

CLAVI：LLMエージェントを用いた意義不明バリエーションの文脈的解釈フレームワーク

矢野由絹 (1), 岡田愛 (1), 飯田直子(2), 千葉健一 (1), 白石友一 (1)

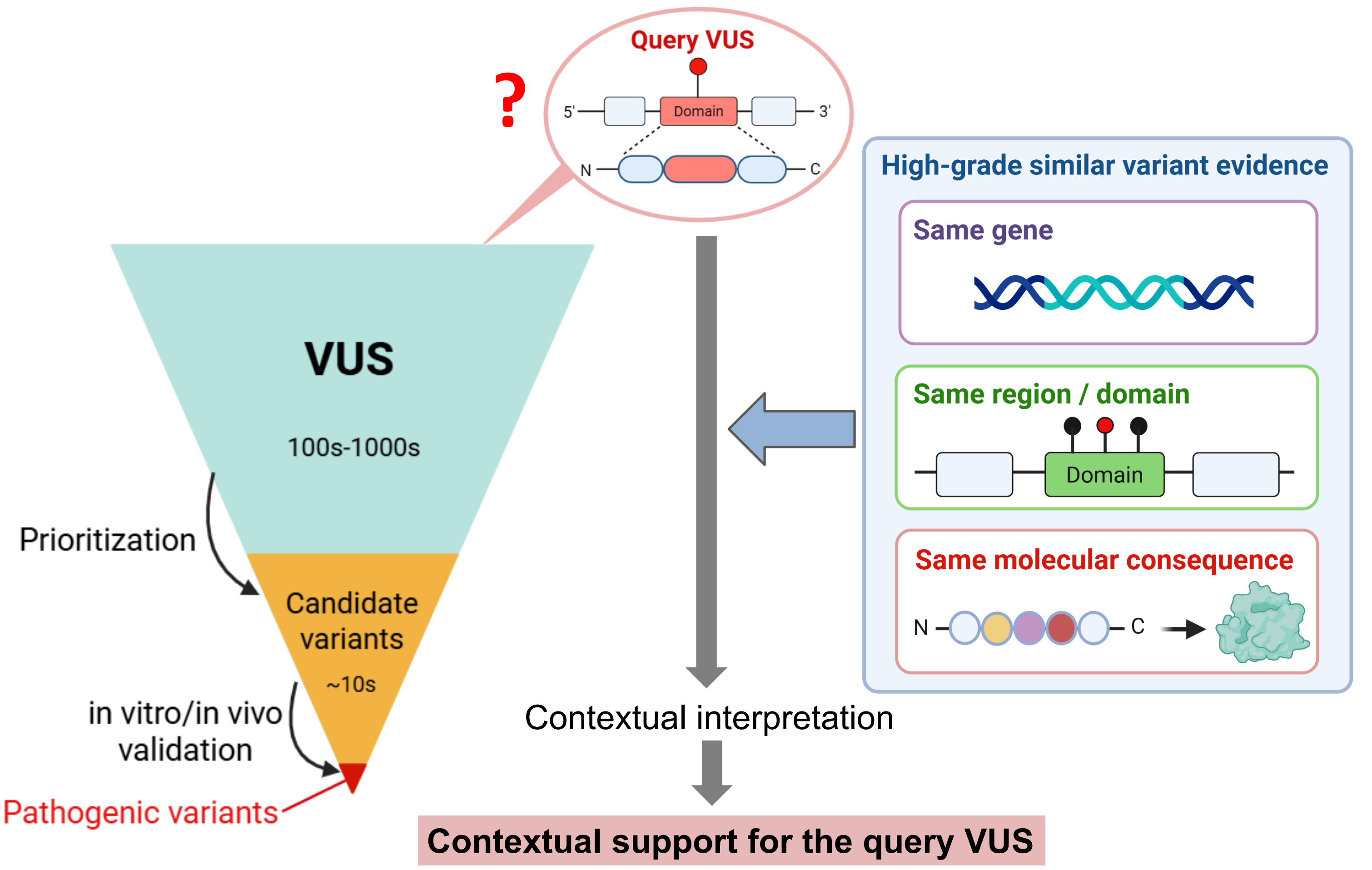
(1) 国立がん研究センター研究所 ゲノム解析基盤開発分野 (2)国立がん研究センター東病院 トランスレーショナルリサーチ支援室

Introduction

近年、遺伝性疾患が疑われる患者に対するゲノムシーケンスの適用拡大に伴い、スプライシングバリエーションを含む意義不明バリエーション（VUS）が同定される機会が急増しています。膨大な数のVUSの中から、どの変異を原因変異候補として優先し、in vitro/in vivoでの機能検証に進めるべきかを判断するためには、適切な優先順位付けが不可欠です。

しかし、多くのVUSでは、対象バリエーションに対する直接的な臨床的・機能的エビデンスが得られません。この場合、SpliceAIやCADDなどの予測スコアが参照されますが、これらは配列変化からの推定に留まり、より正確な病原性判断には機能的側面を含む解釈が必要です。同一遺伝子・同一ドメイン、あるいは同様の分子学的影響を持つ類似バリエーションの情報は、対象VUSの機能的影響を推測するうえで文脈的な支持情報となり得ます。

そこで我々は、VUSの機能的影響および疾患関連性に関する仮説生成を目的として、LLMエージェントに基づくフレームワークであるCLAVIを開発しました。



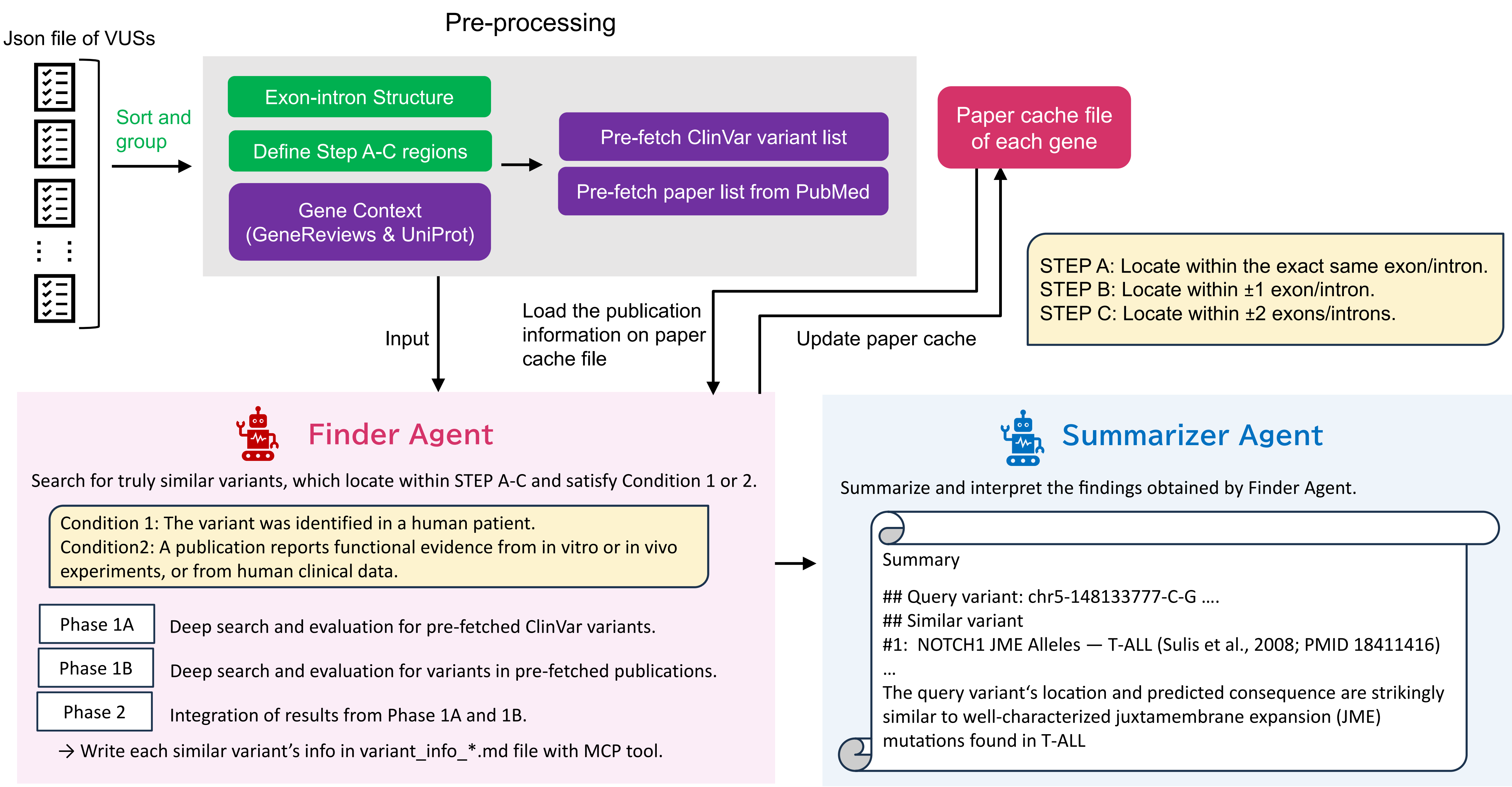
Workflow of CLAVI

フレームワーク

CLAVIは、対象VUS（クエリバリエーション）の類似バリエーションを収集するFinder Agentと、収集された類似バリエーションの情報を統合し解釈するSummarizer Agentの2つのエージェントによって構成されています。

Finder Agentは入力されたクエリバリエーションの情報に対して、文献およびバリエーションデータベースを横断的に検索します。対象VUSと同一遺伝子、近傍領域/ドメイン、または同様の分子学的影響を持つ既報バリエーションのうち、機能解析や患者報告によって支持されたものを、high-grade truly similar variant evidenceとして同定します。

Summarizer Agentは得られた情報を遺伝子・ドメイン情報と統合して要約し、対象VUSの潜在的な機能的影響、ヒト疾患との関連性を解釈します。

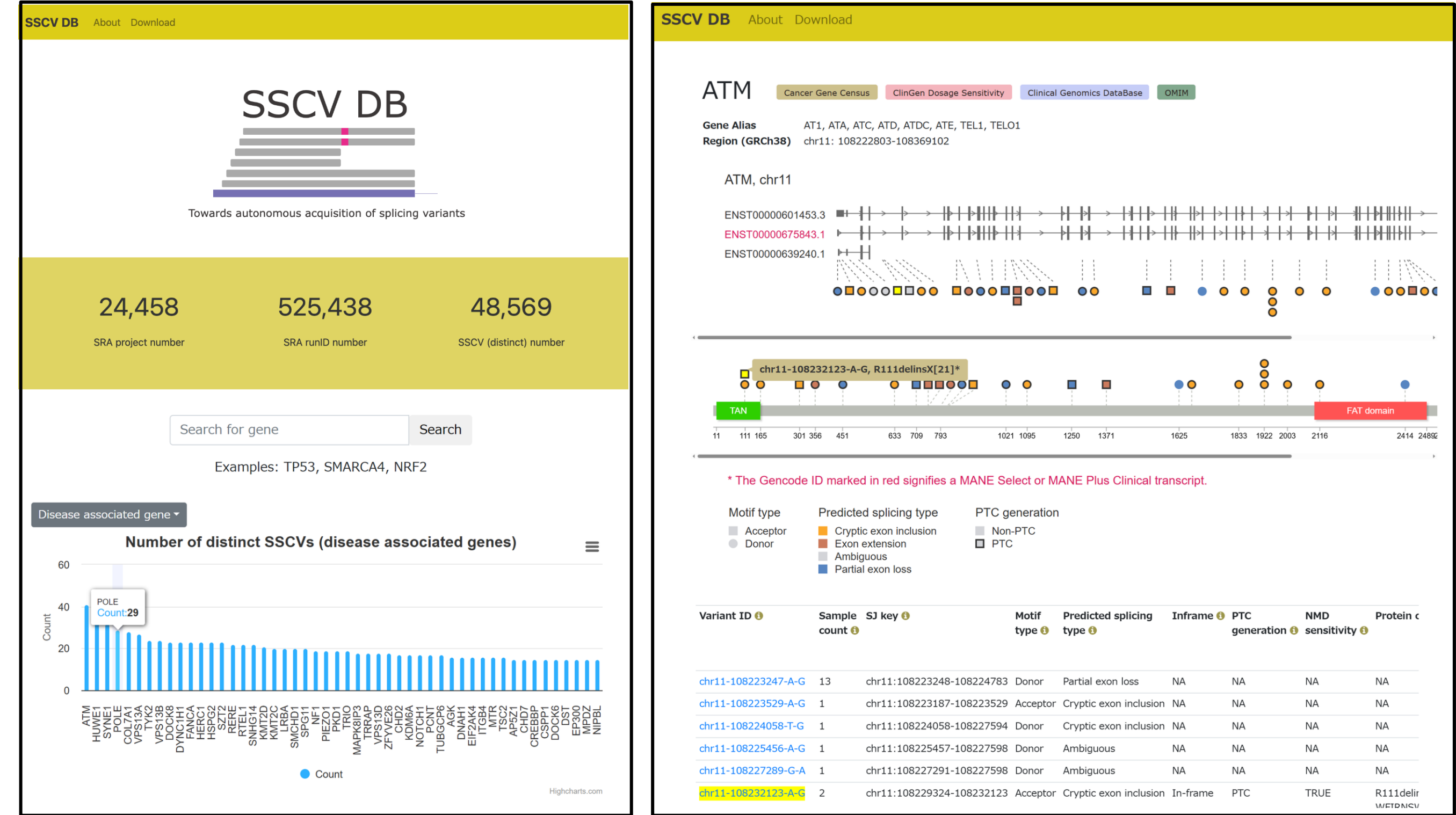


SSCV DB ポータルサイト

我々はトランスクリプトームデータのみを用いてスプライスサイト生成変異(SSCV)を同定することができるアルゴリズム「juncmut」を開発しました。(Iida et al., Nat Commun., 2025)

現在までに約 52 万件のトランスクリプトームデータを juncmut で解析し、約 4 万 8 千件のスプライスサイト生成変異（SSCV）を同定し、カタログ化しています。

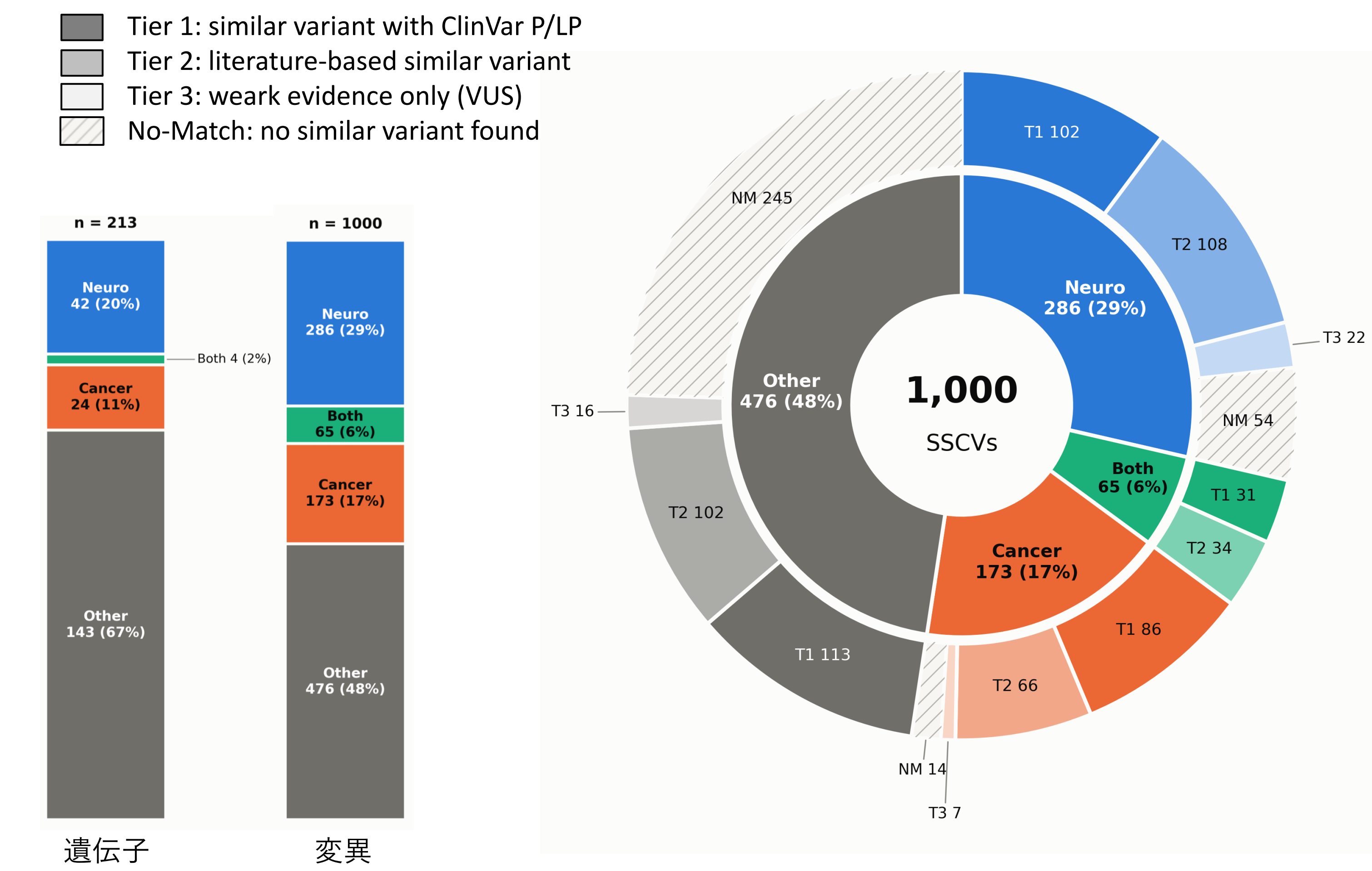
<https://sscvdb.io>



引用文献

Iida et al., Systematically developing a registry of splice-site creating variants utilizing massive publicly available transcriptome sequence data, Nat Commun., 2025 Jan 9;16(1):426 .

SSCV DBに搭載されている1,000 SSCVのアノテーション



SSCV DBに収載された1,000件のSSCV (213遺伝子；神経疾患関連 42 遺伝子、がん関連 24 遺伝子、がん・神経疾患関連 4 遺伝子、その他143遺伝子)を対象に、CLAVIを用いて機能的影響、疾患関連性に関するアノテーションを行った。

結果として、687件（68.7%）で類似バリエーションが同定された。このうち、332件ではClinVarでPathogenic/Likely pathogenic（P/LP）と分類されている類似バリエーションが、310件は文献報告に基づく類似バリエーションが同定された。神経疾患・がん関連遺伝子では、その他の遺伝子と比較して類似バリエーションが同定される割合が高い傾向を認めた。