

知識統合から知識探索へー TogoIDとTogoDX



守屋 勇樹, 池田 秀也, 片山 俊明, TogoID/TogoDX 開発チーム

大学共同利用機関法人 情報・システム研究機構 国立遺伝学研究所 バイオデータ研究拠点 (BSI) ライフサイエンス統合データベース部門 (DBCLS)

moriya@dbcls.rois.ac.jp



DBCLSでは、生命科学データ統合基盤としてTogoIDとTogoDXを開発している。TogoIDは、NCBI Gene、UniProt、Ensembl、OMIMなど異なるデータベース間の識別子 (ID) の対応関係を統合し、遺伝子・タンパク質・疾患・化合物など多様な生命科学データを横断的に接続するサービスである。

一方、TogoDXは、遺伝子、タンパク質、疾患、化合物などについて、発現、機能、疾患関連性、相互作用などの属性情報を統合し、多面的な知識探索を可能にする。これらの基盤により、複数データベースを横断した解析や知識発見を容易にし、疾患メカニズムの理解、創薬標的探索、バイオマーカー探索など、データ駆動型生命科学研究の加速が期待される。

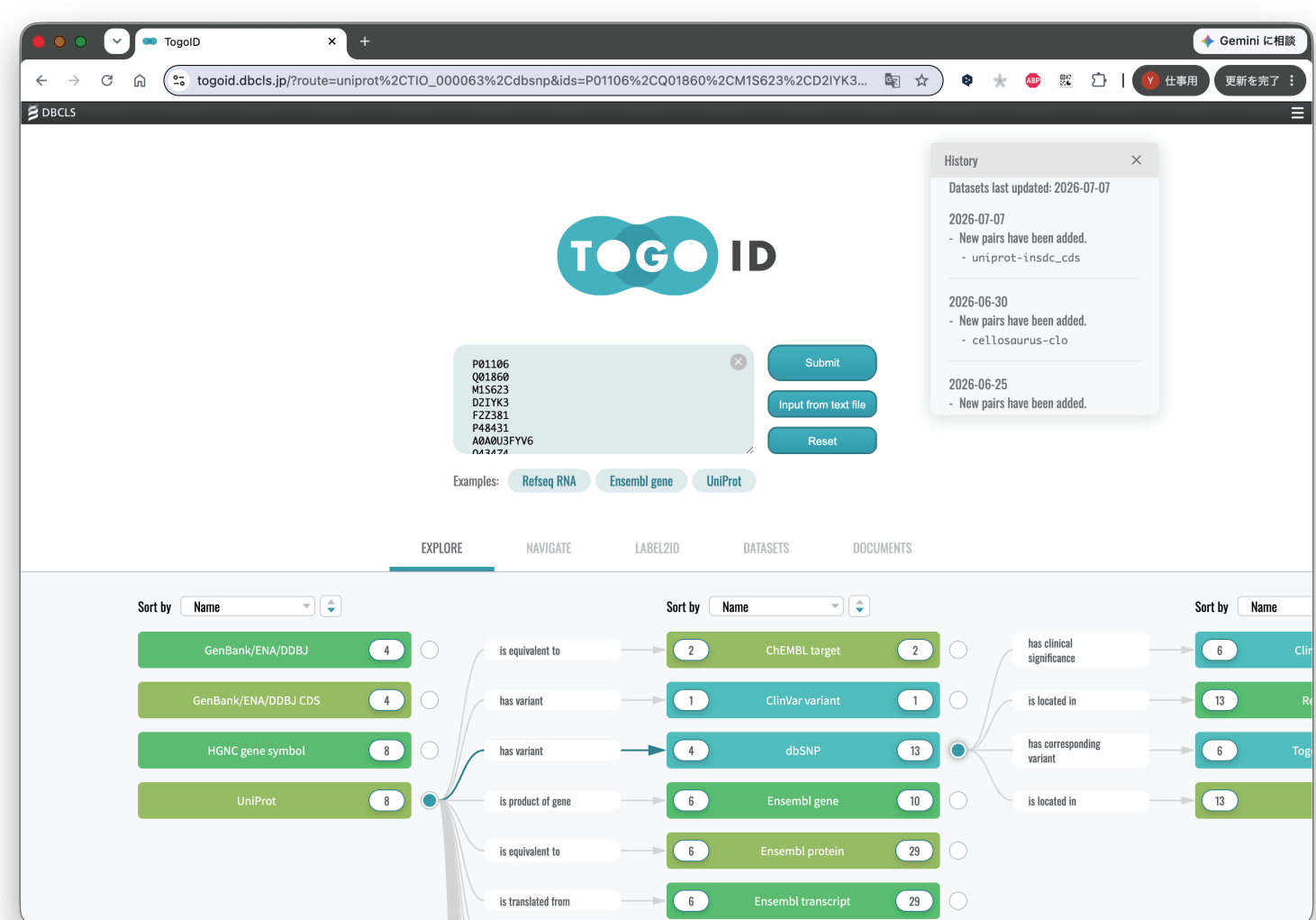
生命科学分野における多様なデータベース間の ID を相互変換するサービス



ヒトに関するデータを統合的に探索・俯瞰・抽出するためのアプリケーション

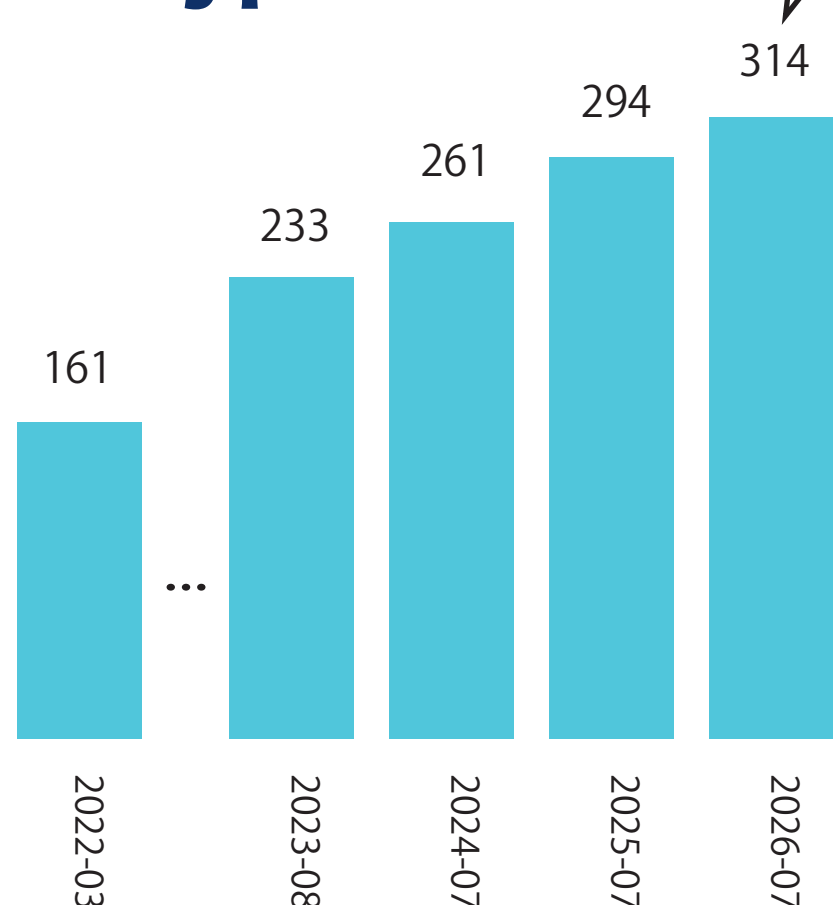
特徴

- 異なる生命科学データベース間の ID 対応を統合
- カテゴリ跨いだ変換
- 多段階 (multi-hop) の ID 変換
- 1 対多・多対多の変換に対応
- 変換経路の探索
- ID へのラベルやアノテーション付与
- 関係の意味をオントロジーで定義
- 定期的なデータ更新



データ数推移

ID type ペア数



各データセットの取得方法や更新方法、ID 対応関係の定義を togoid-config と呼ぶシステムで管理している。各データセットはシンプルな設定ファイルとして記述されており、新しいデータセットやデータセット間の対応関係も、設定ファイルを追加・編集するだけで容易に登録できる。設定は GitHub 上でオープンに管理されているため、コミュニティからの Pull Request による追加・改善を受け付けており、誰でも TogoID の対応データセットの拡充に貢献できる仕組みとなっている。

<https://github.com/togoid/togoid-config>

Year	Database	ID types	Pairs	Ontology	Labels
2022	48	65	161	77	0
2024	73	105	263	133	65
2026	82	119	314	147	75

プログラマブルな問い合わせ

TogoID は ID 変換の他にラベル取得やアノテーション取得機能を備えているが、異なるバックエンドサービスによって提供されているため、これらの API のプログラマブルな利用が困難であった。そこで Python および R ライブラリ、ならびにコマンドラインインターフェース (CLI) を開発した。これらにより、ユーザーが各バックエンドと個別にやり取りすることなく、TogoID の機能に統一的にアクセスできるようになる。

<https://github.com/togoid/togoid-lib-r>
<https://github.com/togoid/togoid-lib-python>

PyPI / Bioconductor に登録予定

自然言語からの問い合わせ

TogoID は自然言語から生命科学データベースにアクセスするためのリモート MCP、TogoMCP に組み込まれており、WebUI から離れたチャット形式での利用も可能となっている。(TogoMCP の詳細は別ポスターで)

NCBI Gene の 1 と 9 は
Ensembl Gene では何？
TogoMCP を使って教えて

TogoMCP で NCBI Gene → Ensembl Gene の変換を実行しました。

まず、変換経路を確認すると、

* ncbigene → ensembl_gene
* 関係 : is nearly equivalent to

となっています。

変換結果は以下の通りです。

NCBI Gene	Ensembl Gene
1	ENSG00000121410
9	ENSG00000171428

遺伝子、タンパク質、立体構造、分子間相互作用、化合物、糖鎖、疾患、変異の 8 カテゴリに関連する 65 の属性を収録し、各データセットは TogoID から専門家が選択したネットワークを介して、お互いに紐づいている。

開発・検討中の属性

- TiSMed : 組織特異的メチル化と発現
- IDG Development Level Summary (Pharos) :
研究・創薬の進行
- Metabolite location (HMDB) : 代謝物の局在
- GRAVY score : タンパク質の疎水性
- Instability Index : タンパク質の安定性
etc.

TogoDX はこれまで、人が Web を操作してデータ探索を行うことを想定して開発してきた。昨今の AI 技術の発達を受け、LLM との対話による探索方法を模索するため、TogoDX MCP の開発を進めている。

MCP の導入

<https://dx.dbcls.jp/togodx/mcp>

各種 LLM で上記 URL に接続することで利用できる。また、ローカルMCPとしても運用できる。

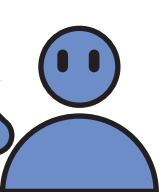
<https://github.com/togodx/togodx-mcp>

MCP の仕組み

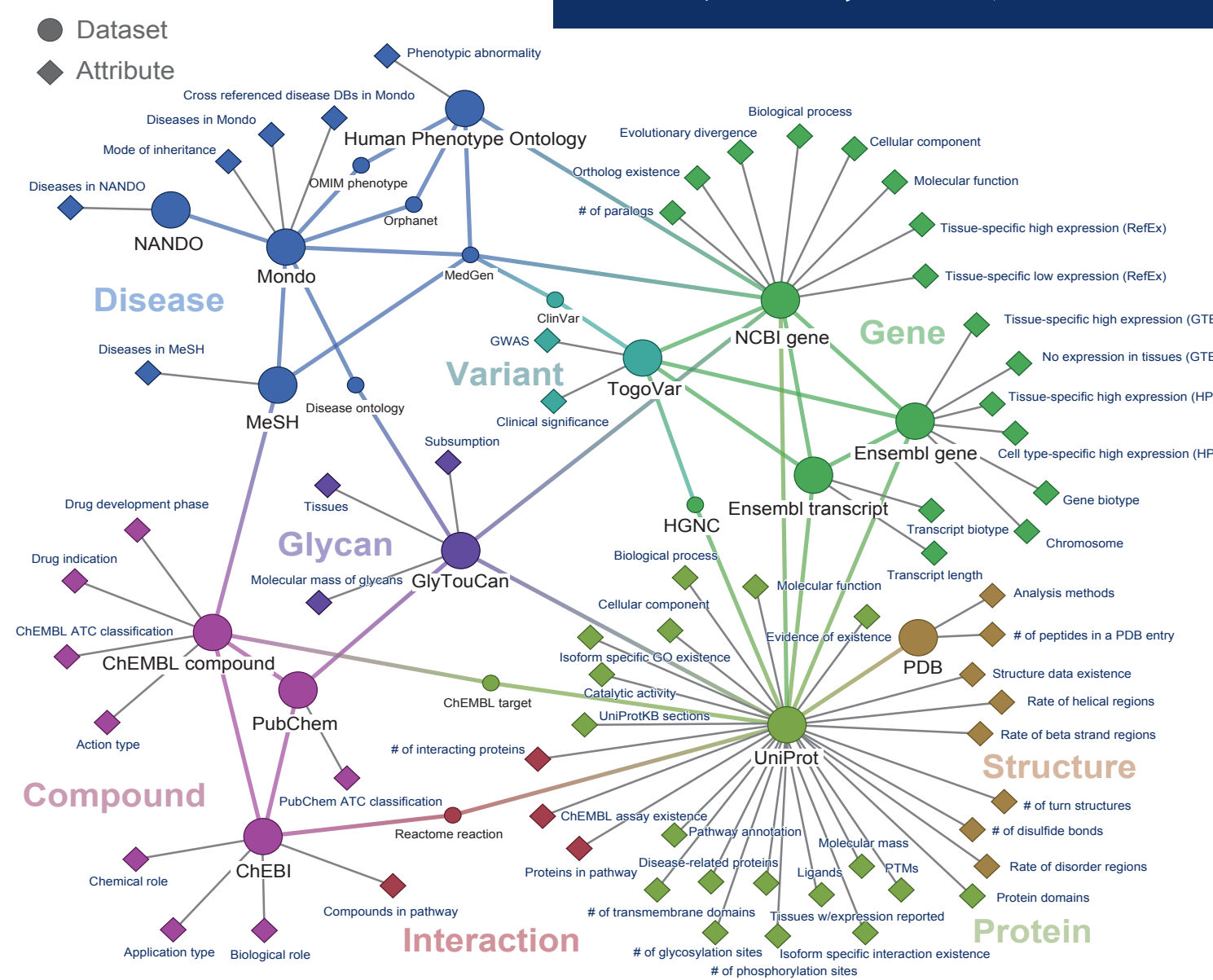
自然言語の問いから、属性選択・条件構築・結果取得・共有 URL 生成までを AI が実行。黄断や ID 変換を人が意識せず操作でき

問い合わせ

ツールを順次実行
属性検索
ノード解決
条件構築
結果取得
共有 URL 取得
結果の解釈と WebUI での共有



データセット

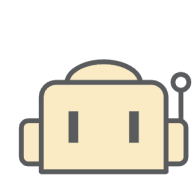


TogoDX MCPとユースケース

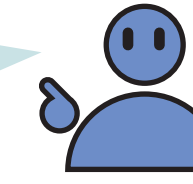
ユースケース：相分離 (LLPS) タンパク質の横断マッピング

*注釈： 液 - 液相分離 (liquid-liquid phase separation: LLPS) とは、水と油のドレッシングのように、細胞内の特定の分子 (タンパク質や RNA など) が集まって「濃い液体の部分」と「薄い液体の部分」に分かれる現象。バランスが崩れることによってタンパク質凝集や神経疾患につながるものが近年報告されている。

PhaSePro 由来のヒト LLPS ドライバー 55 タンパク質を、既存属性へ横断マッピング。新規データを足さず、近年報告の特徴が再現できるかを検証。



55 タンパク質のリストを渡すので、LLPS の特徴が見れる条件で探索して。



相互作用

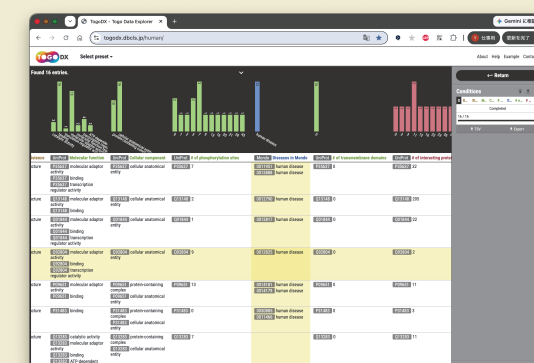
機能と局在

疾患

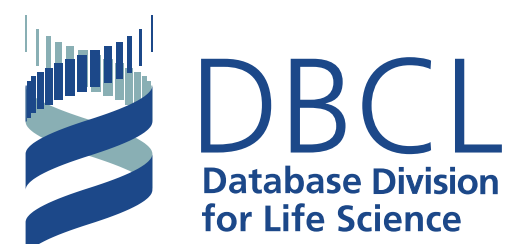
一部には多数のリン酸化部位を持つものや、多数の相互作用を有するハブタンパク質が含まれていた。これは、多価相互作用が LLPS 形成に重要であるという既報と整合する。

RNA 結合や DNA 結合タンパク質が多く含まれ、細胞内局在では核小体、ストレス顆粒、細胞質など、LLPS が形成されることが知られている細胞内構造に関連する属性が多数認められた。

疾患関連属性では Parkinsonism、Alzheimer disease など神経変性疾患との関連を持つタンパク質が多く含まれていた。



従来は複数のデータベースを個別に参照する必要があった情報を、一つのインターフェース上で俯瞰できることは、新たな生命現象を理解する際の探索効率向上に有用であると考えられる。



Database Division for Life Science (DBCLS)
BioData Science Initiative (BSI), National Institute of Genetics, Research Organization of Information and Systems
Univ. of Tokyo Kashiwanoha-campus Station Satellite 6F, 178-4-4, Wakashiba, Kashiwa, Chiba 277-0871, JAPAN

Licensed under CC-BY 4.0 ©2026 Yuki Moriya (DBCLS)

