

TogoVarの拡充: 構造多型・JGA寄託データ由来バリエント・スプライスサイト生成バリエント

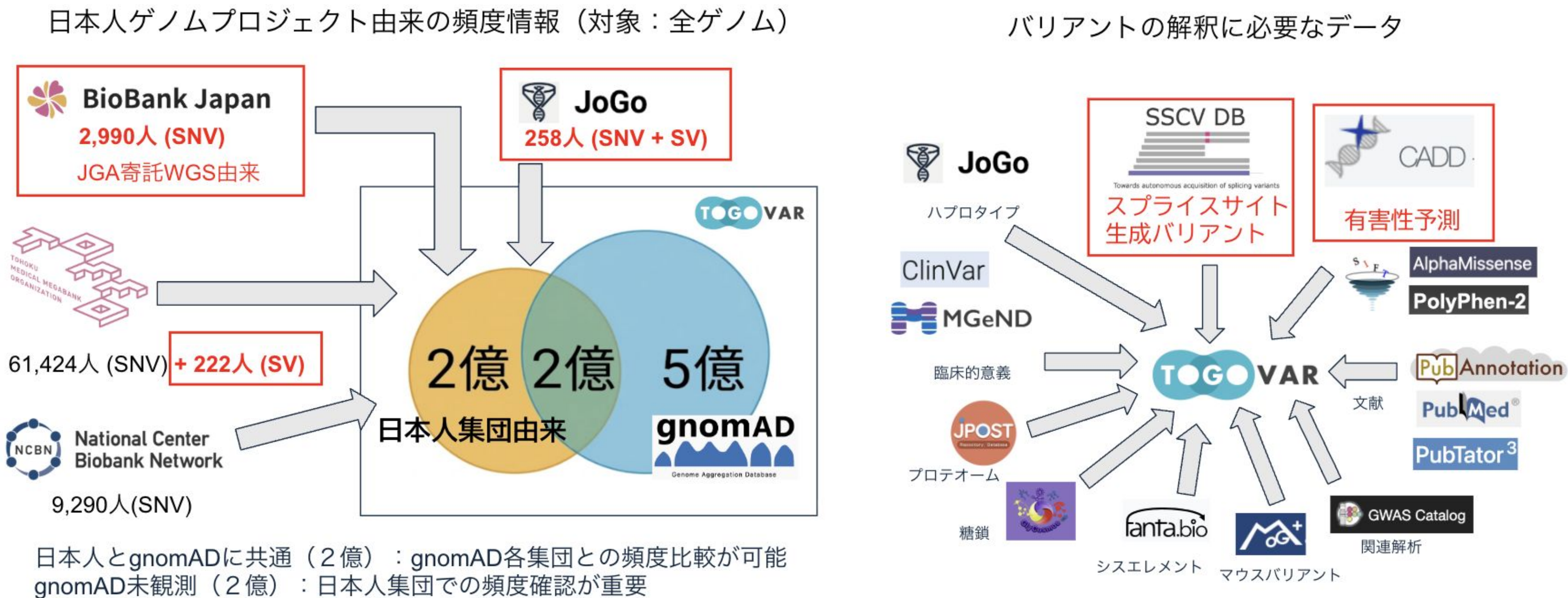


三橋 信孝、守屋 勇樹、川嶋 実苗、細田 正恵、川島 秀一、片山 俊明
国立遺伝学研究所 バイオデータ研究拠点 (BSI) ライフサイエンス統合データベース部門 (DBCLS)

TogoVarは日本人集団を中心とするバリエント頻度と臨床的意義・機能予測・文献情報などを統合し、バリエントの解釈を支援するデータベース

TogoVar 2026.1: 日本人集団の頻度・構造多型・解釈情報の拡充
(JoGo／ToMMo JSV1、JGA寄託WGS、SSCV DB・CADDを統合)

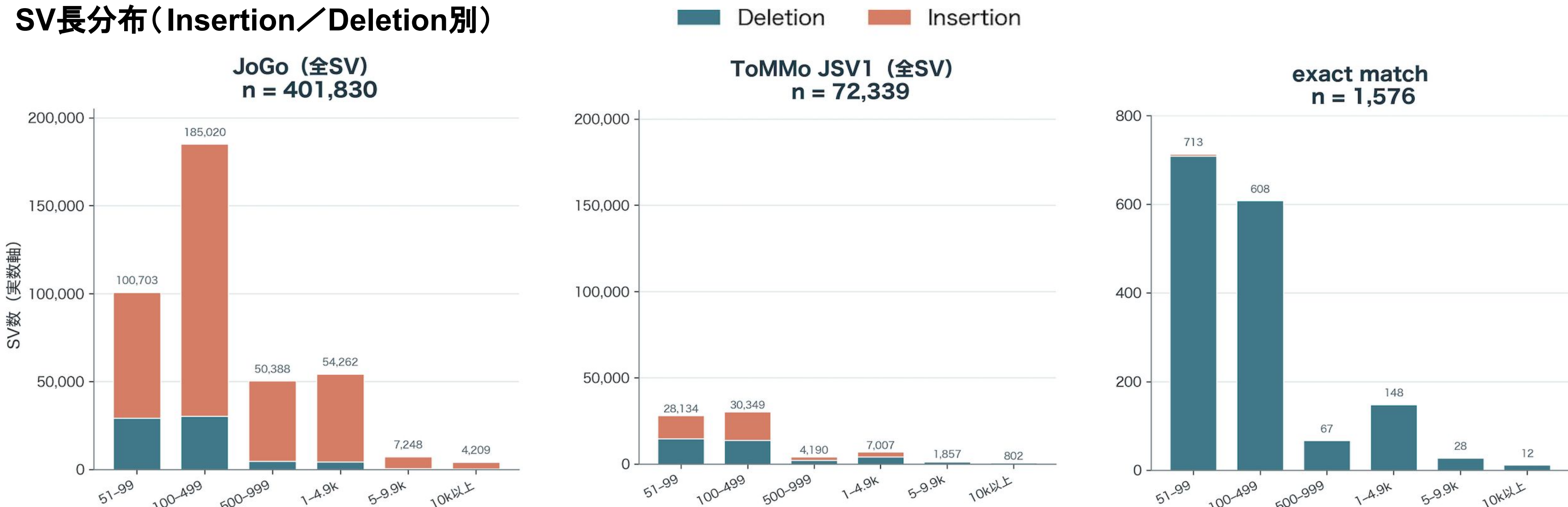
赤字・赤枠: 2026.1 リリースで追加



構造多型 (structural variant, SV) の追加

ToMMo JSV1 (222人)とJoGo(258人)の構造多型(SV)頻度を追加した。両データセットのSVを個別バリエントとして収載し、CHROM_NO + POS + REF + ALTが完全に一致する場合(exact match)に同一SVとして統合した。なお、51 bp以上をSVと定義した。

SV長分布 (Insertion／Deletion別)



- 高精度なロングリードから決定したハプロタイプ配列からバリエントを検出しているJoGoでは多くのInsertionを検出している。
- 両データセット間で完全一致したSVは1,576件で、Deletionが1,572件、Insertionが4件であった。Insertionは挿入配列の完全一致を条件としたため、完全一致として統合されにくい。

課題

- 同一SVの判定: Deletionなどにはreciprocal overlap、Insertionにはブレイクポイントの近接性などのSV型に応じた統合方法を検討する。ゲノムグラフを用いた手法も検討する。
- 解釈情報の拡充: AnnotSV等の活用を検討し、遺伝子・エクソン・調節領域への影響、疾患との関連などのアノテーションを整備する。

JGA (Japanese Genotype-phenotype Archive) 寄託WGSデータ由来のアレル頻度を追加

NBDCヒトデータベースがGATK Best Practicesに基づいて解析したJGA寄託WGSの加工済みデータを利用した。BBJ1K(1,026人)とBBJ2K(1,964人)、計2,990人のアレル頻度とジェノタイプ数を収載し、旧リリースと比較して少なくとも174万件を新規収載した。

加工前のWGSデータ (JGAD000220) 利用可能 (要: 利用申請)

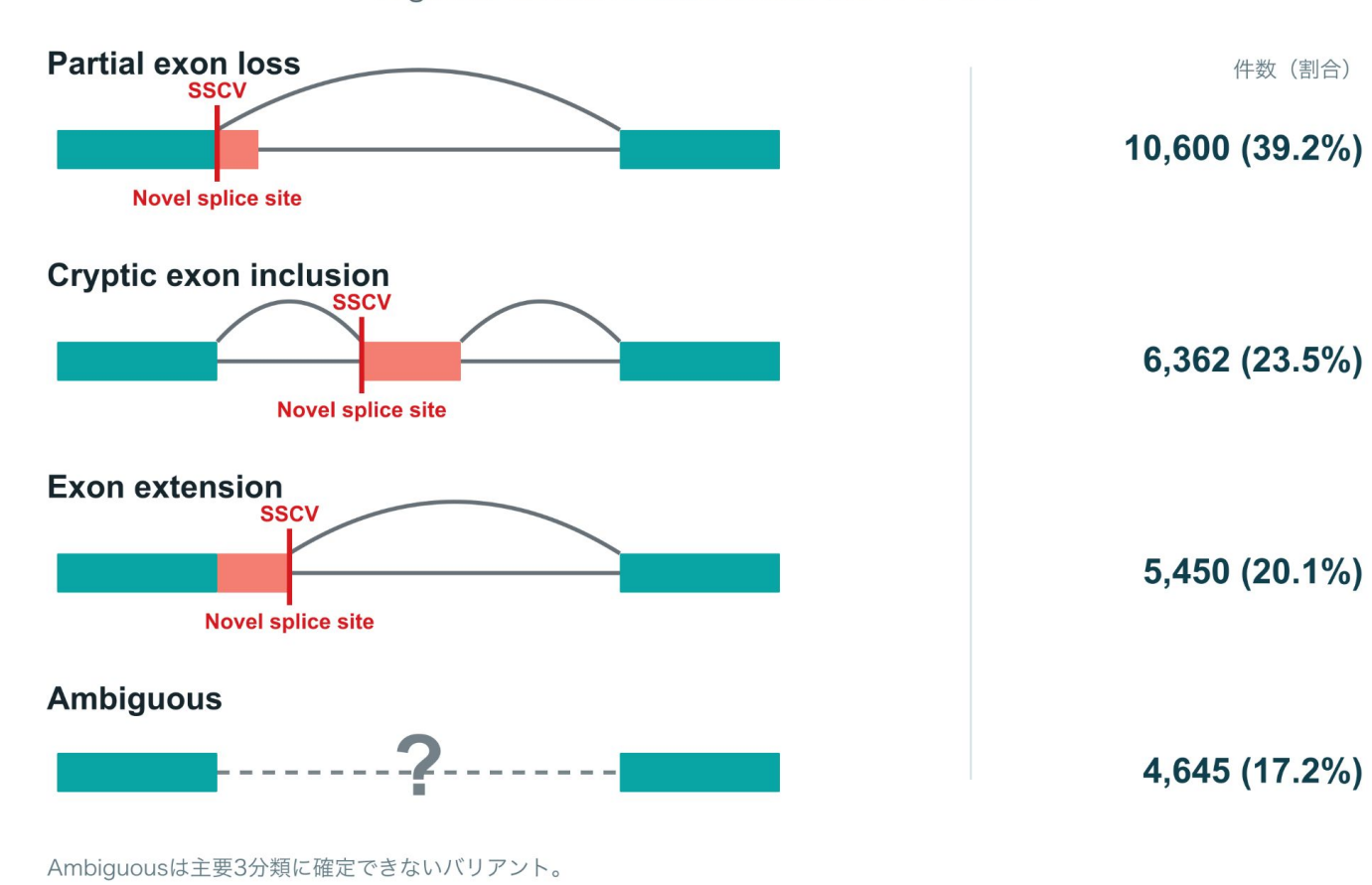


<https://humandbs.dbcls.jp/dataset/JGAD000220>

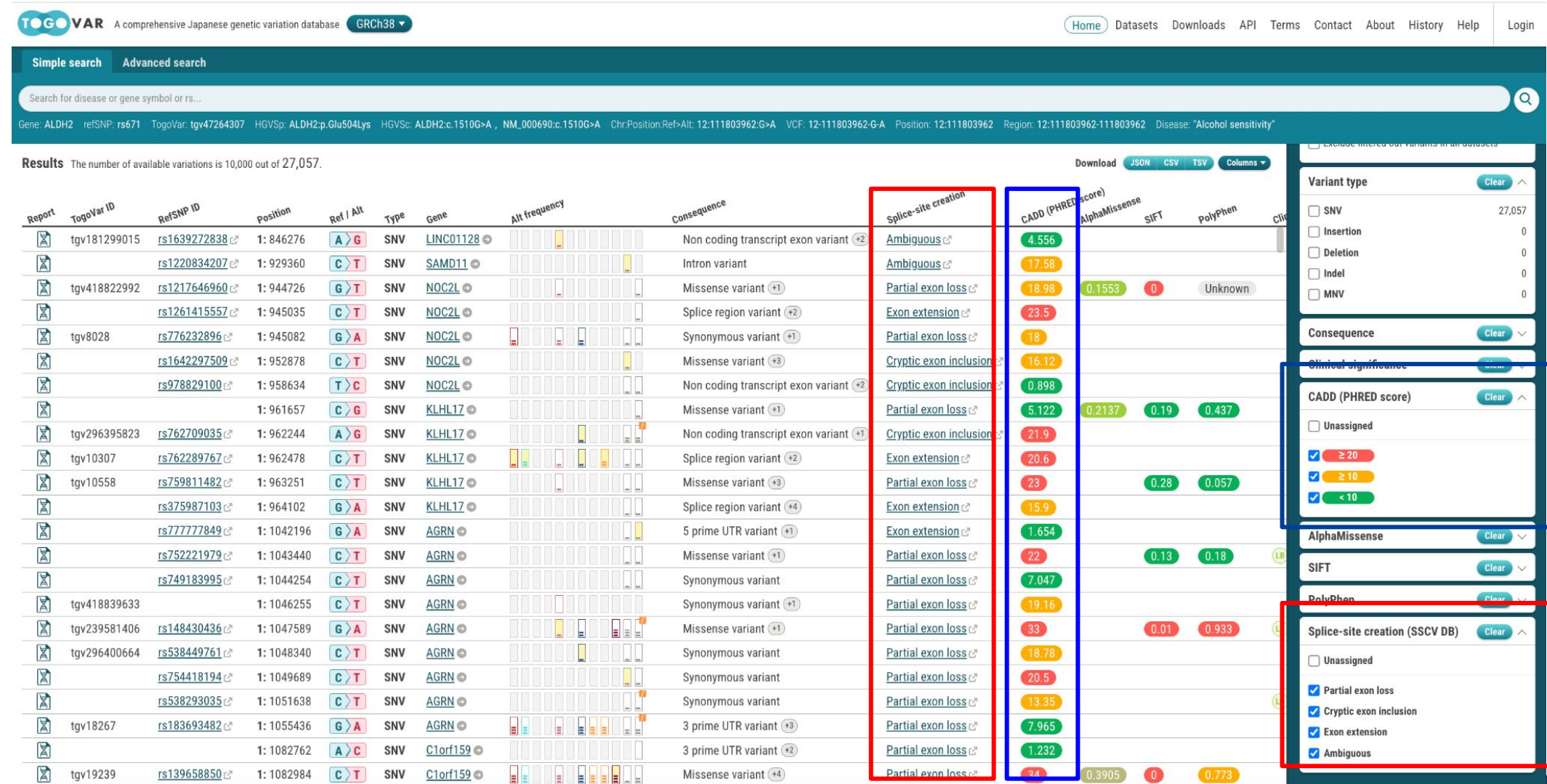
スプライスサイト生成バリエントと CADDスコアの追加

SSCV DB v1.1 (Iida et al.) が公開RNA-seqデータから同定したスプライスサイト生成バリエントを統合した。48,569件のうち27,057件 (すべてSNV) がTogoVar収載バリエントと一致し、主要3分類およびAmbiguousのスプライシング影響情報を付与した。さらに、非コーディング領域を含むバリエントの有害性予測情報として、CADD PHREDスコアを追加した。

SSCV DBのスプライシング影響分類



Iida et al., Nat Commun. 2025;16:426.



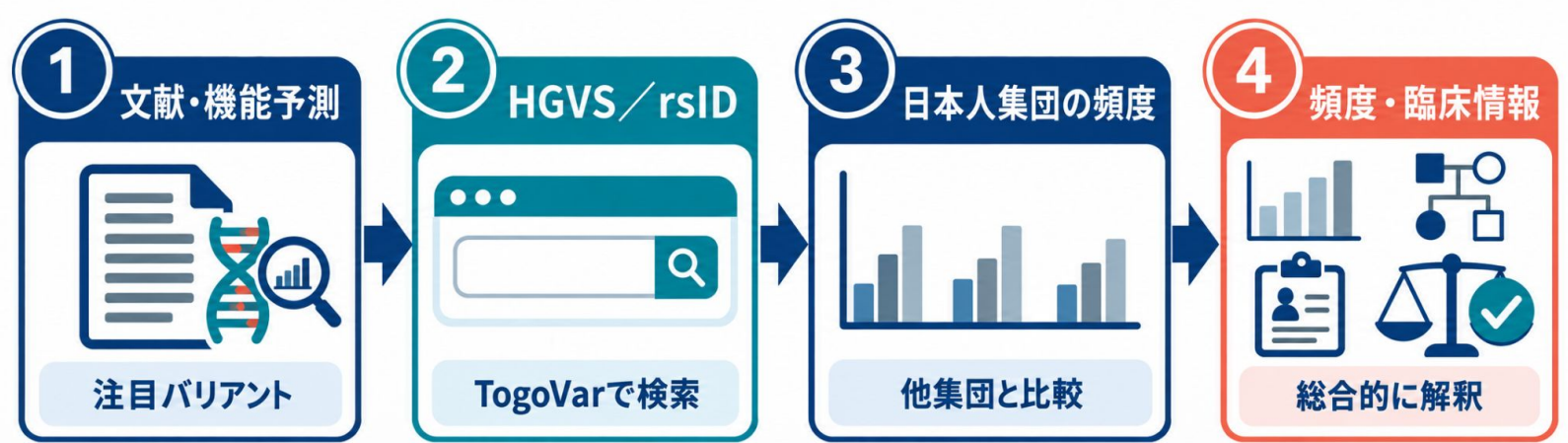
ファセット検索

TogoTVで新動画を公開予定

ポスターNo. 24: TogoTV feat. TogoVar

研究現場でのユースケースを紹介

TOGO TV



動画制作を通じて得られた利用者視点をTogoVarの使い勝手の改善や収載データの検討に反映する。

Licensed under CC-BY 4.0 ©2026 Nobutaka Mitsuhashi (DBCLS)

