

「全ゲノム解析等に係る厚生労働科学研究班」からの 資料

厚生労働科学研究班
「がん全ゲノム解析等の推進に向けた患者還元、解析・データセンター、
ELSI 等に係る技術評価、体制構築についての研究」
からの報告

中釜 齊 (研究班長、国立がん研究センター・理事長)

患者還元WG

河野 隆志 (WG長、国立がん研究センター研究所・分野長)

織田 克利 (東京大学医学部附属病院・教授)

解析・データセンターWG

井元 清哉 (WG長、東京大学医科学研究所・教授)

白石 友一 (国立がん研究センター研究所・分野長)

ELSI WG

横野 恵 (WG長、早稲田大学社会科学総合学術院・准教授)

田代 志門 (東北大学大学院文学研究科・准教授)

令和3年度のAMED研究を開始するために方針決定が必要な事項

担当	大項目	中項目	小項目	策定方針
患者還元WG	患者還元	対象疾患、患者数	対象疾患および対象ステージ詳細	再発頻度や学会意見等に従い、ゲノム情報の有用性を考えて対象症例を選出する。
		医療機関要件	EP、全ゲノム解析体制の要件など	がんゲノム医療中核拠点・拠点病院での施行が望ましい。連携病院は中核拠点・拠点病院と協力して遂行する。
	利活用	保管、管理ルール	採取、処理、移送、保管、管理SOP	検体採取・保管の推奨手順を定める。品質保証、かつ、臨床現場で可能な手法で行う。国内完結で行う。
	シーケンス	受託要件	NGS台数、WGS実績等	品質保証、かつ、ゲノム解析・検査の実績を重要視する。
		技術的要件	WGS深度、RNAseq範囲	全ゲノム解析の深度は（腫瘍120, 正常30）を目標とする。RNAseqでは2千万リード以上を取得する。
		QC	QC方法、タイミング等	世界標準を見据え、昨年の予備解析で定めた手法を活かす。解析・データセンターによるcoverage等のQCデータを共有し、均質データの持続的な取得法の確立に活かす。
解析・データセンターWG	解析・データセンター	ゲノム解析	ゲノムデータベース構築	速やかな一次解析データの共有の仕組みとする。各研究班の提出義務とするゲノム元データを集約する。利便性とセキュリティの双方に目を配る。
			統一パイプライン	研究用一次解析としてのパイプラインを構築し、速やかにデータ共有する。患者還元には、臨床研究者と結果の検証やパイプライン最適化などを議論する。
		臨床情報等の活用	臨床情報DB構築（API自動収集）	臨床現場でfeasibleな項目内容・数や入力手法とする。API自動収集については可能な施設より実行を開始する。
		データ共有	共有ルール	国内外のアカデミア・企業が一定ルールのもとデータを共有する(WG枠を超えたメンバーで策定)。
			利活用ポリシー	国内外のアカデミア・企業の利用を考えたポリシーを制定する(WG枠を超えたメンバーで策定)。
ELSI	ELSI	ICF	患者還元に係る統一ICF作成	モデル文案を作成し、各研究班のプロトコルおよびICFへの反映を依頼する。各研究班からのフィードバックを得て中長期的にモデルICFを拡充する。
		ガイダンス	IC手法、2次的所見・結果還元在り方等	「研究により得られた結果等の説明」を含め、統合指針を遵守して行う。ICF以外に本事業についての周知・広報が必要。二次的所見対応については小杉班ガイダンスが基本となる。

患者還元WG

○班長：河野隆志（国立がん研究センター）

患者還元WG (1)

全ゲノム解析結果のうち科学的に妥当で実臨床、臨床試験等の形で診療に役立つデータを当該患者に還元する。

全ゲノム解析等の対象とする患者、実施機関について

- ・ **対象患者は再発頻度や学会意見等に従い、ゲノム情報の有用性を考えて対象症例を選出する。**
具体的には以下の項目を考慮する
 - ・ (前向き解析の場合) 新たに、患者に対して十分な説明のうえ、同意が得られること。
 - ・ 手術、生検、採血 (血液腫瘍) 等により全ゲノム解析に必要な質・量の検体が得られること。
 - ・ 手術等において根治の可能性が低い難治がんであること。
 - ・ 解析開始時に生存しており、何らかの治療の提供が期待できる状態であること。
ただし、希少がん・難治がん等で十分なゲノム情報が集積されていないがんについては、この限りではない。
- ・ **ICやエキスパートパネル、2次的所見対応など患者還元の体制が確保された定められた施設において実施する。**
 - ・ がんゲノム医療中核拠点・拠点病院での施行が望ましい。連携病院はこれらと協力して遂行する。
 - ・ 統一解析パイプラインによる解析結果の改良を解析・データセンターとともに行う。
 - ・ 臨床情報収集の項目や手法について、解析・データセンターと決定・改良していく。
(項目の選出、ネットワークを用いた選択式入力など、データの均一化、かつ、負担軽減の方策を含めて)
 - ・ 遂行の経験を手順、IC手法、データ解析手法・レポート作成システムの策定や継続的な改良に活かす。
 - ・ OJTにより、医療従事者のゲノムリテラシーを向上させる。

患者還元WG (2)

臨床現場でfeasibleな実行体制の構築

- ・ **試料の採取、処理、移送、保管、管理については、品質保証かつfeasibleな手法で行う。そのすべての過程を国内完結で行う。**
- ・ 十分な品質のDNA・RNAを確保するため、組織検体の採取・保管に関しては以下の手順を遵守する。
組織採取後、30分以内に液体窒素に浸漬して急速凍結を行う。速やかな処置が困難である場合、検体は4℃の冷蔵庫に保管し、3時間以内に急速凍結を行う(日本病理学会「ゲノム研究用病理組織検体取扱い規程」)。
 - ・ 手術検体だけでなく、生検試料の解析手技を合わせて開発する。
 - ・ 患者還元のため、統一解析パイプラインによる解析結果の検証を行う。
- ・ **ゲノム解析受託施設は、品質保証、かつ、ゲノム解析もしくは検査の実績を重視して選出する。**
 - ・ 均一で高品質なシーケンスデータ確保のため実績を重視すべき。
 - ・ ゲノム解析の先進諸国との国際共同研究でも活用可能なシーケンスデータの取得が可能であること。
 - ・ 遺伝子検査にかかる精度管理を実施している衛生検査所等でシーケンスを行うこと。
ゲノム検査でISO15189 認定（公益財団法人日本適合性認定協会）、CAP-LAP 認定（米国臨床病理医協会、臨床検査プログラム）、CLIA 認定（CLIA認証検査室改善法）のいずれかを取得していることを重視する。
 - ・ 前向き解析に向けて、検体回収や検体識別番号付けなど、運用面の対応も考慮する。
- ・ **腫瘍内不均一性を考え、全ゲノム解析の深度は（腫瘍120, 正常30）とする。RNAseqでは2千万read以上を取得する。ただし、サンプルの質によることを鑑み目標値として設定する。**
- ・ **世界標準を見据えたQC手法を設定し、実行する。具体的には、昨年の予備解析の経験を活かす。**
 - ・ 受託企業に対しては、QV30/20以上の塩基割合が75%/90%以上であること、重複リードを除いたのちに N: 90G塩基、T: 360G塩基以上のデータを取得することを標準化手法での委託条件となる。
 - ・ 全ゲノム解析を必要とする患者の試料の制限を考え、質・量に従った手法を合わせて準備、委託する。その際には上記に準ずるデータ量に関する委託条件を定める。
 - ・ 解析・データセンターによるcoverage等のQCデータを共有し、サンプル、手順や委託要件の見直しを含め、均質データを継続的に取得できる手法を確立する。

解析・データセンターWG

○班長：井元清哉（東京大学医科学研究所）

解析・データセンターWG (1)

全ゲノム情報を患者還元するためのデータ解析・提供・保管を実施する研究・創薬等の拠点。

- ・統一解析パイプラインを構築し全ゲノムシーケンスにより得られたFASTQデータの一次解析を行う。
- ・臨床情報を収集し、ゲノム情報と合わせ診療に有用な情報を患者に還元するためのデータ提供を行う。
- ・全ゲノムデータベースを構築し、研究・創薬等における利活用についてのシステム、ルールを確立する。

統一解析パイプラインについて

- ・ 統一解析パイプラインを制定する目的は、多くの研究者が行う最大公約数的な解析処理を中央で一括して行うことで解析・データ共有の効率化を図ることである。
- ・ 解析のメニュー・パラメータについては、国内外の動向を鑑み、将来のデータ共有、各種データベースとの適合性を図るために、一般的なものを選定すること。
- ・ 統一解析パイプラインでの処理内容はオープンソースで管理される。
- ・ 統一解析パイプラインの品質を保つために、中長期的に管理・運営をする枠組みを考案する必要がある。
- ・ FASTQデータのアライメント率等の品質に関する項目をレポートする機能を有すること。
- ・ 各々の研究班、がん種での個別の解析目的のために、独自の後処理・偽陽性フィルタ等を加えることが想定される。変異検出ツールの結果の後処理・偽陽性を除去するフィルタリング方法などは解析センターと各研究班で連携して協力して構築しつつ、最終的な判断はユーザーが行うことができる体制を構築すること。

解析・データセンターWG (2)

ゲノム解析について

- 年間1万症例程度の全ゲノムシーケンスデータ（腫瘍120x、正常30x）の一次解析（マッピング、バリエーションコール）を実施できる体制を構築すること。
- 一次解析した結果は、データ提供機関と速やかに共有すること。
- 解析の進捗については、データポータルを通してデータ提供機関と共有できる仕組みを検討する。
- 生殖細胞系列変異においては、データの偏り等に十分注意を払った解析を検討すること。劣勢遺伝性疾患に関する公表など、その範囲については、ELSIと十分に検討すること。
- 現在のインフラを有効活用しつつ本事業を進めるため、スタート時はスパコンによる解析を先行で進めるが、順次クラウドに移行する。
オンプレミスのスパコン上でもGrid Engine等で計算ジョブを並列差配し、処理を分割し複数の計算ノードで解析が行える環境を構築し、スムーズにクラウドに移行できるよう検討すること。イルミナのシステムなどを参考にしながら、サプライチェーンリスクへの対応としてひとつのベンダーに依存しない高度な並列化処理が可能な形式でのシステム構築を進めていくことが必要。
- がん種横断的解析など高度な2次解析の体制の構築についても解析センターのフォーカスにある。研究者が高度な2次解析を実施するための体制を検討すること。
- コンピューティングリソースの管理ができる人材やデータ解析人材など専門性の高い人材を確保した上で、その育成を長期的な視野のもとに進める。

解析・データセンターWG (3)

データセンター機能の構築

- 1次解析から2次解析まで解析ワークフローと多様な解析資源の提供をオンプレ・クラウドのハイブリッドも活用して実現する。
- 医療機関からの臨床情報、シーケンスセンターからのシーケンスデータは、解析・データセンターでデータベース化する。
 - 解析・データセンターがコンプライアンス・ガイドラインへ適合しているかを判断する仕組みを設けること。
 - 情報管理に関しては、政府統一基準群や医療安全性ガイドライン、本事業における情報セキュリティに関する取り決め事項に従った情報管理を行う必要がある。定期的な監査を検討する。
認証機能の強化（多要素認証の導入）、データの無害化の方法、不正アクセスの検知や自動遮断等々についての方法論を検討すること。また、基盤運用の負担軽減に必要な運用管理の自動化も考慮すること。
 - データのアップロードなどを管理できるテクノロジー窓口の設置が必要。
 - 医療機関、シーケンスセンター、解析・データセンターとの間には、セキュリティを確保したシステムやネットワーク接続環境の整備が必要となる。また、データセンターは物理的に国内にあるセンターの利用を基本とする。
 - 複数のシーケンス拠点とのデータのやり取りを可能とするために、データ転送・フォーマットの方式を規格化を行う。
 - 解析処理は、同時並行で複数の処理が並列に実施できる実装とする。さらに、ワークフロー言語、仮想コンテナ技術、コンテナオーケストレーション技術等を活用し、効率的な実行形式を検討すること。
 - 使用頻度に応じたデータ保管形態の工夫による保管コストの削減と分散保管によるセキュリティ確保が必要である。
 - データストレージについては、既存のリソースを活用しつつ、クラウドへの移行を検討すること。

解析・データセンターWG (4)

- 臨床情報の収集においては、臨床現場で feasible であることが必要。API自動収集については、可能な施設より実行を開始する。その際に、REST-APIを活用することを前提とし、本事業に必要な臨床データ項目やデータの送受信を机上ではなく、電子的な構文や送受信の仕組みを開発する。
また、特定のアプリケーション技術やライセンスを活用することなく、オープンなAPIとして実現する。そのための、制度的な課題なども研究する。
- 人工知能技術の活用を見据えた各種データの統一化が必要である。また、人工知能の最適化に必要な人工知能維持開発基盤の構築も進める
- 常に新たな技術を取り入れられるように、情報システムは特定の技術に依存しないように、拡張性が高い方式で整備する。
- データポータルを通して事業の進捗や成果を定期的に公開すること。
- 臨床情報の収集や臨床に必要な知識のデータ保管、ゲノム解析基盤（シーケンスセンターとのネットワーク等の接続を含む）、人工知能を活用した変異等の検出、解析データの管理と利活用、システム全体の保守・運用や高度なセキュリティ防御など、各々の分野は技術的専門性が違う。これらの要素毎に最適な開発技術者の確保と開発体制の整備が必要であり、そのための調達構造やシステム開発・運用のガバナンス維持のためのガイドライン等の整備も行う。
- ゲノムデータの利活用システムについては、ストレージに蓄積されているシーケンスデータ、変異情報などへのリソースに対して必要なアクセスを行うWeb APIの整備を行う。また、当該APIを利用したポータルサイトなどのアプリケーション開発を通じ、ゲノムデータの効果的な利活用に資するAPIの規格化、オープンな利活用アプリケーション開発の際の制度的な課題の検討を行う。

ELSI WG

○班長：横野恵（早稲田大学）

ELSI WG (1)

ICFの作成方針

- 統合指針の遵守が前提となる。
- 必須とする共通事項の範囲と内容を確認してモデル文案を作成し、各研究班のプロトコールおよびICFへの反映を依頼する。

必須とする共通事項

- 全ゲノム解析等実行計画の事業概要等
- 本事業の一部として実施する上で必要な研究目的についての記載¹
- 臨床情報収集の共通項目（右記）
- データの流れ²
- 検体の流れ
- 知財の取扱い
- 患者還元

1. 一次利用として想定されるデータの利用目的がカバーできる内容とする
2. 解析機関，データ・センター，公的データ・ベースおよび二次利活用の方針

- 各研究班からのフィードバックを得て中長期的にモデルICFを拡充する。
- 既存検体の利用については昨年度ELSI WGで留意点を取りまとめており、必要に応じて研究班と共有する。
- ICF以外に本事業についての周知・広報（ウェブサイトや説明資材（パンフレット，院内掲示用ポスター等））が必要。ICFでは各研究班のプロトコールの説明が主体になる。

臨床情報収集の共通項目（2月時点の案）

- ① 症例基本情報・同意情報
- ② 検体情報
- ③ 患者背景・既往歴・併存疾患・家族歴
- ④ 診断情報
- ⑤ 治療情報（薬物療法等）
- ⑥ 治療経過（最良治療効果，転帰，G3以上の有害事象等）

ELSI WG (2)

患者還元のガイダンス

- 本事業における患者還元は、研究として実施するため「人を対象とする生命科学・医学系研究に関する倫理指針」（2021年6月30日施行。以下、「統合指針」）を遵守する。
- 統合指針では、ゲノム指針の「遺伝情報の開示」および「遺伝カウンセリング」の規定に代わる医学研究全般に共通の内容として、新たに「研究により得られた結果等の説明」の項目が設けられた（ここでいう「結果等」には、一次的所見・二次的所見の双方が含まれる）。

<研究により得られた結果等の取扱い関係>

① 「研究により得られた結果等の取扱い」に係る規定の変更

ゲノム指針「第3の8 遺伝情報の開示」「第3の9 遺伝カウンセリング」の規定を改訂し、研究により得られた結果等を研究対象者に説明する上で必要な概念や手続を、人を対象とする生命科学・医学系研究全般を実施する上で配慮を要する事項として規定する。

Source: 医学研究等に係る倫理指針の見直しに関する合同会議第7回参考資料1 <https://www.mhlw.go.jp/content/000670423.pdf>

- 統合指針の規定により
 - ① 結果等の研究対象者への説明方針を定めて研究計画書に記載すること
 - ② 遺伝カウンセリングや遺伝医療の専門家との連携を含め、研究に係る相談実施体制を整備することが求められる。

ELSI WG (3)

患者還元のガイドンス (つづき)

- 本事業でも、各研究班において、上記規定に基づく対応方針を検討・整備することが基本となる。
- なお、説明方針の検討にあたっては、関連する以下の報告・提言を参照するものとする。

作成者	対象	作成時期等	
日本臨床腫瘍学会・日本癌治療学会・日本癌学会	「次世代シーケンサー等を用いた遺伝子パネル検査に基づくがん診療ガイドンス」	主に固形がん細胞・組織に生じる変異を検出するがんゲノムプロファイリング検査	2020年3月改訂
AMED小杉班	「ゲノム医療における情報伝達プロセスに関する提言」	がん遺伝子パネル検査 生殖細胞系列網羅的遺伝学的検査	2019年12月改定 厚労科研で改定検討中
AMED長神班	「研究における個人の遺伝情報の結果返却 検討および留意すべき事項と今後の議論・検討に向けた課題に関する提言」	研究における個人の遺伝情報の結果返却のうち、生殖細胞系列に関する遺伝情報を返却対象とするもの	2019年3

*二次的所見の開示対象については、小杉班提言の推奨度別リストを基本とすることが想定される。

- 研究班向けのセミナーを開催して、以下を共有し共通理解と課題の把握を図る。
 - ① 統合指針の関連規定および全ゲノム解析の場合の留意点 (昨年度ELSIワーキングにおいて論点整理)
 - ② 関連する報告・提言の内容
- 全ゲノム解析結果の患者への還元に関しては、本邦での経験の蓄積が乏しいため、本事業による患者還元体制構築の実践を通じて得られた知見を今後のガイドンス作成・改定に活用する。

データ共有ルール（データシェアリングポリシー）

- ・収集したデータ（臨床情報、ゲノムデータ）の迅速な利用を実現するため、必要な事項を定める。当該ルールについては、国際的動向や研究の進展にともない、適宜、方針修正等を行う。
- ・収集したデータの共有に際しては、取得された臨床情報、シーケンス情報を解析・データセンターで詳細に分析し、データベース化したものを共有する。
アカデミアが参画するフォーラム（アカデミアフォーラム）、産業界が参画するフォーラム（産業フォーラム）を形成してデータ共有を行う。
- ・アカデミアフォーラム
データ取得者とデータ解析者で構成される。
計画立案を目的とした全データの俯瞰・簡易解析が可能である。
自データ以外を用いた詳細解析には、データ取得者との合意とデータ利活用審査委員会への報告を必要とする。
自由な出願・知財の占有ができる。
論文等での成果公表、出願においては、データの公的DB移行の時期によらず、フォーラムへの届け出を行う。
- ・産業フォーラム
計画立案を目的とした全データの俯瞰・簡易解析が可能である(有償)。
詳細解析には、データ利活用審査委員会の承認を必要とする(有償)。
自由な出願・知財の占有ができる。出願に関して、フォーラムへの届け出は不要。
成果を公表できるが、公表までの少なくとも24か月、36か月は超えない範囲の制限期間※1,※2,※3を設ける

※1 各がん種において100例程度のデータ登録時点(希少がんは別途定める)を、起始ポイントとする。

※2 出願公開により公表される場合(出願公開制度による出願後18カ月の公開)を除く。

※3 論文公表等でデータ取得者・解析者が非公表の必要性が乏しいと判断した場合は、制限期間を24か月未満に短縮する。

成果公表においては、データの公的DB移行の時期によらず、フォーラムへの届け出を行う。

- ・一定期間(3年間を想定)以降、ゲノムデータおよび基本的な診療情報は公的DBに移行する。
詳細な診療情報は個人同定の配慮から公開しない。
企業等による詳細な臨床情報の利用には、データ利活用審査委員会の承認(と企業においては追加料金)を必要とする。
利用者は知財を占有できる。
論文公表までの制限期間なし。

データ利活用ポリシー骨子

目的

- ・本ポリシーは、収集したデータ（臨床情報、ゲノムデータ）を幅広く、公正かつ円滑に利活用するために必要な事項を定める。

データ利活用について

- ・データ利活用審査委員会の承認のもと、利用許諾等を行う。
- ・学術研究や医薬品等の開発、科学的なエビデンスに基づく予防を目的とする申請に対し許諾される。
- ・利用者は国内外[※]の学術研究機関や企業を想定する。
※日本と同等の水準にあると認められている個人情報の保護に関する制度を有している国または地域
- ・データの利用許諾等を受ける者がポリシーの定めに違反した場合、その者の氏名の公表やデータの利用許諾等の停止、新たな利用申請の拒絶、差止請求、損害賠償請求等の措置を講ずることができる。
- ・データの利用許諾等を受ける者が、自己以外の第三者に当該データの提供や転売等を行うことを認めない。ただし、自己が本ポリシー及びデータの利用許諾に関する契約で負うのと同等の守秘義務を課した上での自己の委託先や関係会社への提供についてはこの限りではない。

データ利活用審査委員会

- ・第三者への利用許諾等について公平性を担保することを目的に設置する。
- ・データ利活用審査委員会のメンバーは男女両性を含むものとする。そのほかの要件は別途定める。
- ・データ利活用審査委員会のメンバーは、適切なCOI管理を受ける。また利活用審査会メンバーとして知り得た情報については守秘義務を負うものとする。

データ利活用審査委員会（続き）

- ・利用許諾等の可否に関して以下の審査項目について審査する。
申請者による利用目的が、学術研究や医薬品等の開発、科学的なエビデンスに基づく予防を目的とするものであること
個人*・血縁者の同定、血縁関係の存在・不存在の確認、又は個人、小集団そして地域に対して不利益をもたらす可能性のある利用にあたらぬこと
*臨床試験参加など、患者還元を目的とする際はこの限りではない。
研究計画の科学的妥当性及び利用するデータ範囲の妥当性があること
申請者が研究計画を遂行するにあたり、十分な実績又は能力を有していること
申請者が所属する機関が十分な研究設備や情報の保管・廃棄などの管理体制を有していること
その他、データ利活用審査委員会メンバーが必要と判断する事項
- ・データの利用者は、データ利活用審査委員会より許諾を受けた利用目的以外の目的にデータを使用してはならない。

知的財産権

- ・データの利用から生じる知的財産及び知的財産権は、創出した利用者に帰属するものとする。

公表

- ・利用者は、当該データを用いた研究成果を公表することができる。
- ・データのうち患者臨床情報について、成果発表に必要な範囲で論文等に掲載できる。
- ・（特定）個人の識別につながる可能性等について十分に検討し、その可能性がある場合には発表前に再度利活用審査会に報告し協議することとする。

秘密保持

- ・データの利用許諾等を受ける者は、当該データに係る情報を秘密として管理し、第三者に開示又は提供をしてはならない。ただし、データ利活用審査委員会において許可が下りた利用目的の範囲内においては、当該利用者自らが負うのと同等の守秘義務を課すことを条件に解析委託先等にも開示することができる。
- ・データの利用許諾等を受ける者は、データを取り扱うにあたり、不正アクセス、紛失、破壊、漏洩などのリスクに対し、合理的な安全対策を講じなければならない。

情報公開

- ・データの利用許諾等を行った相手先の名称等について公開する。