, while minimizing their weaknesses. The Translator Consortium has been actively exploring approaches for integrating LLMs and generative AI into the Translator system (also see "LLM Integration" within the [Supporting Information](#), pp. 12–13). A number of approaches are in early development or under consideration, including a "summarization service" that generates high-level summaries of results with key mechanistic explanations highlighted; an "edge summarization service" that generates publication-supported summaries of specific edges in results; and a variety of embedding approaches such as a meta-data standard created by the Translator Consortium called *venomx* that captures important metadata about embeddings.

4.1 | Limitations

While Translator has numerous strengths and attractive features, it also has limitations. Even though Translator aims to return answers to user queries in a timely manner, the system can—for some queries—be slow and computationally expensive. A major goal of the next phase of development is to focus on performance improvements. Translator also is limited by the data modeling constraints inherent in KGs, for instance, the context, temporality, and high-dimensionality of data. The Translator Consortium is working with the Biolink Model community to improve their modeling approaches in support of nuanced queries and richly decorated answers. Related, Translator currently does not consider quantitative factors such as dose, potency, and duration of exposure in queries (e.g., side-effects of drug X at doses above Y mg/kg or toxicity of chemical X at concentrations above Y µg/mL). The system also does not support user filtering by quantitative metrics. However, Translator will return

quantitative metrics in results, when provided by the underlying knowledge sources. Approaches for leveraging quantitative metrics are a topic of discussion with the Translator Consortium. Additionally, users should be aware that Translator intentionally conflates certain biomedical concepts such as genes and proteins, reflecting the conflation in the underlying knowledge sources. Likewise, provenance is limited by the underlying knowledge sources, sometimes resulting in an inability to accurately trace complex and incomplete trails of ownership and algorithmic transformations of primary data and knowledge. Finally, due to the federated nature of the system and ongoing optimization of scoring and filtering methods, the results obtained for the same query may change over time.

4.2 | Potential Future Directions

Translator remains in active development. While the Translator Consortium has not yet finalized or prioritized next steps, several possibilities are under discussion.

For example, we plan to extend the number and types of templated queries that are supported by the UI. One such query that is in development is the "pathfinder" query. Instead of asking "what drugs may treat disease X?" or "what gene's activity may be increased/decreased by chemical Y?", the pathfinder query asks "how are entities X and Y related?" or "why does entity X interact with entity Y?" As such, pathfinder queries focus on explaining the relationship between entities by constructing paths between them. Paths are not required to be mechanistic, nor are they required to be direct, and predicates are agnostic, meaning that they are not specified by a user. Instead, pathfinder queries identify general paths between two entities, which may traverse through multiple intermediate nodes and invoke various predicates.

Another new type of query is the "set-input" query, which allows a user to select a set of specified biomedical entities as input terms, as well as a general type of output entity, and Translator returns paths that match many, but not necessarily all, of the input terms. In contrast, a batch query, which is currently supported by Translator, allows users to enter multiple input terms for a query, but the query is then executed sequentially or in parallel when possible, with independent results returned for each input term. The set-input query allows users to ask, for instance, "what genes are shared by a set of input phenotypes that characterize a patient with an undiagnosed rare disease?"

We also plan to improve and expand the scoring, ordering, and organizing of results that are displayed to users. For

example, we may group gene sets based on their membership to the same biological pathway. In addition, we may add new filters and facets to the UI, for example, exposing the Clinical Information Score. This score represents drug-disease associations as observed in real-world patient data and is currently being used for scoring and ordering of results (see [Supporting](#)

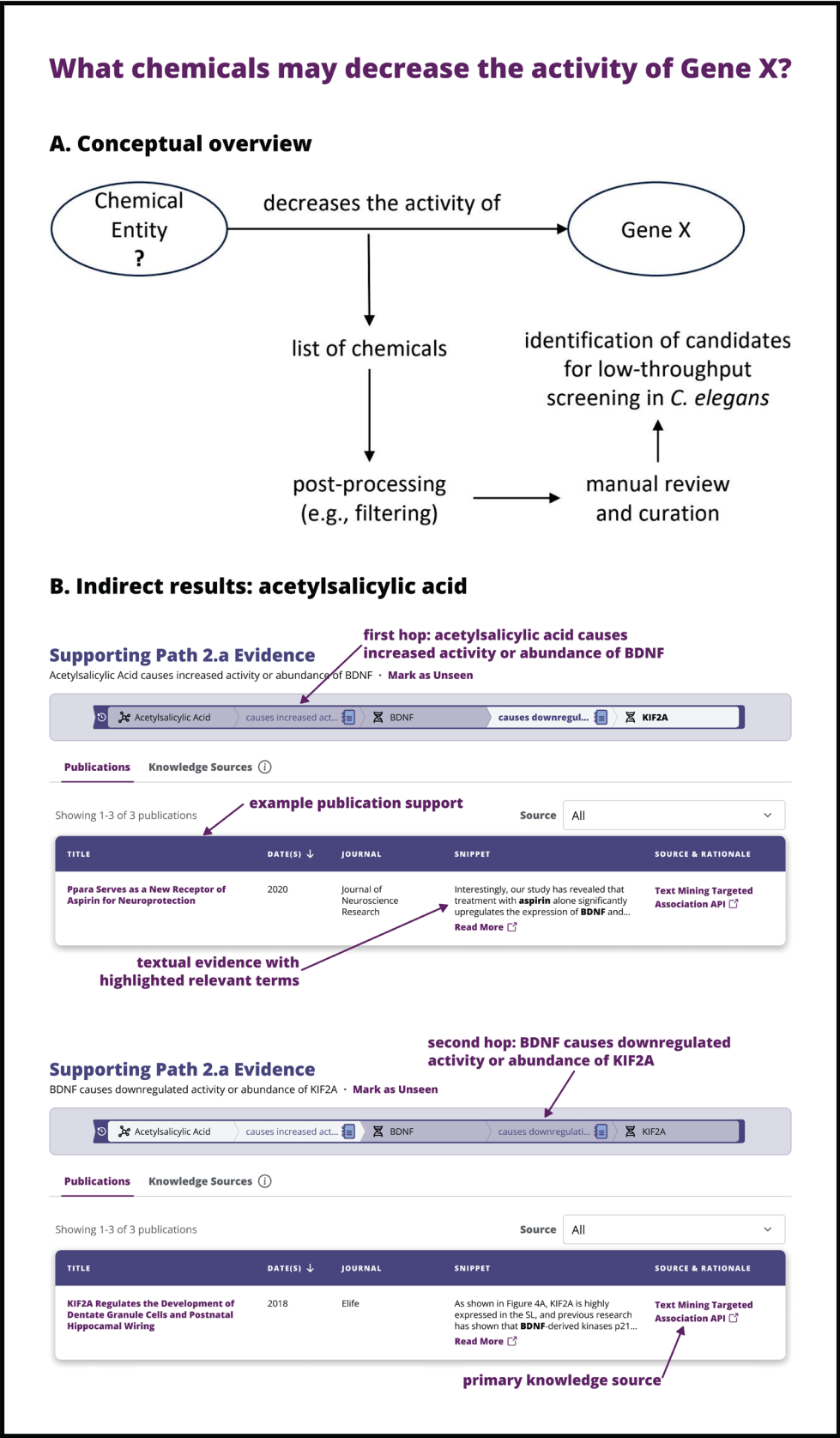


FIGURE 4 | Legend on next page.

FIGURE 4 | Use case: Screening and validating drug candidates using a model organism. (A) Conceptual framework. This use case focused on a rare syndrome associated with seizures and other neurological symptoms and caused by bi-allelic, loss-of-function variants in the *CAMSAP1* gene. Translator was used to propose drug candidates targeting molecular signaling pathways known to involve CAMSAPs for subsequent in vitro screening in a *C. elegans* model of the disease. (B) One of the drugs identified as a target for *KIF2A* was acetylsalicylic acid. An indirect supporting path asserted that “acetylsalicylic acid causes increased activity or abundance of BDNF, which causes downregulated activity or abundance of KIF2A”. Results can be found at: [https://ui.transltr.io/results?l=KIF2A%20\(Human\)&i=NCBIGene:3796&t=2&r=23d52e95&q=b3562865-bda6-4e8c-a83f-065004983843](https://ui.transltr.io/results?l=KIF2A%20(Human)&i=NCBIGene:3796&t=2&r=23d52e95&q=b3562865-bda6-4e8c-a83f-065004983843) (generated 03:03 PM CST, November 06, 2024).

TABLE 1 | Qualitative comparative analysis of Translator and several other biomedical KG-based systems.

System	Open source	Reasoning support	Centralized or federated	Evidence, provenance, confidence	AI/LLM integration
Translator [39]	Yes	Integrates multiple types and sources of biomedical data for comprehensive analysis across biomedical entity types and relationships, leveraging several distinct reasoning algorithms	Federated (integrates and semantically harmonizes data from hundreds of diverse sources)	Provides detailed evidence and provenance information, as well as confidence scores	Exploring AI integration
Causaly [40]	No (proprietary)	Uses AI for causal relationship identification, but limited in terms of the available biomedical entity types	Centralized (proprietary database)	Offers evidence and confidence metrics	Uses AI for data analysis
EmBiology [41]	No (proprietary)	Focuses on biological relationship exploration between select biomedical entities	Centralized (proprietary database)	Provides evidence and provenance details	Integrates AI for data insights
BIOTEQUE [42, 43]	Yes	Supports queries and reasoning targeting biological pathways	Centralized (integrates datasets into a single, unified platform)	Offers detailed provenance and confidence scores	Uses AI for data processing
Open Targets [44]	Yes	Provides tools for protein target identification and validation	Centralized (integrates datasets into a single platform)	Offers comprehensive evidence and provenance	Exploring AI integration
STRING [45]	Yes	Integrates known and predicted protein–protein interactions	Centralized (integrates multiple data sources)	Provides confidence scores for interactions	Not focused on AI/LLM integration

Abbreviations: AI = artificial intelligence; LLM = Large Language Model.

Information “Results Scoring and Annotation”, pp. 9–11 for details).

Another possible future direction is to expand the portfolio of workflow operations that are supported by Translator. For instance, we may want to chain workflow operations together in a manner that supports iterative and more complex queries.

Finally, with the major initial public release complete and the supporting infrastructure in place, the Translator Consortium is considering approaches to expand beyond its current team of software developers and end users in order to extend the reach of the program and the use of the Translator system, in support

of open-source software development, open team science, open community engagement, and long-term sustainability. For example, we may offer “Bring Your Own data/knowledge to Translator” (BYOT) capabilities. One BYOT prototype, the Gene Network Inference For Expression Regulation (GenNIFER) application [49], overlays Translator gene regulation knowledge for validation of inferred gene regulatory networks derived from single-cell RNA-seq datasets. BYOT capabilities will allow external users to more readily engage with the Translator Consortium and contribute to development of the Translator system by expanding the knowledge sources supported by Translator, while also helping to sustain the underlying data and software infrastructure and facilitating the external user’s own research.

Acknowledgments

The authors are grateful to members of the Publications Committee at NCATS for their review and approval of the manuscript for submission. Additionally, the authors would like to acknowledge Translator program leadership and the extramural and intramural support provided by NCATS. We note that ChatGPT was used to generate the first draft of Table 1. The authors subsequently refined the content, independent of any artificial intelligence tool. The following persons served as collaborating/consortial authors: Michel Arab, Shervin Abdollahi, Nichollette Acosta, Ayushi Agrawal, Stanley Ahalt, Nada Amin, Michael Bada, Krish Balar, Jim Balhoff, William Baumgartner, Jon-Michael Beasley, Emily Blake, Namdi Brandon, Richard Bruskiewich, Noel Burt, Jackson Callaghan, Terese Camp, Marco Alvarado Cano, Kathleen Carter, Remzi Celebi, Cheng-Han Chung, Larry Chung, Maria Costanzo, Margot Cousin, Tim Curry, Vlado Dancik, Jennifer Dougherty, Marc Duby, Stephen Edwards, Joe Farris, Nate Fehrmann, Marta Figueiral, Jason Flannick, Charlie Fox, Amy Glen, Prateek Goel, Skye Goetz, Joseph Gormley, Perry Haaland, Kristina Hanspers, Nomi Harris, Jeff Henrickson, Anthony Hickey, Eugene Hinderer, Maureen Hoatlin, Tursynay Issabekova, Rhea Karty, Yaphet Kebede, Keum Joo Kim, Simon King, Eric Klee, Michael R. Knowles, Ryan Koesterer, Daniel Korn, Ashok Krishnamurthy, Laura Lambert, Margaret W. Leigh, Jason Lin, Max Lupey, Chunyu Ma, Andrew Magis, Mallika Mainali, Meisha Mandal, Luis Mendoza, Matthew Might, Evan Morris, Kenny Morton, Sandrine Muller, Kamileh Narsinh, Olawumi Olanakanmi, Deepak Panwar, Michael Patton, David B. Peden, Amber Peters, Alexander Pico, Filippo Pinto e Vairo, Nishad Prakash, Guthrie Price, Sundareswar Pullela, Anthony Ragazzi, Navya Ramakrishnan, Jason Reilly, Everaldo Rodolpho, Greg Rosenblatt, Irit Rubin, Rayn Sakaguchi, Eugene Santos, Ava Schaak, Johnathan Schaff, Kevin Schaper, Shepherd Schurman, Brett Smith, David Smith, Sarah Stemmann, Michael Strasser, Harsha Sureshbabu, Mohsen Taheri, Alexander Tropsha, Sarah Tyndall, Adam Viola, Kevin Vizhalil, Madison Walker, Chen Wang, Max Wang, Braiden Warden, Paul B. Watkins, Ximing Wen, Chunhua Weng, Elizabeth White, Stephanie Won, Erica Wood, Braiden Worden, Yao Xiao, Rena Yang, Hong Yi, Qian Zhu, Tom Zisk.

Conflicts of Interest

S.E.B. and S.H. have received support from the NSF Convergence Accelerator Open Knowledge Networks to develop applications related to the SPOKE KG. All other authors declared no competing interests for this work.

Data Availability Statement

The Translator system is publicly accessible at <https://ui.transltr.io/>.

References

1. C. P. Austin, C. M. Colvis, and N. T. Southall, "Deconstructing the Translational Tower of Babel," *Clinical and Translational Science* 12, no. 2 (2019): 85, <https://doi.org/10.1111/cts.12595>.
2. Wikipedia, "Knowledge Graph," (2024), accessed November 21, 2024, https://en.wikipedia.org/w/index.php?title=Knowledge_graph&oldid=1255523142.
3. K. Fecho, A. E. Thessen, S. E. Baranzini, et al., "Progress Toward a Universal Biomedical Data Translator," *Clinical and Translational Science* 15, no. 8 (2022): 1838–1847, <https://doi.org/10.1111/cts.13301>.
4. D. R. Unni, S. A. T. Moxon, M. Bada, et al., "Biolink Model: A Universal Schema for Knowledge Graphs in Clinical, Biomedical, and Translational Science," *Clinical and Translational Science* 15, no. 8 (2022): 1848–1855, <https://doi.org/10.1111/cts.13302>.
5. The Biomedical Data Translator Consortium, "The Biomedical Data Translator Program: Conception, Culture, and Community," *Clinical*

and Translational Science 12, no. 2 (2019): 91–94, <https://doi.org/10.1111/cts.12592>.

6. Google, "Introducing the Knowledge Graph: Things, Not Strings," (2012), accessed November 21, 2024, <https://blog.google/products/search/introducing-knowledge-graph-things-not/>.
7. R. Dworski, G. A. Fitzgerald, J. A. Oates, and J. R. Sheller, "Effect of Oral Prednisone on Airway Inflammatory Mediators in Atopic Asthma," *American Journal of Respiratory and Critical Care Medicine* 149, no. 4 (1994): 953–959, <https://doi.org/10.1164/ajrccm.149.4.8143061>.
8. Wisecube AI – Research Intelligence Platform, "Knowledge Graphs: RDF or Property Graphs. Which One Should You Pick?," accessed November 21, 2024, <https://www.wisecube.ai/blog/knowledge-graphs-rdf-or-property-graphs-which-one-should-you-pick/>.
9. T. E. Putman, K. Schaper, N. Matentzoglou, et al., "The Monarch Initiative in 2024: An Analytic Platform Integrating Phenotypes, Genes and Diseases Across Species," *Nucleic Acids Research* 52, no. D1 (2024): D938–D949, <https://doi.org/10.1093/nar/gkad1082>.
10. K. R. Sharma, C. M. Colvis, G. P. Rodgers, and D. M. Sheeley, "Illuminating the Druggable Genome: Pathways to Progress," *Drug Discovery Today* 29, no. 3 (2024): 103805, <https://doi.org/10.1016/j.drudis.2023.103805>.
11. "NCATSTranslator/ReasonerAPI," (2024), accessed November 21, 2024, <https://github.com/NCATSTranslator/ReasonerAPI>.
12. MET, "MET Proto-Oncogene, Receptor Tyrosine Kinase [*Homo sapiens* (Human)]," NCBI, accessed November 22, 2024, <https://www.ncbi.nlm.nih.gov/gene/4233>.
13. G. E. Wood, H. Hockings, D. M. Hilton, and S. Kermorgant, "The Role of MET in Chemotherapy Resistance," *Oncogene* 40, no. 11 (2021): 1927–1941, <https://doi.org/10.1038/s41388-020-01577-5>.
14. K. H. Kim and M. S. Lee, "Pathogenesis of Nonalcoholic Steatohepatitis and Hormone-Based Therapeutic Approaches," *Frontiers in Endocrinology* 9 (2018): 485, <https://doi.org/10.3389/fendo.2018.00485>.
15. *LiverTox: Clinical and Research Information on Drug-Induced Liver Injury* (National Institute of Diabetes and Digestive and Kidney Diseases, 2012), Etoposide. [Updated 2018 Feb 25], accessed March 19, 2025, <https://www.ncbi.nlm.nih.gov/books/NBK548102/>.
16. "Etiposide," DrugBank, accessed March 19, 2025, <https://go.drugbank.com/drugs/DB00773>.
17. P. H. Byers, "Vascular Ehlers-Danlos Syndrome," in *GeneReviews*, ed. M. P. Adam, J. Feldman, G. M. Mirzaa, R. A. Pagon, S. E. Wallace, and A. Amemiya (University of Washington, 1993), accessed November 22, 2024, <http://www.ncbi.nlm.nih.gov/books/NBK1494/>.
18. J. J. Nawarskas, A. Cheng-Lai, and W. H. Frishman, "Celiprolol: A Unique Selective Adrenoceptor Modulator," *Cardiology in Review* 25, no. 5 (2017): 247–253, <https://doi.org/10.1097/CRD.0000000000000159>.
19. "Celiprolol," DrugBank, accessed November 22, 2024, <https://go.drugbank.com/drugs/DB04846>.
20. Science & Pipeline/Zevra Therapeutics. (2022), accessed November 22, 2024, <https://zevra.com/science-pipeline/>.
21. "A Strategic Roadmap for Establishing New Approaches to Evaluate the Safety of Chemicals and Medical Products in the United States," National Toxicology Program, accessed November 22, 2024, <https://ntp.niehs.nih.gov/whatwestudy/niceatm/natl-strategy>.
22. "NICEATM: Alternative Methods," National Toxicology Program, accessed November 22, 2024, <https://ntp.niehs.nih.gov/whatwestudy/niceatm>.
23. "Drug Repurposing Hub," (2023), NCATS Translator GitHub Repository, accessed November 22, 2024, <https://github.com/NCATSTranslator/Translator-All/wiki/Drug-Repurposing-Hub>.

24. “DrugBank,” (2023), NCATS Translator GitHub Repository, accessed November 22, 2024, <https://github.com/NCATSTranslator/Translator-All/wiki/DrugBank>.
25. E. H. Yao, N. Fukuda, T. Matsumoto, et al., “Effects of the Antioxidative Beta-Blocker Celiprolol on Endothelial Progenitor Cells in Hypertensive Rats,” *American Journal of Hypertension* 21, no. 9 (2008): 1062–1068, <https://doi.org/10.1038/ajh.2008.233>.
26. “DrugCentral,” (2023), NCATS Translator GitHub Repository, accessed November 22, 2024, <https://github.com/NCATSTranslator/Translator-All/wiki/DrugCentral>.
27. “Text-Mined Assertion KP,” (2024), NCATS Translator GitHub Repository, accessed November 22, 2024, <https://github.com/NCATSTranslator/Translator-All/wiki/Text%E2%80%90mined-Assertion-KP>.
28. “COL1A1 Gene. MedlinePlus Genetics,” (2019), accessed November 22, 2024, <https://medlineplus.gov/genetics/gene/col1a1/>.
29. R. Khalaf-Nazzal, J. Fasham, K. A. Inskeep, et al., “Bi-Allelic CAMSAP1 Variants Cause a Clinically Recognizable Neuronal Migration Disorder,” *American Journal of Human Genetics* 109, no. 11 (2022): 2068–2079, <https://doi.org/10.1016/j.ajhg.2022.09.012>.
30. J. D. Marcette, J. J. Chen, and M. L. Nonet, “The *Caenorhabditis elegans* Microtubule Minus-End Binding Homolog PTRN-1 Stabilizes Synapses and Neurites,” *eLife* 3 (2014): e01637, <https://doi.org/10.7554/eLife.01637>.
31. C. E. Richardson, K. A. Spilker, J. G. Cueva, J. Perrino, M. B. Goodman, and K. Shen, “PTRN-1, a Microtubule Minus End-Binding CAMSAP Homolog, Promotes Microtubule Function in *Caenorhabditis elegans* Neurons,” *eLife* 3 (2014): e01498, <https://doi.org/10.7554/eLife.01498>.
32. M. Chuang, T. I. Hsiao, A. Tong, S. Xu, and A. D. Chisholm, “DAPK Interacts With Patronin and the Microtubule Cytoskeleton in Epidermal Development and Wound Repair,” *eLife* 5 (2016): e15833, <https://doi.org/10.7554/eLife.15833>.
33. R. Hayakawa, T. Hayakawa, K. Takeda, and H. Ichijo, “Therapeutic Targets in the ASK1-Dependent Stress Signaling Pathways,” *Proceedings of the Japan Academy. Series B, Physical and Biological Sciences* 88, no. 8 (2012): 434–453, <https://doi.org/10.2183/pjab.88.434>.
34. M. Arpin, D. Chirivino, A. Naba, and I. Zwaenepoel, “Emerging Role for ERM Proteins in Cell Adhesion and Migration,” *Cell Adhesion & Migration* 5, no. 2 (2011): 199–206, <https://doi.org/10.4161/cam.5.2.15081>.
35. J. Atherton, K. Jiang, M. M. Stangier, et al., “A Structural Model for Microtubule Minus-End Recognition and Protection by CAMSAP Proteins,” *Nature Structural & Molecular Biology* 24, no. 11 (2017): 931–943, <https://doi.org/10.1038/nsmb.3483>.
36. N. Homma, R. Zhou, M. I. Naseer, A. G. Chaudhary, M. H. Al-Qahtani, and N. Hirokawa, “KIF2A Regulates the Development of Dentate Granule Cells and Postnatal Hippocampal Wiring,” *eLife* 7 (2018): e30935, <https://doi.org/10.7554/eLife.30935>.
37. D. Patel, A. Roy, and K. Pahan, “PPAR α Serves as a New Receptor of Aspirin for Neuroprotection,” *Journal of Neuroscience Research* 98, no. 4 (2019): 626, <https://doi.org/10.1002/jnr.24561>.
38. R. Pandey, S. Gupta, S. Tandon, O. Wolkenhauer, J. Vera, and S. K. Gupta, “Baccoside A Suppresses Epileptic-Like Seizure/Convulsion in *Caenorhabditis elegans*,” *Seizure* 19, no. 7 (2010): 439–442, <https://doi.org/10.1016/j.seizure.2010.06.005>.
39. “Translator User Interface,” accessed March 19, 2025, <https://ui.transltr.io/>.
40. Causaly, 2024, accessed November 22, 2024, <https://www.causaly.com/>.
41. *EmBiology, Biological Data Structured for Insights* (Elsevier, 2025), accessed November 22, 2024, <https://www.elsevier.com/products/embiology>.
42. “BIOTEQUE: Metapath Explorer,” accessed November 22, 2024, <https://bioteque.irbbarcelona.org/>.
43. “MetaPath Explorer GitHub Repository,” (2022), accessed November 22, 2024, <https://github.com/meb-team/MetaPathExplorer>.
44. “Open Targets,” accessed November 22, 2024, <https://www.opentargets.org/>.
45. “STRING: Functional Protein Association Networks,” (2024), accessed November 22, 2024, <https://string-db.org/>.
46. R. Haas, “A Survey of Biomedical Knowledge Graphs and of Resources for Their Construction,” (2024), accessed November 22, 2024, <https://github.com/robert-haas/awesome-biomedical-knowledge-graphs/blob/main/target/bmkg.pdf>.
47. D. J. Hand, *Dark Data: Why What You Don't Know Matters* (Princeton University Press, 2020).
48. M. Shaker and D. Mauger, “Applying the Clinical Literature to a Science of Uncertainty and an Art of Probability,” *Journal of Allergy and Clinical Immunology. In Practice* 9, no. 12 (2021): 4233–4234, <https://doi.org/10.1016/j.jaip.2021.08.024>.
49. Gene Network Inference for Expression Regulation (GenNIFER), accessed November 25, 2024, <https://di2ag.github.io/gennifer/>.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.