

## 電子カルテ情報を用いた証拠性のある臨床研究手法に関する研究

(H27-医療-指定-016)

- 測定、画像データ証拠性・安全運用環境検討 -

遺伝子解析研究への証拠性の導入

研究分担者 作佐部 太也

藤田保健衛生大学医療科学部臨床工学科 准教授

**研究要旨:** **目的:** 遺伝子解析研究における証拠性を確保する上でのワークフロー上のチェックポイントを見だし、適切な方法を提案することを目的とする。**方法:** 次世代シーケンサーを運用している研究現場においてワークフローについての聞き取り調査を行う。**結果:** ワークフローの概要は判明したが手作業が多く体系化したものではなく、情報システムによる支援が必要なことが明らかになった。**結論:** 証拠性確保のためのチェックポイントを見出すためには、ワークフロー運用を支援する情報システムを開発することでワークフローの体系化を行う必要があることがわかった。

### A.研究目的

今日において遺伝子解析研究における次世代シーケンサー（NGS）は、基礎研究を超えて臨床研究のツールとなりつつある。しかし、基礎研究のツールとしての十分な運用経験の蓄積による成熟期間を経ていないため、データの管理方法などにおいて未成熟な状態であることが想定される。このことを別の視点からとらえると、証拠性に留意したデータ管理方法を広く普及させるには、むしろ望ましいタイミングであると考えられる。本分担研究において、医療機関の付属研究施設における実際のNGSの運用やデータ解析処理の状況について調査分析し、証拠性確保のためのワークフロー上のチェックポイントを見だし、また、適切な方法を提案することを目的とする。

### B.研究方法

分担研究者が所属する藤田保健衛生大学に付属する総合医科学研究所の遺伝子解析関連の研究部門におけるNGSの運用状況について聞き取りなどの調査を行い遺伝子解析の基本的なワークフローのパターンを明らかにする。次にワークフローを支援するようなデータ管理システムの開発し、研究所だけでなく大学病院の臨床研究者に利用してもらうことで、より一般的なワークフローのパターンを明らかにする。その上で、過誤によるデータに対する誤った操作が発生する可能性や悪意に基づいた改竄などの入る余地を分析し、適切なチェックポイントおよび証拠性を確保する方式を検討する。

### 倫理的配慮

平成27年度の分担研究においては、特に個人情報を取扱うなどの倫理的な課題は発

生しなかった。

### C.研究結果

聞き取り調査によりデータ管理の体制などの情報が得られた。

まず人的体制については、計算機技術を独学で学んだ若手の生物系研究員が一人でデータ管理および基本的な解析処理を行っていた。このような状況は大規模な研究機関以外では一般的な状況であることもわかった。データの解析処理についてはNGSから出力されるデータに対してサーバ規模の計算機上で各種処理コマンドを研究員が手動で動作させていた。

解析結果のデータを用いて複数の研究者が探索的な研究を行っているが、研究プロジェクトのワークフローの管理は、表計算ソフトウェア、SNS、電子メールなどのプリミティブな情報システムを組み合わせで行っていた。データ管理については、全てのデータはファイル形態でNASに蓄積され、研究員が開発した簡易なWebアプリケーションで管理を行っていた。

データ処理のワークフローや研究プロジェクトのワークフローの管理についてはほとんどが手作業となっており、情報システム化が望まれていることがわかった。このような状況は一般的なものと考えられ、実際、NGSを開発・販売しているメーカーや大手の研究機関などが解析ワークフローおよびデータの管理を行うクラウド・サービスを提供している。しかし、患者のデータを扱う臨床研究においては、患者からの承諾の取得から研究倫理審査の社会的プロセスの整備だけ

でなく、そもそも学外へのヒトゲノム情報の転送の可否についての運営的な方針の整備が必要となる。実際、本学においてはそのような整備は十分ではないため、クラウド・サービスを利用することはできず、他の医療機関においても同様であると推測される。

また、検査センターなどのように定型業務としてNGSをランさせているわけではなく、その後の解析処理を含め、研究プロジェクトなどに応じてプロセスをその都度変更するため、自動化やアウトソーシング自体が難しい。

一方、情報処理技術的な観点からみると、データ量についてはNGSから出力となるFASTQ形式のファイルが1GB程度、その後の解析マップ処理などの出力となるBAM形式のファイルが5GB程度。リファレンスに対する変異の抽出した結果のVCF形式のファイルが数100MB程度と比較的サイズの大きいファイルが扱われてはいるが、発生するデータ数は多くはなく、今日のMDCTによる放射線診断画像と比べれば処理が困難になるほどの量ではない。

### D.考察

聞き取り前の想定通り、データ管理の方式は洗練されたものではなく改善の余地があった。何より手動でコマンドを実行することが多く、動作の記録や再現性の確保などに不安があった。よって、解析処理のワークフローを支援する情報システムが必要と考えられる。

一方、研究プロジェクトの管理として、対象患者の選択から検体の採取、管理、

数理的な解析や探索的な分析、研究の指揮や管理などのワークフローについても未整備な状況であり、やはり研究のワークフローを支援する情報システムが必要と考えられる。

今後は聞き取り調査の段階から協同での作業の段階へと作業を進め、ワークフローの詳細な洗い出しと、そのワークフローの運用を支援するプロトタイプ・システムの開発を行う。情報システムとしては二種類必要となり、一つはNGSの出力から臨床研究者が探索的分析を行うことができるデータに処理する計算機上でのプロセスを支援するシステム、もう一つはより大きな視点から研究プロジェクトの中での探索的分析のワークフローをデータ中心に管理するシステムが考えられる。

このような半自動化したシステムを開発し運用することで、詳細な動作記録を自動的に採取することができるため、証拠性確保のためのチェックポイントを見出すための各種の情報が得られることが期待される。なお、開発するワークフロー支援システムについては諸事情が許せばオープンソースとして配布することを計画しており、他の研究機関で試用してもらうことで、より本学固有の事情に左右されない、一般的な情報が収集できることも期待される。また、ワークフロー支援システムがNGSを用いた臨床研究の効率化、活性化に寄与することも期待

される。

## E. 結論

実際の研究機関におけるNGSおよびデータの運用状況について調査より、改善の余地があることが明らかになった。

証拠性の確保の前段階としてデータ処理や研究プロジェクト運営についてのワークフローの詳細な分析と、ワークフローの運用を支援する情報システムの導入が必要であることが明らかになった。

## F. 健康危険情報

平成27年度の本研究においては、生命、健康に重大な影響を及ぼすと考えられる新たな問題、情報は取り扱わなかった。

## G. 研究発表

### 1. 論文発表

なし

### 2. 学会発表

なし

## H. 知的財産権の出願・登録状況

### (予定も含む)

- |           |    |
|-----------|----|
| 1. 特許取得   | なし |
| 2. 実用新案登録 | なし |
| 3. その他    | なし |