

厚生科学研究費補助金（がん対策推進総合研究事業）

総括研究年度終了報告書

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

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、付随する ELSI（倫理的・法的・社会的課題）について検討し ICF（説明同意文書）案作成、収集臨床情報の内容・収集方法・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の作成、全ゲノム解析における適切なシーケンス解析条件の検討、データ共有・患者還元に関する課題点抽出、解析パイプラインの開発・実装と運用開始を行った。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として、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。

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. がん全ゲノム解析に付随する ELSI（倫理的・法的・社会的課題）の検討、先行解析における既存試料・情報の利用に関する留意点及び ICF（説明同意文書）案作成、2. 解析検体・臨床情報収集の検討、3. 全ゲノム解析におけるシーケンス解析手法の検討（深度と検出感度の関係調査、深度を高めた全ゲノム解析（腫瘍 200x, 正常コントロール 100x）の実行可能性と、変異やサブクローン構造の検出性能の検証、ロングリード技術を用いたがん全ゲノム解析のフィージビリティ検討）、4. データ解析ワークフロー、5. 全ゲノムデータ利活用および共有ルール、その実行体制に関する検討、6. 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、を目的とする。

B. 研究方法

1. がん全ゲノム解析に付随する ELSI（倫理的・法的・社会的課題）の検討

ゲノム解析および ELSI の専門家や患者の立場の委員計 11 名で構成されるワーキンググループを組織して検討を行った。計 10 回のワーキング会合を開催して論点の検討・意見交換を実施したほか、並行して文献調査を実施した。また、「がん全ゲノム体制班」の他のワーキング・グループからの要請に応じて ELSI の観点から助言・情報提供を行った。調査・検討の状況・成果については、厚生労働省「がん

全ゲノム解析等連絡調整会議」において随時報告した。

2. 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。

3. 全ゲノム解析におけるシーケンス解析手法の検討

ヒト大腸癌 5 症例（腫瘍 5 検体、正常コントロール 5 検体、計 10 検体）について、従来法（深度：腫瘍 50x、正常 30x）よりもよりディープにシーケンスを行い（深度：腫瘍 200x、正常 100x）、得られたシーケンスデータをヒトリファレンスゲノムにマッピングする際に、従来の CPU を用いた演算法と新たな GPU を用いた演算法の実施時間を計測し、解析に要する時間を比較する。サブクロンの検出能力の比較は、得られたシーケンスデータをランダムにダウンサンプリングし（深度：腫瘍 100x および 50x、正常 50x および 30x）、データ量によって検出できるサブクロンの腫瘍内分画サイズを検討した。時系列的なクローン構造の変化を評価するため、複数のタイミングで検体が採取された 25 症例の造血器腫瘍の症例について、異なるタイミングの検体の全ゲノムシーケンスを実施した。このうち 7 例は、2020 年の調査で生存が確認されている。ローグリード解析はナノポアシーケンシング解析を PromethION (Oxford Nanopore) を用いて実施した。

4. データ解析ワークフロー

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した。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った。

5. 全ゲノムデータ利活用および共有ルール、その実行体制に関する検討

全ゲノムデータを共有・活用するための考え方、インフラ等についての検討を行うべく、学界及び産業界のメンバー11名によってワーキンググループを構成した。

6. 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、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。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、腫瘍組織 162 検体（再発、転移検体等を含む）に

ついて RNA シークエンス解析を実施した。

(倫理面への配慮)

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

「全ゲノム解析等実行計画」におけるがんの全ゲノム解析に付随する ELSI として、1) ICF 案の作成、2) 既存試料・情報の利用に関する留意点、3) 解析結果の患者への返却に関する留意点、4) 遺伝的特徴・情報に基づく差別禁止に関する法制度の検討、の4点を抽出し、論点整理と対応方針の検討を行った。検討に基づき全ゲノム解析を想定した ICF および留意点の案を取りまとめた。

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。

東京大学医科学研究所ヒトゲノム解析センターのスーパーコンピュータ Shirokane において、得られたシーケンスデータをヒトリファレンス配列にマッピングを行った。この際、演算装置として従来の CPU を用いたパイプラインと、新たに GPU を用いたパイプラインによる解析を実施し、両方で解析に要する時間を比較した。深度 100x では GPU は CPU に比較して 4.9-7.2 倍、深度 200x では 11-13 倍高速であった。

germline SNP と判明している variant が腫瘍細胞検体に誤ってコールされることがある。その偽陽性コールはシーケンス深度に依存するため、正常コントロールの深度によって偽陽性コールがどの程度検出されるかを検討した。正常コントロールの深度が 30x でも SNP のコンタミネーション数は~4 個と非常に少なかった。この観点からは正常コントロール検体の深度は 30x で十分であると考えられた。

腫瘍検体の深度を一定としたときに、正常コントロール検体の深度が体細胞変異コールに与える影響を検討した。各ペア検体において、深度：腫瘍 200x、正常コントロール 100x でシーケンスした際の変異コール結果をゴールドスタンダードとして用いた。偽陽性(アーチファクト)の割合は正常コントロールの深度が増えることで減少した。一方、感度は正常コントロール検体の深度が 50x

で飽和しつつあった。正常コントロールの深度 50x でも偽陽性コールの割合は十分に低く、検出感度・偽陽性率の観点から正常コントロールの深度は 50x 強でよいと考えられた。正常コントロールの深度を 50x としたときの、腫瘍検体の深度による遺伝子変異の検出感度を検討したところ、腫瘍の深度 100x でも約 10%の変異を検出できておらず、メジャークローンおよびサブクローンの変異を高感度で検出するためには腫瘍の深度として 100~200x が必要と考えられた。以上の結果から、本検討の結論として深度：腫瘍 100~200x、正常 50x 強でのシーケンスにより腫瘍内サブクローンが有する遺伝子変異を高感度に検出できると考えられた。

標的シーケンスと比較した場合、全ゲノムシーケンスにおける遺伝子変異の検出感度はクローンサイズに依存し、アレル頻度の低い変異では感度の低下を認めたものの、アレル頻度 30%以上では 9 割以上の感度を得た (図 10)。一方でコピー数異常については、標的シーケンス解析で検出可能な異常の全てが全ゲノムシーケンスでも検出可能であった。また、時系列解析を行なった症例については、体細胞変異のアレル頻度の経時的な変化に基づいて、クローン構造の変化を明確に捉えることが可能であった。

検体に用いた患者由来膵がん細胞は三次元培養にて樹立されたオルガノイド細胞株 8 例および当該患者末梢血白血球から抽出されたゲノム DNA である。ゲノム DNA サイズは概ね 50kb 前後であったが、一部検体はより大きなサイズまで含まれていた。DNA の断片化は g-TUBE (Covaris) あるいは細径シリンジによるフラグメンテーションを用いて行った。前者は断片化されない分子の残存も観察されたが、安定してデータが得られた。PromethION フローセルを用いた解析で、全検体で 1 ランあたり 30Gb 以上の塩基情報が得られ、最大 110Gb、すなわち 30X ゲノムの情報収集が可能であることが確認された。PCR 増幅がないため、コピー数変化は鋭敏に検出可能であった。

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-cpat-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。

大規模全ゲノム解析研究において先行する英国 Genomics England と NHS Genomic Medicines Service, および米国 All of Us について調査を行い、とりわけ企業によるデータ利活用の状況について調査した。

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 シークエンスを行った。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クエンス解析を行った。

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クエ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 シークエンス解析を実施した。これらのシークエンス解析から得られた FASTQ データから、データ解析を進めた。

D. 考察

同意書は、保険診療として実施している「がん遺伝子パネル検査」とは違い、本事業の性質上、データベース公開や長期利用も含めて一括した同意を得る方式としている。企業の単独利用も可能としている。同意撤回があった場合、データについては速やかに「使用は停止」することとしている。インフォームド・アセントフォームは別に作成が必要であるほか、先進医療として実施する場合には全体的に書き直しを要する。生殖細胞系列の結果返却については、結果返却の希望、家族との共有可否・連絡先に関しては、C-CAT 同意書作成時の議論を踏まえて、任意の項目として含めている。意思変更申出については、生殖細胞系列の結果返却、家族との共有意思、再連絡可否を取り上げている。ただし、生殖細胞系列の結果返却については、ICF に留まらない課題が多く、別途、詳細な留意点の作成が必要である。さらに、事業計画が明確になった場合には、実施体制、どのような情報を収集するのか、データ二次利用方針等、データの利用目的、再連絡の可能性についても再検討が必要である。

がん全ゲノム解析については、深度：腫瘍 100~200x、正常 50x 強でのシークエンスにより腫瘍内サブクローン構造の解析が可能であること、100x の深度では低アリル域の感度が低いこと、経時的検体の解析によってクローン構造の変化が観察できることなどが示された。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

クエンズデータの解析を終了できる見込みである。RNA シークエンス/ 全ゲノムメチル化シークエ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。

データの患者還元にあたっては、以下のような検討が必要である。1)還元する情報については現行のがんパネル検査の結果とも比較検証しながら選定する。2)データの解釈に必要なレファレンス情報として、各がん種について日本人のがんゲノム情報が先行解析で収集されることが求められる。3)全例を患者還元とするかどうかは慎重な議論が必要である。現在は日本人がんの全ゲノムデータベースが存在しないこともあるため、解釈が困難であることから、当面は還元内容について限定することが望ましい。4)全ゲノム情報の患者還元のロジスティクスを含め十分な検討が必要であり、体制整備を検討する研究班の設置が望ましい。

E. 結論

1. がん全ゲノム解析に付随する ELSI（倫理的・法的・社会的課題）について検討し ICF（説明同意文書）案作成した。
2. 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。
3. 全ゲノム解析におけるシークエンス解析手法を検討し、適切な解析条件を見出した。
4. 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。
5. データ共有・患者還元に関する課題点を抽出した。
6. 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。

F. 研究危険情報

特記事項なし

G. 研究発表

1. 論文発表
2. 学会発表

H. 知的財産権の出願・登録状況

該当なし

作成上の留意事項

1. 「A. 研究目的」について
 - ・厚生労働行政の課題との関連性を含めて記入すること。
2. 「B. 研究方法」について
 - (1) 実施経過が分かるように具体的に記入すること。
 - (2) 「(倫理面への配慮)」には、研究対象者に対する人権擁護上の配慮、研究方法による研究対象者に対する不利益、危険性の排除や説明と同意(インフォームド・コンセント)に関わる状況、実験に動物に対する動物愛護上の配慮など、当該研究を行った際に実施した倫理面への配慮の内容及び方法について、具体的に記入すること。倫理面の問題がないと判断した場合には、その旨を記入するとともに必ず理由を明記すること。なお、ヒトゲノム・遺伝子解析研究に関する倫理指針(平成25年文部科学省・厚生労働省・経済産業省告示第1号)、人を対象とする医学系研究に関する倫理指針(平成26年文部科学省・厚生労働省告示第3号)、遺伝子治療等臨床研究に関する指針(平成31年厚生労働省告示第48号)、厚生労働省の所管する実施機関における動物実験等の実施に関する基本指針(平成18年6月1日付厚生労働省大臣官房厚生科学課長通知)及び申請者が所属する研究機関で定めた倫理規定等を遵守するとともに、あらかじめ当該研究機関の長等の承認、届出、確認等が必要な研究については、研究開始前に所定の手続を行うこと。
3. 「C. 研究結果」について
 - ・当該年度の研究成果が明らかになるように具体的に記入すること。
4. 「F. 健康危険情報」について
 - ・研究分担者や研究協力者の把握した情報・意見等についても研究代表者がとりまとめて総括研究報告書に記入すること。
5. その他
 - (1) 日本工業規格A列4番の用紙を用いること。
 - (2) 文字の大きさは、10～12ポイント程度とする。