

研究開発課題別中間評価結果

➤ 課題情報

研究開発課題名 「ヒトゲノム・病原体ゲノムと疾患・医薬品をつなぐ統合データベース」

研究代表名 金久 實

➤ 中間評価結果：

ゲノムと疾患や医薬品を分子ネットワークで統合した「KEGG MEDICUS」を拡張し、ヒトゲノムおよびウイルスその他の病原体ゲノムの情報を社会で活用するための基盤データベースとして高度化するとともに、より多くの疾患とネットワークの対応付け、収録データの高品質化を行い、いまだ機能が解明されていない遺伝子・タンパク質を多く含むウイルスゲノムについて知識を広く体系化する研究開発課題である。

疾患遺伝子について調査し、DISEASE データベースのネットワークバリエーションマップのネットワーク要素として定義した。2024 年 8 月末時点で登録した DISEASE エントリ数は、研究開発開始時点の 175 から 819 に増加し、新規に定義したネットワーク要素数はバリエーションマップのみの要素も含めて 1,300 ほどになった。NETWORK データベースと DRUG データベースとの統合利用を前提として DISEASE データベースの構成と内容を見直し、医薬品とネットワークの関連について、ヒト遺伝子バリエーションを標的またはマーカーとするものに対象を限定していたものを、ネットワーク上で関連があれば通常遺伝子でも医薬品・標的関係を定義し、ネットワークバリエーションマップに表示できるように再編成した。

ウイルスタンパク質とヒトタンパク質の相互作用についてヒト生体内システムの観点からデータ収集し、主に自然免疫系を中心にヒトの免疫系に関するネットワークバリエーションマップを 8 個作成するとともに、ウイルスその他の病原体およびヒト遺伝子バリエーションが関与する疾患を 134 個データベースに収載した。ウイルスオーソログクラスターをウイルスオーソロググループ（VOG）と名称変更し、配列類似性と系統関係の計算処理で作成した VOG をキーワード検索ツールとともに公開した。また、VOG クラスターの遺伝子の並びをウイルス間で調べるツールとウイルスと生物種の間で調べるツールを公開した。新たな展開として、当初、KEGG の 68 万個のウイルスタンパク質から計算によりオーソロググループを定義していた VOG を、KEGG 生物種全体 5000 万個のタンパク質に広げた。さらにこれらに属するものを調べたデータセットを定期的に作成し、遺伝子の並びを比較するアルゴリズムを用いて、ウイルスと宿主あるいはその他の生物種で共通に存在する遺伝子クラスターを調べることができるようにした。

医薬品添付文書サービスに関し、商品一覧の利用件数が 2022 年度 4 月の研究開発開始時点の 2.5 倍に達した。

上述の通り、進捗状況と今後の成果見込みは優れている。KEGG MEDICUS は国際的にも広く認知され、研究者以外の一般利用者も含めて非常に多くの人に使われている社会的にも重要な情報基盤となっている。研究計画は適切であり、今後も概ね現計画通りに推進すべきである。今後も KEGG NETWORK の充実を着実に進めつつ KEGG MEDICUS の有用性を向上させ、生命科学分野におけるさらなる波及効果の拡大を期待する。

以上