

次世代シーケンサ Web 解析システム Exome SNV screening system

多検体の SNV 情報から候補遺伝子を検出

コールされた多検体の SNV 情報をリレーショナルデータベースに格納し、種々の絞り込み条件を Web から設定、検索を行うことで、表現型と関連する候補遺伝子のリストアップを可能とします。

Exome SNV screening system システムの特長

・SNV データ（VCF 形式）を RDB にインポートします

コールされた SNV のデータを RDB にインポートし、その後、アノテーション付与、絞り込み検索のための前処理を行います。何種類かの VCF 形式に対応していますが、必要な VCF 形式で未対応であれば対応します。

・重複のないバリエーションデータを作成

インポートの際に、ひとつの SNV コールが複数のバリエーション情報を含む場合（アレルのどちらかがリファレンスと異なる、INDEL コールが複数のバリエーションの組み合わせである等）、これを別々のバリエーションとして、システム内のバリエーションテーブルに格納します。また、異なるサンプルで同じバリエーションが存在する場合は、バリエーションテーブルにはひとつのバリエーションとして登録し、重複のないバリエーションデータを生成します。

・厳選されたアノテーション情報

重複のないバリエーションについて、アノテーションを付与します。アミノ酸置換の有無を含む遺伝子との関係情報 1 種類、1000genome のバリエーション情報（人種ごとの頻度、JPT の頻度は独自で計算）、UCSC の dbSNP 情報が付与され、絞り込み検索条件として使用できます。

・検索対象サンプルのグループ化と検索条件の保存、さらに、SNV の集合演算が可能

インポートされた複数のサンプルから、絞り込み対象のサンプル群を選択し定義します。サンプル群に対し、種々の条件を設定し SNV 単位・遺伝子単位で SNV を絞り込みます。サンプルグループ内での頻度を条件にすることもできます。さらに、解析対象サンプル間の集合演算（積、和、引き算、有意差集合）も Web の簡単な操作だけで可能です。

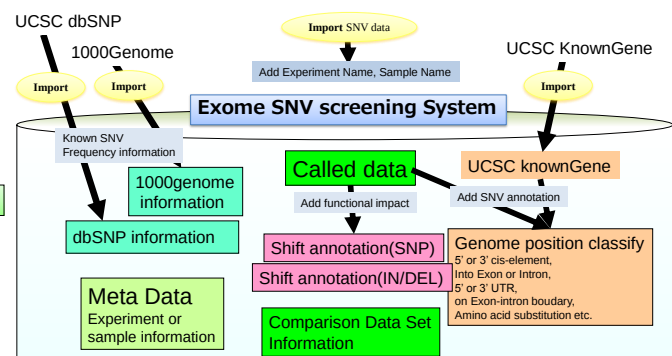
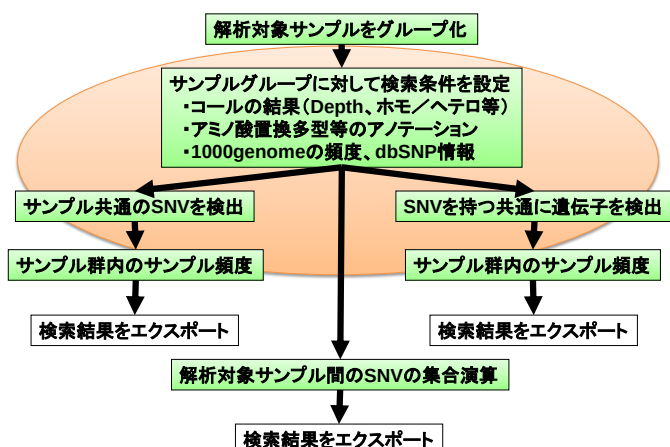


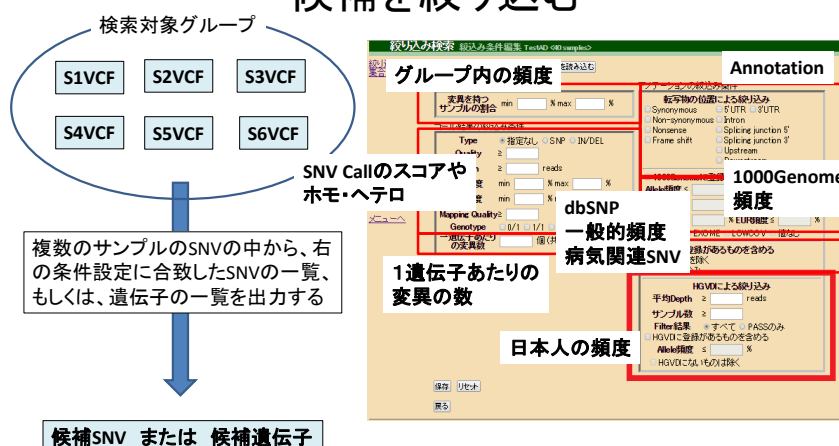
図 1：絞り込みシステムの概要

図 2：絞り込み検索の手順

利用ソフトウェア SIFT

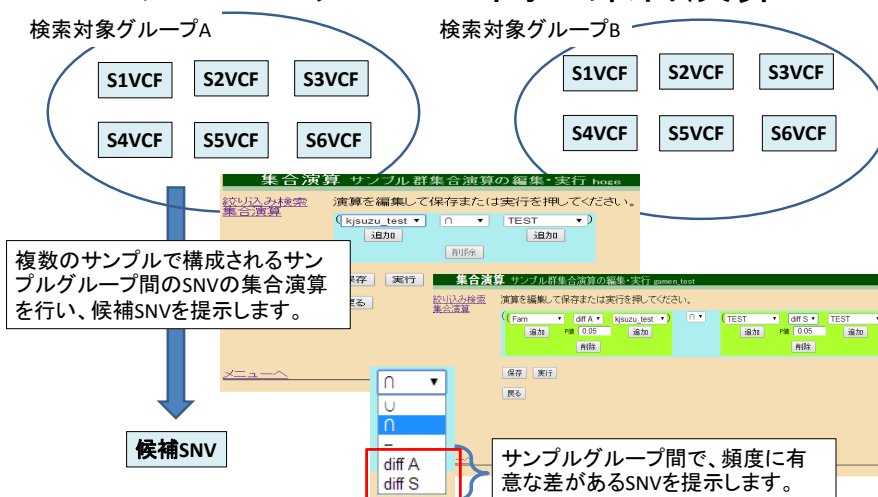
利用 DB dbSNP, 1000genome

複数サンプルのデータから 候補を絞り込む



18

SNV集合間の演算(1) サンプルグループ間の集合演算



SNV集合間の演算(2) 家系のサンプルの集合演算

