

厚生科学研究費補助金（がん対策推進総合研究事業）

分担研究年度終了報告

遺伝子パネル検査による遺伝子プロファイリングに基づく

複数の標的治療に関する患者申出療養に関する研究

山本 昇 国立がん研究センター中央病院 先端医療科・科長

分担研究者名

柴田龍弘 国立がん研究センター研究所・がんゲノミクス研究分野・分野長

間野博行 国立がん研究センター研究所・細胞情報学分野・分野長

白石友一 国立がん研究センター研究所・ゲノム解析基盤開発分野・分野長

平田 真 国立がん研究センター中央病院・遺伝子診療部門・医員

研究要旨

がん全ゲノム解析計画のための体制構築として、1. 解析検体・臨床情報収集の検討、2. データ解析ワークフロー、3. 難治性がん（膵がん）並びに希少がんの試料を用いたパイロット解析を行うことを目的とした。収集臨床情報の内容・収集方法（固形腫瘍・造血器悪性腫瘍それぞれのサブWGにより収集方法の確定）、現場負担軽減策の検討、検体処理・収集・保管等のワークフロー等について原案を作成した。がん全ゲノム解析パイプラインの開発・実装を行い、実際のデータを用いた運用を開始した。難治性がん（膵がん・白血病）並びに希少がんの試料を用いて、全ゲノム解析・RNA シークエンス解析・全ゲノムメチル化シークエンス解析についてのパイロット研究を行い、総計で 445 ペア症例の全ゲノム解析、515 サンプルの RNA シークエンス解析、160 サンプルの全ゲノムメチル化シークエンス解析を終了した。

A. 研究目的

がん全ゲノム解析計画のための体制構築として、1. 解析検体・臨床情報収集の検討、2. データ解析ワークフロー、3. 難治性がん（膵がん）並びに希少がんの試料を用いたパイロット解析を行うことを目的とする。

B. 研究方法

1. 解析検体・臨床情報収集の検討

バイオバンク WG を立ち上げ、収集臨床情報の内容・収集方法（固形腫瘍・造血器悪性腫瘍それぞれのサブWGにより収集方法の確定）、現場負担軽減策の検討、検体処理・収集・保管等のワークフロー等につ

いて討議を行った。

2. データ解析ワークフロー

がんゲノムシーケンス解析においては、アラインメントや変異検出に利用するソフトウェア、オプション、リファレンスゲノムなどの定義ファイルの選定により、最終的に検出される後天的変異のリストが大きく変わり得る。解析結果について、他のプロジェクトとの比較、横断的な解析を可能とするためには、統一した形式で解析実行を行うことが不可欠である。また、変異検出のためには近年では、感度や正答率を上昇させるために、複数の変異検出ツールの結果を組み合わせるといったアプローチが頻繁に採用されている。上記の背景を踏まえ、国際的に広く利用されているソフトウェア、定義ファイルを選定しつつ、所定の一通りの解析処理を行うソフトウェアを開発する。またその他にも、大規模なゲノムデータの効率的な解析を達成するための運用手順の整備を行う。

3. 難治性がん（膵がん・白血病）並びに希少がんの試料を用いたパイロット研究

国立がん研究センターバイオバンクに保存されている膵がん・胆道がん凍結検体を用いて、全ゲノム解析・RNA シーケンス解析・全ゲノムメチル化シーケンス解析を行った。骨軟部腫瘍について、別研究により PDC (Patient-derived Cell)、PDX (Patient-derived Xenograft) が樹立されている症例を中心に症例、組織型の選定を行い、160 症例について凍結腫瘍組織検体および凍結正常組織検体（もしくは抽出済みの正常組織由来 DNA 検体）の収集を行った。収集した凍結腫瘍検体は OCT 包埋後、薄切し病理組織評価のための HE 染色標本作製するとともに DNA、RNA 抽出を行い、NanoDrop、Qubit、Bioanalyzer による評価を行った。DNA は腫瘍組織 196 検体（再発、転移検体等を含む）、正常組織 160 検体について全ゲノムシーケンス解析を行い、RNA は腫瘍組織 162 検体（再発、転移検体等を含む）について RNA シーケンス解析を実施した。また、収集症例について臨床情報の収集を行い、この集計作業を行った。

（倫理面への配慮）

今回の収集症例は各医療機関附属のバイオバンクに保管されている症例を中心に検体収集を進めた。バイオバンク保管の検体は研究目的での利用について包括的同意がとられた後のものであり、本研究計画の遂行は各医療機関の IRB 承認を得て、研究計画を情報公開しオプトアウト機会を保障して進めている。また、一部の症例については医療機関での IRB 承認のうえ、個別に同意が取得され、検体収集が行われているものを用いた。検体収集においては各医療機関で匿名化符号を付与し対応表を各医療機関で保管しつつ、個人識別に直結する情報を研究解析機関側に伝えない様管理して、検体収集、臨床情報収集を行った。

C. 研究結果

全ゲノム解析に供する検体について、収集すべき臨床情報の内容・収集方法（固形腫瘍・造血器悪性腫瘍それぞれのサブ WG により収集方法の確定）について検討を行った。前向き収集検体処理について推奨手順書案を作成した。

がん全ゲノムシーケンスの解析ワークフローのソフトウェアの開発を行なった (<https://github.com/ncc-ccat-gap/GCATWorkflow>)。本ソフトウェアにおいては、GPU サーバーを利用したアクセラレーターである Parabricks の利用をサポートしており、通常の CPU クラスターとの併用により格段の高速化が可能である。また、変異検出の方法として、mutect2, manta, GRIDSS といった国際的に広く利用されているツールに加えて、国内で比較的頻繁に利用されている Genomon を選定し、ワークフローの中でこれらを一括で実行可能な形式での実装を行なった。さらに、シーケンス拠点から受け取るシーケンスデータについて、統一したフォルダ構成、命名規則を取り決めることで、効率的な運用形式の整備を遂行するなど、非機能面においても、ゲノム解析を実行するための工夫を行なった。

膵がん・胆道がん凍結検体として 400 症例の検体を確保した。膵がんは腫瘍率が低いことが知られているので、最初に異常が起こることが知られている KRAS 遺伝子変異から腫瘍率を算定することにした。変異特異的 PCR 法により得られた KRAS 変異アリル頻度 (Variant Allele Frequency: VAF) を算出し、腫瘍率 30%以上の 285 症例については腫瘍部並びに同一患者正常組織検体について全ゲノム解析を行った。また腫瘍率が低い検体も合わせて 355 症例については腫瘍部から RNA を抽出し、RNA シーケンスを行った。全ゲノム解析を行った検体のうち、十分 DNA 量が確保できた 160 症例について更に全ゲノムメチル化シーケンス解析を行った。

骨軟部腫瘍 160 症例について腫瘍組織、正常組織検体を収集し、最終的に腫瘍組織由来 DNA196 検体（原発に加えて再発、転移時の腫瘍組織検体を含む）、正常組織由来 DNA160 検体（バイオバンクにて DNA 抽出済み検体を含む）を用いて全ゲノムシーケンス解析を実施した。また同様に、腫瘍組織由来 RNA162 検体を用いて RNA シーケンス解析を実施した。これらのシーケンス解析から得られた FASTQ データから、データ解析を進めている。また、解析対象となった症例の臨床情報を収集し、その集計作業を行った。軟部腫瘍の組織型の内訳は未分化多型肉腫 19 症例、悪性末梢神経鞘腫 17 症例、脱分化型脂肪肉腫 17 症例、滑膜肉腫 16 症例、皮膚隆起性線維肉腫 11 症例、骨外性骨肉腫 10 症例、粘液線維肉腫 8 症例、胞巣状軟部肉腫 7 症例、低悪性度線維粘液肉腫 6 症例、粘液型脂肪肉腫 5 症例、BCOR 再構成肉腫 2 症例、CIC-DUX 再構成肉腫 2 症例、その他の組織型 12 症例。骨腫瘍の組織型の内訳は骨巨細胞腫 11 症例、脱分化型軟骨肉腫 2 症例、軟骨肉腫 1 症例、脊索腫 1 症例、Ewing 肉腫 1 症例、平滑筋肉腫 1 症例、骨肉腫 1 症例、分類不能肉腫 1 症例であった。集計時点での観察期間中央値は 39 か月（1-163 か月）、発症時年齢中央値は 49 歳（4-93 歳）であった。

D. 考察

構築したがん体細胞全ゲノム解析パイプラインを GPU 上で実行可能なアクセラレーター (Parabricks) により高速化を行うことで、1 台の GPU で 6-7 ペア (腫瘍部 x100+正常部 x30)/day の計算処理が可能という見積もりを得た。今後、1. 数百、数千検体規模のデータで運用を行い、問題点を抽出する、2. 効率的な偽陽性フィルタリングの方法の開発、等を検討していく。本ワークフローを用いることで、1 台の GPU サーバーについて、1 日に 8 検体の腫瘍・正常検体のペア (腫瘍:100x, 正常:30x) の解析が可能であることを実証した。すでに 75 検体ほどの解析が終了しており、今後数ヶ月で本研究プロジェクトにおいて生成された全シーケンスデータの解析を終了できる見込みである。RNA シーケンス/全ゲノムメチル化シーケンスデータについては、解析 WG での議論も参考にしつつ、in-house での解析を実施する予定である。

E. 結論

1. 収集臨床情報の内容・収集方法 (固形腫瘍・造血器悪性腫瘍それぞれのサブ WG により収集方法の確定)、現場負担軽減策の検討、検体処理・収集・保管等のワークフロー等について原案を作成した。
2. がん全ゲノム解析パイプラインの開発・実装を行い、実際のデータを用いた運用を開始した。
3. 難治性がん (膵がん・白血病) 並びに希少がんの試料を用いて、全ゲノム解析・RNA シーケンス解析・全ゲノムメチル化シーケンス解析についてのパイロット研究を行い、総計で 445 ペア症例の全ゲノム解析、515 サンプルの RNA シーケンス解析、160 サンプルの全ゲノムメチル化シーケンス解析を終了した。

F. 研究危険情報

特記事項なし

G. 研究発表

1. 論文発表

該当なし

2. 学会発表

該当なし

H. 知的財産権の出願・登録状況

該当なし