

研究開発課題別事後評価結果

▶ 課題情報

| | |
|---------|---------------------------------------|
| 研究開発課題名 | 疾患ヒトゲノム変異の生物学的機能注釈を目指した多階層オーミクスデータの統合 |
| 研究代表名 | 菅野 純夫（東京医科歯科大学 難治疾患研究所 非常勤講師） |

▶ 事後評価結果

| | |
|---|---|
| 1. 総合評価 | <p>当初の実実施計画はほぼ達成され、一定の成果が出された。</p> <p>測定技術が急速に発展し研究ニーズも大きく変化する中で、可視化から分析機能までを実装したサイト構築を目指す野心的なプロジェクトだった。シングルセルやロングリードなど、新しい技術で取得されたオーミクスデータのビューワが提供され、それらのデータからどのような知識が得られるかということを研究者がイメージできるようになった点は評価できる。</p> <p>本データベースが対象とする研究分野は研究者や潜在的ユーザーが多いことに鑑みると、データの提示方法の工夫やデータベースの活用例の創出が利用促進につながると思われる。企業との共同研究の成果はインパクトがあるものと思われ、特許公開等の将来の適切なタイミングで公開していく必要がある。</p> |
| 2. 研究コミュニティを含むデータ提供者やDB利用者との連携・協業 | <p>Nucleic Acids Research 誌に DBKERO の論文発表をするなど、研究コミュニティに対して適切に周知された点を評価する。国内の大型プロジェクトや先進ゲノム支援等で生み出されたデータを公開し、その成果を広く共有するする仕組みとして機能した。講習会や学会でのワークショップが実施され、さまざまな研究機関や企業との連携に繋がった点を評価する。</p> |
| 3. 利用者にとって有用なデータ基盤、持続的なDB運用体制構築に向けた取り組み | <p>ロングリード、シングルセル、空間トランスクリプトームといった新しい技術で取得されたデータも含め、さまざまなオーミクスデータの可視化基盤が提供され、国内のゲノム関連プロジェクトのデータが掲載された点を評価する。</p> |
| 4. 生命科学研究所やイノベーションへの波及効果（DBを利用して得られた研究成 | <p>5,000件を超える多層オーミクスデータセットが DBKERO に収載された。複数企業との共同研究や外部プロジェクトでのデータ活用等、一定の波及効果があったものと推察される。</p> |

| | |
|---------------|---|
| 果) | |
| 5. その他特記すべき事項 | 共同研究先の企業から大学院生を受け入れる等により、人材育成に貢献した点を評価する。 |
| 6. 実施計画の達成状況 | <p>DBKERO を研究開発対象とし、1) ヒトおよびモデル生物の多階層オーミクスデータの収集と加工、2) 他のデータベースとの統合、3) 新規解析技術で取得されたデータを表示するブラウザの開発、4) これらのデータの RDF 化、5) ダウンロードポータルの開発を目的として、研究開発が実施された。</p> <p>1)については、配列情報が公開された CREST IHEC 日本チームのエピゲノム情報ならびに Bio Bank Japan の GWAS データ、国立国際医療研究センターが取得した HGVDB のデータ、東北メディカルメガバンクが取得した公開可能なデータ、国立遺伝学研究所人類遺伝研究室が取得した HLA 領域を含む多型データおよび国立がん研究センターが取得したがん細胞への薬剤投与による摂動データが収録された。また先進ゲノム支援で取得されたマウス等のモデル生物種の多層オーミクスデータが収録された。以上を含め、合計 5,510 件の多層オーミクスデータセットが収録された。</p> <p>2)については、国立がん研究センターと東京大学が取得したロングリードシーケンシ、シングルセル RNA-seq データ、空間トランスクリプトーム解析データ等を模擬データセットとして、これらを表示するブラウザ開発が実施された。また健康人末梢血シングルセル発現データの個人差、個人の日動変動を閲覧できるようにするブラウザを開発し、研究期間終了までに 171 ライブラリーのデータが公開される予定となっている。</p> <p>3)については、ChIP-Atlas のピークデータが表示できるように開発が行われた。</p> <p>4)については、トランスクリプトームとエピゲノムのデータが RDF 化され、NBDC RDF ポータルに登録された。また先進ゲノム支援関連データを模擬データとしてメタデータの RDF 化の枠組みが構築された。</p> <p>5)については、二次配布可能なデータとして、DBKERO 収録データの 47%にあたる 2,573 件がダウンロードポータルに収録された。また ICGC、TCGA、ENCODE 等のデータ 2,723 件（同 49%）は元データのリンクサイトが明示された。</p> <p>上記の通り、当初の実施計画はほぼ達成された。</p> |