

## オミクスデータ共有促進ガイドンス(詳細版)

先導的 ELSI 研究プログラム オミクス研究班

### はじめに

医療のあり方を世界的に変革しつつある ICT や AI、ビッグデータなどを用いた健康・医療分野のデジタル化の流れにおいて、オミクスデータ（ゲノム、エピゲノム、トランスクリプトーム、プロテオーム、メタボロームなど）を含む広範な健康・医療データを用いた研究開発が注目されており、データの品質と信頼性の向上、アクセスや共有の強化が目指されている。目標として、EU では研究者に対して、2022 年までに少なくとも 100 万人分のゲノムデータへのアクセスを提供し、2025 年までに少なくとも 1000 万人分のオミクス、診断画像、電子医療記録を含むコホートデータへのアクセスを提供するとし、EU 全域の調整が行われている<sup>1</sup>。日本でも、研究分野全体で必要なクリティカルマスに到達するために、AMED を中心にデータの共有化促進への取り組みが行われている。

さらに、日本では個人情報保護法（個人情報の保護に関する法律（平成 15 年法律第 57 号）の改正（2017 年施行）、EU では一般データ保護規則（以下「GDPR」）の成立（2018 年施行）など、デジタル世界におけるデータ保護の強化、個人のエンパワメント、データ主体の権利拡大といった機運の高まりも見られている。

このような状況のなか、特に、個人情報保護法をはじめとする規制環境の変化が、データ共有の促進のみならず、従来のデータ共有の状況に対しても、制限をかける方向に働くのではないかと懸念する声があるが、研究者の間で聞かれるようになった。

そこで本研究班では、従来行われてきたオミクスデータの共有方法に対し、改正個人情報保護法<sup>2</sup>や、関連のガイドンス、指針などによる規制状況の変化が、実際に制限的に機能し、共有への障壁になるのかどうかを検討し、本ガイドンスとして報告することにした。

本ガイドンスは、日本医療研究開発機構（以下、AMED）による「ゲノム医療実現推進プラットフォーム事業（先導的 ELSI 研究プログラム）」平成 28 年度採択研究課題「学際連携に基づく未来志向型ゲノム研究ガバナンスの構築」（研究代表者：大阪大学 加藤和人）のうち、「ゲノム・オミクス・臨床データの保護と共有に関する ELSI 研究」（以下「オミクス班」）の研究成果である。本ガイドンスはオミクスデータの共有促進に資することを目的としたもので、ベタープラクティスであり、医学系研究をはじめとするヒト由来試料を対象とした研究を遂行する際に産出されたさまざまな種類のオミクスデータに

---

<sup>1</sup> European Commission, 2018. Communication from the commission to the European parliament, the council, the European economic and social committee and the committee of the regions, on enabling the digital transformation of health and care in the Digital Single Market; empowering citizens and building a healthier society, COM/2018/233 final, Brussels, 25. 4. 2018.

<sup>2</sup> 2019 年 5 月現在、現行の改正個人情報保護法は、個人情報保護委員会において、新たな見直し作業が行われているが、本ガイドンスはその見直し内容を取り込んだものではない。

適用されるものである<sup>3</sup>。このガイダンスに沿ったオミクスデータ共有を進めることで、円滑にデータを共有でき、ゲノム医療研究開発が一層促進されることが期待される。

## 目次

1. 本ガイダンスの基本的な考え方
2. オミクスデータ共有促進へ向けた考え方
タイプ1
タイプ2
タイプ3
《参考資料》ヒトデータ共有の一般的な枠組み
謝辞および執筆者
《別紙》オミクスデータ共有促進ガイダンス(概要版)

### 1. 本ガイダンスの基本的な考え方

#### 1) オミクスデータ共有の背景

次世代シーケンス (NGS) 技術の登場によりゲノム解読のコストが低下したことでゲノム医療の研究開発が国内外で急速に進展している。「『ゲノム医療』とは、個人の『ゲノム情報』をはじめとした各種オミクス検査情報をもとにして、その人の体質や病状に適した『医療』を行うことを指す」<sup>4</sup>。ゲノム医療の研究開発において、ゲノムデータを含む様々なオミクスデータを用いることで、疾患の発症マーカー探索などの研究開発を促進することができると共に、その人の病状に適した医療を選択・提供することが可能となる。そのためには、オミクスデータが適切に広く共有され、利活用されることが重要である。

#### 2) 従来行われているオミクスデータの共有方法

オミクスデータの共有方法は、大きく「非制限公開データベース」（オープンアクセス）と「制限公開データベース」（管理されたアクセス）の2つに分類される（末尾の《参考資料》ヒトデータ共有の一般的な枠組みを参照）。

オミクスデータのうち、解析後のデータをまとめた集計データや統計解析結果、一過性あるいは後天的に変化する発現データや代謝物データなどから、氏名等の個人識別情報を除去したものは、特定の個人を識別できなくしたもの<sup>5</sup>と考えることができ、一般に、「非制限公開データ」として共

<sup>3</sup> 本ガイダンスは、データの共有を対象としており、試料そのものの共有については対象としていない。

<sup>4</sup> ゲノム医療実現推進協議会中間とりまとめ（平成27年7月）より。

<sup>5</sup> 基本的にはこのように考えるが、非常に稀な値が含まれていて個人識別性が失われない可能性がある場合を、後述の「タイプ2」で議論する。

有される。

一方、個人の識別性を失っていないオミクスデータ（生殖細胞系のゲノムデータなど生涯変わらず一意に個人と結びつくと考えられるデータや、本人とのリンク（対応表）を残しておく必要のあるデータなど）は、「制限公開データ」として共有されることが一般的である<sup>6</sup>。その他、データを公開しない「共同研究」ベースでの共有も行われている。

### 3) 個人情報保護法の改正等によるデータ共有への影響

本研究班は、オミクスのうち、主にゲノム、エピゲノム、トランスクリプトーム、プロテオーム、メタボロームについて、研究者からの具体的なデータ共有の状況の聞き取りを行い、それらが改正個人情報保護法等の規制の状況と、矛盾する点があるかどうかを検討した。

詳細は後に述べるが、結論としては、従来のデータ共有方法は、2017年に施行された改正個人情報保護法の規制内容と大きく矛盾するところはなく、従来の方法でデータ共有できることがわかった。本ガイダンスでは、共有方法を変更せず、従来通りの方法で共有を続ける場合を、「タイプ1」とした。

研究者への聞き取りでは、現状だけでなく、オミクス研究における技術の進展やコストの増減といった将来の研究の展望についても検討した。その結果、データによっては、共有方法の変更が、共有の促進のために有効であると考えられる場合があることがわかった。そのようなデータのうち特に重要であると考えられたものは、プロテオームやメタボロームを網羅的に解析した際のスペクトルデータ（いわゆる「Raw データ」）と、個人識別符号に該当するゲノムデータ（ゲノム配列）を取得あるいは推測することができるエピゲノム、トランスクリプトームやプロテオームの配列データであった。そこでこれらを、それぞれ「タイプ2」「タイプ3」として、今後のデータ共有の際に配慮すべき点について考察を加えた。

なお、タイプ1～3の関係については、《別紙》オミクスデータ共有促進ガイダンス(概要版)に図示するとともに、概要をまとめて記載した。

#### タイプ1

従来のオミクスデータ（ゲノム、エピゲノム、トランスクリプトーム、プロテオーム、メタボロームのほとんどすべてのデータ）  
従来通りの方法で共有することが可能である。

<sup>6</sup> 特定個人の識別性を失わせることができないデータは、個人情報と考えられ、個人情報保護法や研究倫理指針によって、原則として本人の同意内容の範囲内でしか利用できない。そのため、制限公開データベースにおいて、利用目的と同意内容の適合性やアクセス可能な者などが管理される。

### タイプ2（今後配慮が必要）

個人の網羅的なスペクトルデータ（メタボロームとプロテオームの網羅的に解析された個人毎のスペクトルデータ。いわゆる「Raw データ」）

特徴的な情報（特殊な値）を含む場合に再識別リスクが高まる可能性、偶発的に機微な情報が得られる可能性に配慮して共有する必要がある。

現在は共同研究での共有が一般的だが、制限公開データベースによる共有促進が可能。

### タイプ3（今後配慮が必要）

個人のゲノム配列を取得あるいは推測することができるデータ（エピゲノム、トランスクリプトームのリードデータ、プロテオームの変異ペプチド配列データなど）

原則従来通りの方法で共有可能だが、個別のケースごとに個人識別リスクを考量し、制限公開データベースでの共有も検討。特に研究遂行過程で、エピゲノム、トランスクリプトームのリードデータやプロテオームの変異ペプチド配列データから推測されたゲノム DNA 塩基配列のうち、個人識別符号に相当するような個人識別性のあるデータを共有する場合は制限公開を行う。またその際には、医学系指針ではなく、ゲノム指針に従う。

## 2. オミクスデータ共有促進へ向けた考え方

### タイプ1

タイプ1は、ゲノム、トランスクリプトーム、プロテオーム、メタボロームなどのオミクスデータが、従来の方法で共有される場合である。

これらは、以下に述べるように、改正個人情報保護法等による影響はほとんど発生していないと考えられ、従来どおりの方法で共有可能である<sup>7</sup>。

#### 1) 個人情報保護法

個人情報保護法は 2013 年前後に見直しが始まった。見直しの過程では、定義を拡大し、個人情報に該当する情報の範囲を広げる方向への議論もあったが、結局、成立した改正個人情報保護法（2015 年 9 月成立、2017 年 5 月 30 日施行）では、個人情報に該当する情報の範囲を広げることはなく、定義の「明確化」ととどめることとなった。したがって、従来から個人情報であったオミクス

<sup>7</sup> ただし、「タイプ1」の説明の後段で述べるように、海外にあるデータベースや海外の研究者に（国内のデータベースを介さず）直接データを提供する場合には、研究倫理指針により、本人がその旨を認識できるようにしておくことが必要になった。また、記録の作成・保存義務も強化されているので、これまで記録が不十分であった場合には、対応が必要である。

データは法改正後も個人情報であり、従来から個人情報でなかったオミクスデータは法改正後も個人情報ではないと考えることができるため、改正前後でデータ共有方法に矛盾は生じないと考えられる。

ただし、個人情報の該当性をより明確に判断できるよう、個人情報の定義のカテゴリーに「個人識別符号」の枠組みが設けられた。オミクスデータの個人識別符号への該当性について問題になることがあるが、結論からいうと、ゲノムデータの一部やエピゲノムのリードデータ（定量前の元データ）のみが該当し、トランスクリプトームの RNA 塩基配列データやプロテオームのアミノ酸配列データ、メタボロームデータは、個人識別符号に該当しない。

個人識別符号とは、「身体の特徴のいずれかを電子計算機の用に供するために変換した文字、番号、記号その他の符号であって、特定の個人を識別するに足りるものとして個人情報保護委員会規則で定める基準に適合するもの」である（個人情報保護法 政令第1条）。したがって、データが個人識別符号を含む場合、住所・氏名等の個人を識別する情報を除去しても、なお個人識別符号によって個人を識別できるので、そのデータは個人情報になる。個人識別符号の代表は、マイナンバーや旅券番号である。生体情報に関連するものでは、生体認証に用いられる（本人を認証することを目的とした装置やソフトウェアにより、本人を認証することができるようにした）指紋等のほか、「細胞から採取されたデオキシリボ核酸（別名 DNA）を構成する塩基の配列」があり、さらにそのうち、「全核ゲノムシーケンスデータ、全エクソームシーケンスデータ、全ゲノム塩基多型（single nucleotide polymorphism : SNP）データ、互いに独立な 40 箇所以上の SNP から構成されるシーケンスデータ、9 座位以上の 4 塩基単位の繰り返し配列（short tandem repeat : STR）等の遺伝型情報により本人を認証することができるようにしたもの」が指定されている。

個人識別符号に該当するオミクスデータは、ゲノムデータやエピゲノムのリードデータのうち、上記の条件に当てはまるもののみである。「細胞から採取されたデオキシリボ核酸」と限定されているため、RNA データやタンパク質データはもとより、RNA やタンパク質のデータから推測されたゲノムデータは個人識別符号には該当しないと解釈される。

したがって、オミクスデータのうち個人識別性のあるデータは、従来と同様、（個人識別符号ではなく）通常の個人情報である。その例としては、住所・氏名のような特定の個人を識別する情報が付いているオミクスデータ、個人識別符号に該当するゲノムデータや、生体認証の用に供した指紋などの情報などが付いているオミクスデータ、対応表などの個人を識別可能な情報が存在するオミクスデータなどである。繰り返しになるが、この状況は従来と変わっていないので、個人情報保護法の改正によって個人情報該当性が変化したわけではない。

また、改正個人情報保護法では、本人に対する不当な差別、偏見その他の不利益が生じないようにその取扱いに特に配慮を要するものを「要配慮個人情報」に規定した。要配慮個人情報とは、個人情報のうち、「人種、信条、社会的身分、病歴、犯罪歴等、その他の政令指定情報」が含まれるものである。「その他の政令指定情報」には、医師等により行われた健康診断の結果や医療を提供

する施設の診療記録などが含まれる。つまり、オミクスのうち、診療目的で解析を実施したゲノムデータや遺伝子検査の結果（消費者直販型を含む）であって特定の個人を識別できるものは、要配慮個人情報に該当する。一方、ゲノム以外のオミクスデータは、現段階では、本人の診断や治療に利用されることがほとんどない。これらのことから、純粋な研究目的で得られたオミクスデータは、病歴や、健診・診療目的で得られたデータ、遺伝子検査結果を含まない限り、要配慮個人情報には該当しないと解釈される。なお、要配慮個人情報に該当することとなったデータについては、原則として明確な本人同意のもと（オプトイン）でしか取り扱うことができないが、2)に述べるように研究においては、要配慮個人情報に該当したとしても、大きな影響は生じない。

## 2) 研究倫理指針

個人情報保護法が改正されたことを受け、研究倫理指針（ヒトゲノム・遺伝子解析研究に関する倫理指針[以下、ゲノム指針]、人を対象とする医学系研究に関する倫理指針[以下、医学系指針]など）が改正された。結論を先に述べると、改正研究倫理指針についても、従来のオミクスデータの取り扱いや共有方法に関して、大きく変更された部分はないといえる。研究倫理指針の主要な変更点に、①「連結可能・不可能匿名化」の用語の廃止、②要配慮個人情報のオプトイン原則、③トレーサビリティの確保、④外国にある第三者へ個人情報を提供する際のオプトイン原則、がある。以下に記載する通り、①②については大きな影響がないものの、③④については、適切な対応が必要となる。

「連結可能・不可能匿名化」用語廃止では、個人識別符号に該当するゲノムデータについては対応表の有無にかかわらず、個人情報として取り扱うこととなった。しかし個人のゲノムデータは、従来より個人情報に準じた安全管理措置を講じるよう定められており、誰でも閲覧可能な形（オープンアクセス）での共有は行われてこなかったこと、また、生体試料取得時から『研究成果を公的データベースに登録し、二次利用される可能性がある』ことについて同意を受けていることから、従来のデータ共有にとって大きな影響はないと考えられる<sup>8</sup>。

要配慮個人情報のオプトイン原則については、研究倫理指針によって個人情報保護法の例外規定等を用いたうえで、従来とあまり変更なくデータを利用できるよう配慮されており、要配慮個人情報に該当する場合でも、従来と同じように、オプトイン、オプトアウトあるいは匿名化等の条件のもとで通知・公開等による利用が可能であるため、特に大きな影響は生じていないといえる。

一方、改正個人情報保護法において、個人情報の流通状況をトレースできるようにすることが定められたことより（個人情報保護法第 25 条）、個人データの授受の記録作成と一定期間の保管が義

---

<sup>8</sup> ただし、既存試料・情報から得られるゲノムデータのうち、公的データベースへの提供とデータの二次利用に関して、十分な同意を得ていなかったデータも存在する。どのような同意内容であれば十分であるかという点については、本研究とは別の AMED のプロジェクトである MGeND データベースへのデータ提供ガイダンスにおいて基本的な考え方が公表されている。（<https://mgend.med.kyoto-u.ac.jp/submission>）

務付けられた。つまり、個人情報の該当性判断については変更されていないが、記録義務の追加によりトレーサビリティの強化がなされた。個人情報に該当するオミクスデータの共有に際しては、研究倫理指針のガイダンス等に記載されている、記録が必要な項目や、記録の保管年限を確認しておくことが必要である。

更に、改正個人情報保護法において新たに外国にある第三者へ個人情報を提供する際にも明確な同意が必要となった（個人情報保護法第 24 条）。そのため、海外の共同研究者等とデータを共有する場合や、海外のデータベースへデータを登録する場合、あらかじめ本人の同意を受けることが必要になった。研究倫理指針では、原則として、適切な同意を受けなければならない、としているが、適切な同意を受けることが困難な場合は、必要な倫理的な手続き（研究目的等定められた項目の通知または公開、拒否の機会の保障等）により外国にある第三者への提供が可能となる。とは言え、昨今は研究遂行過程において国際共同研究の機会が増え、また、研究成果を公的データベースへ登録することが出版社や研究助成機関から求められるようになってきている。今後は、研究開始時より、インフォームド・コンセントの説明文書において『日本国内の研究機関に所属する研究者だけではなく、製薬企業等の民間企業や海外の研究機関に所属する研究者もデータを利用する可能性がある』旨について記載し、明確な同意を受けておくことが大切である。

## タイプ2

タイプ2は、メタボロームとプロテオーム解析研究において、個人の試料が網羅的に解析された際のスペクトルデータ（いわゆる「Raw データ」）が共有される場合である。

現状では、プロテオームやメタボロームの Raw データは、データの質の管理や解釈の難しさなどから、ほとんどが共同研究ベースで共有されている。タイプ1で述べたように、この状況を変えないのであれば、改正個人情報保護法等の内容と矛盾はなく、従来どおり共有することができる。

しかし今後、解析技術の進展やコストの軽減化とともに、ビッグデータ解析による新たな知見の創出のため、あらゆるオミクスデータへのアクセスの需要が高まってくることが見込まれる。また、学術論文雑誌に研究成果を投稿する際、論文作成に使用したデータの公開と共有が求められるようになってきている。そのため、従来の共同研究ベースでの共有だけでなく、制限公開データベースを介した、より広い共有方法への移行を検討する必要がある。その際には、以下のように、網羅的スペクトルデータが、個人に特徴的な情報や機微な情報を含む可能性と、個人が再識別されるリスクへの対策を講じておくことが推奨される。

### 1) 網羅的スペクトルデータ

プロテオームやメタボロームは、質量分析 (mass spectrometry: MS) や核磁気共鳴 (nuclear magnetic resonance: NMR) を用いて解析される。その際、あらかじめ選択した種類の代謝物等を、

定量的に測定する場合（標的解析）と、種類を特定せず、代謝物やタンパク質・ペプチドを網羅的に測定する場合（網羅的解析）とがある。タイプ2で考察する「網羅的スペクトルデータ」とは、主として後者の網羅的解析の場合における、いわゆる「Raw データ」を指す。

Raw データで検出されている物質のピークは非常に多数存在するので、それらが実際にどの代謝物（あるいは外来の物質）やタンパク質であるかを、理論的には無限に存在する化合物やペプチド断片の候補のなかから突き止めて同定する作業は、非常に困難である。そのため、Raw データには、同定されずに残ったままの未知のピークが多数存在することになる。タイプ2では、このようなRaw データを共有することについて、検討を行った。

## 2) 検討内容

メタボロームでは、スペクトルに含まれるピークの一部が、個人が摂取した薬物や食物、環境汚染等による影響を反映した物質に対応することがあり、薬物の成分によってどのような疾患に罹患しているか、また、体内に取り込まれた環境特異的な物質（化学物質、水、土壌など）によって居住地域がわかる可能性がある。また、プロテオームでも、疾患や食生活によって特徴的なピークが現れる可能性は、メタボロームの場合より間接的であるとしても、ありえないことではない。

当該データの二次利用の際にこれらの特徴的な情報が取得され、個人が再識別される可能性がある。またその情報が、同時にその個人の機微な情報であれば、再識別されることと相まって、個人に大きな不利益をもたらすことも懸念される。

プロテオームやメタボロームの Raw データは、ほとんどが共同研究ベースで共有されていることが、今回の調査で再認識された。今後、制限公開データベースを介した共有が増えていくことが予測されるが、そのための環境が整っていないことも、今回の調査から示唆された。制限公開データベースを介したプロテオームやメタボロームの Raw データの共有を促進するためには、以下の2点が重要であると考えられる。

### (1) 共有先でのデータの解釈が正しくできること

プロテオームやメタボロームの Raw データから最終的に物質を同定することは非常に難しく、共有先でデータを正しく解釈されない懸念がある。特にメタボロームにおいては、候補となる物質が、人工的な物質を含めてほぼ無限に考えられる。データを正しく解釈できる人材の育成や解釈方法に関する情報の共有が必要である。

### (2) 特徴的な情報や機微な情報が含まれる可能性について認識し、データ取扱いに関する共通理解がなされること

ここでいう機微な情報とは、個人情報保護法で定義されている要配慮個人情報指着しているのではなく、いわゆる「本人に社会的な不利益を与える可能性のあるさまざまな情報」や「稀で特



殊な情報」を指す。例えば、メタボロームの Raw データからは、違法な薬物の摂取や妊娠等の情報が、予期せず得られる可能性がある。これらの情報の共有促進には、漏洩対策等の安全管理について、研究者の共通理解が必要である。

メタボロームやプロテオームのデータの機微性を検討すると、ゲノムデータのような個人の一生を通じてのデータの不変性や、子孫への継承性という意味での機微性ではなく、病院等における一般的な「臨床検査データ」の機微性に類似している<sup>9</sup>。そこで、これらの Raw データの利用の際には、臨床検査データを研究利用する場合と同様に、医学系指針を遵守すること、電子化された医療情報のセキュリティガイドラインである「医療情報システムの安全管理に関するガイドライン 第5版」(厚生労働省)<sup>10</sup>に準じる扱いをすることを共通理解とすることによって、共有先でのデータの適切な取り扱いが期待できると考えられる。

### 3) その他の検討事項：トランスクリプトデータ

トランスクリプトームも、解析技術の進歩に伴い、次世代シーケンス技術による RNA シーケンシング (RNA-seq) データは高い精度で個人のデータを再現できるようになってきている。それに伴い、非常にまれな変異をもつ転写産物のトランスクリプトデータは、個人識別性や機微性を有する可能性がある。例えば、特定の疾病との関わりが明らかになっている転写産物のトランスクリプトデータは、他の臨床情報とあわせることで個人が特定された場合、大きな不利益をもたらす可能性がある<sup>11</sup>。次項のタイプ3で述べるように、トランスクリプトデータは、現状、非制限公開データベースで共有されることがほとんどだが、上で述べたようなケースについては、誰しもが閲覧可能な非制限の公開データとしてではなく、制限公開データとして共有することが推奨される。

### タイプ3

タイプ3は、エピゲノムデータのうち個人のゲノム DNA 配列を含むリードデータ、および、トランスクリプトームやプロテオームのデータのうち個人のゲノム配列を推測できるリードデータや変異ペプチドのデータの様に、個人識別符号に該当するゲノムデータに相当するデータが共有される場合である。一般的にこれらのデータは、さらに解析された後、特徴量やパターンを抽出し定量したデータが非制限公開データベースを介して共有されるが、近年、研究成果を投稿する際に、論文に

---

<sup>9</sup> オミクスのうち、メタゲノム (腸内や皮膚等の細菌叢、マイクロバイオーム) については、個人の一生にわたるデータの不変性や、子孫への継承性が一定程度存在するという情報もある。

<sup>10</sup> データ管理に民間事業者サービスを用いる場合には、さらに「医療情報を受託管理する情報処理事業者における安全管理ガイドライン」(経済産業省)や「クラウドサービス事業者が医療情報を取り扱う際の安全管理に関するガイドライン (第1版)」(総務省)も参考にすることが推奨される。これら3つのガイドラインは、「3省3ガイドライン」とよばれている。

<sup>11</sup> このような研究の実施に際して新たに同意を取得する際には、データベースを介して国際的に広くデータが活用されるというデータ共有の趣旨を踏まえ、個人の再識別リスクについて説明した上で同意を受けることが望ましい。

使用した元データを公的データベースへ登録することが多くなったことで、個人識別符号に相当するリードデータ等の共有が求められる場合も出てきている。

#### 1) 個人のゲノム配列を含むか推測することができるデータ

エピゲノムは、ゲノム DNA の塩基修飾等に伴う構造変化による遺伝子発現等への影響を解析するもので、それらの特徴を抽出・定量する前の段階のリードデータは、個人のゲノム DNA 配列を含む。これまで、エピゲノムのリードデータは共有の対象ではなく、ゲノム DNA の塩基修飾等の特徴量を定量した後のデータが、非制限公開データベースで共有されてきた。しかし近年では、上記の通り、個人の DNA 配列であるリードデータを共有する機会が増えている。リードデータに含まれる DNA 塩基配列の長さが、個人情報保護法上の個人識別符号に該当すれば、氏名や住所等の個人を識別する情報を除去したとしても、個人情報に該当する。このような場合には、Whole Genome Sequencing データや Whole Exome Sequencing といったゲノムデータと同様、制限公開による共有を行うこととなる。

トランスクリプトームの転写産物の発現量の測定手法には、マイクロアレイ、RNA-seq などさまざまな手法があるが、どの手法を用いたとしても、遺伝子単位の転写産物の「発現量」としてのデータについては匿名化した上で非制限公開データベースを介して共有されている。一方、次世代シーケンス技術を用いたトランスクリプトーム解析研究のなかには、ゲノム塩基配列を推測する内容を含むものも出てきている。例えば、転写産物をマッピングして遺伝子の変異の部位を同定する過程を含む研究等である。タイプ3では、特にこのような研究において、個人のゲノム配列を推測できるデータを共有する場合について検討を行った。

結論としては、個人のゲノム配列を推測できるデータは、個人識別符号に該当するゲノムデータを推測することが理論的に可能であること、また、今後、ICT や AI を含めた個人識別技術の進展が考えられることから、制限公開での共有に変更することを視野に入れて、検討を続けることが望ましい。

#### 2) 検討内容

上述のように、RNA の配列情報をマッピングして遺伝子の変異部位を同定する研究等では、個人識別符号に該当するゲノムデータ（DNA の塩基配列情報）を取得することも理論的には高い確率で可能となる。厳密には、個人情報保護法の個人識別符号に該当しないものの、データから氏名等の個人を識別する情報を除いても、なお特定の個人を識別することができるデータに相当する。ただし、DNA を直接シーケンスするのではないこと、過去に実施されたデータに関しては技術的に成熟していない点があることから、正しい情報を得られているとは言い切れないことが考慮されるべきである。よって、トランスクリプトーム解析研究においては、個人識別符号に相当する DNA の配列情報（変異の情報）を取得したとしても、それが直ちに「個人識別符号と同等である」とまで

は言えず、「個人識別符号に類似した」情報とするのが妥当であると考えられる。

「個人識別符号に類似した」情報が読み取れるとしても、個人識別符号そのものではないので、そこから特定の個人が識別できるかどうかの問題となる。個人情報保護法や GDPR では、個人が識別できるかどうか（個人識別符号以外の個人情報に該当するかどうか<sup>12</sup>）は、完全な匿名化が絶対的な意味で達成されているかどうかではなく、「個人を識別するのに合理的に使用される手段」があるかどうかによって、総合的に判断すべきという立場をとっている。つまり、理論的には個人識別が可能だけで、実社会においてその手段がないのであれば、個人情報に該当するとはいえない。

そこで、現時点では、トランスクリプトームのリードデータからゲノムデータを読み取ることによって個人を特定し、そのコストに見合う利益を得るなどということは通常考えられないとして、「個人識別符号に類似した」情報が読み取れるデータであっても、「個人を識別するのに合理的に使用される手段」が実社会に存在しないという根拠をもって、非制限公開で共有することは可能と考えられる。

また、特に非常に過去に取得されていたデータについては、データのエラー率や技術的精度を考慮し、個人識別性が低いものと考え、必ずしも制限公開とする必要は無いという解釈もありえる。

しかし、ICT や AI の技術は日進月歩であり、新たに「個人を識別するのに合理的に使用される手段」が利用可能になる可能性は常に存在する。したがって、技術の進展度合いによっては、研究ごとに個別にリスクを判断し、非制限公開ではなく、制限公開によるデータ共有を検討していくことが望ましいと考えられる。

なお、米国 NIH (The National Institutes of Health) では、RNA-seq の配列情報は制限公開データとして dbGAP (The database of Genotypes and Phenotypes) への登録を推奨している（発現量データは今まで通り GEO [Gene Expression Omnibus] へ登録）。NIH と三極連携のもと、ライフサイエンス分野で産生されたデータを共有するためのデータベースを運用している欧州 EMBL-EBI (The European Molecular Biology Laboratory-The European Bioinformatics Institute) や日本 NIG-DDBJ (National Institute of Genetics-DNA Data Bank of Japan) でも、配列情報は EGA (The European Genome-phenome Archive) / JGA (Japanese Genotype-phenotype Archive) へ、発現量データを ArrayExpress / GEA (Genomic Expression Archive) へ登録することを推奨している。

### 3) その他の検討事項：医学系指針とゲノム指針

トランスクリプトームやプロテオーム解析研究は、従来、医学系指針に従い実施されている。

ただし、上で述べたような、ゲノム塩基配列を推測する内容を含む研究、例えば、RNA の塩基配列をマッピングして遺伝子の変異の部位を同定する研究等の場合には、ゲノムデータ・ゲノム情報を解析する事に相当すると考えられるため、ゲノム指針のもとでの実施が望ましいと考えられる。

---

<sup>12</sup> 法で個人識別符号とされたものについては、そのみで特定の個人を識別できると定義されているので、実社会において識別のための方法や手段があるかどうかは関係しない。

トランスクリプトームやプロテオーム解析研究に限らず、オミクス研究においては、医学系指針とゲノム指針のどちらの研究倫理指針のもとで研究を行うかの判断が難しいことがある。これについては、今後、両指針に明記するなど、明確な根拠を指針が提供していくべきである。

《参考資料》 ヒトデータ共有の一般的な枠組み

アクセスの分類	共有されるデータ	例
非制限公開 (Open Access)	<ul style="list-style-type: none"> <li>・ 広汎な研究目的について同意を取得しているデータ</li> <li>・ 個人情報や機微情報を含まないデータ</li> <li>・ 研究倫理指針の適用外の試料を用いた研究から取得したデータ</li> </ul>	ゲノム塩基配列データ (SRA, ERA, DRA) 遺伝子発現データ (GEO, Array Express, GEA) 集計情報 (dbSNP, dbVar, JVar, gnomAD, TogoVar) 病的バリエーションデータ (ClinVar, MGeND)
制限公開 (Controlled Access)	<ul style="list-style-type: none"> <li>・ データベースを介した共有や、広汎な研究目的について同意を受けているデータ</li> <li>・ 個人情報や機微情報を含むデータ</li> </ul>	dbGAP, EGA, JGA バイオバンクが保有するデータ (東北メディカル・メガバンク計画統合データベース dbTMM)
制限共有 ※AMED 固有の概念	<ul style="list-style-type: none"> <li>・ データの制限公開に先駆け、限定された研究者間で共有するデータ</li> <li>・ 個人情報や機微情報を含むデータ</li> </ul>	AGD
オンサイトアクセス (Onsite Access)	<ul style="list-style-type: none"> <li>・ 個人毎のデータの譲渡を行わず、オンサイトでのアクセスのみを許可するデータ</li> <li>・ より識別性が高い個人情報やより機微性が高い機微情報を含むデータ</li> </ul>	NDB 医療情報のデータ 位置情報のデータ (LifeLines)

この他に、非制限公開と制限公開の中間にあたる、新しい公開概念として、氏名や所属等の情報を登録し、承認を受けた者がアクセスできる Registered Access による共有方法も提案されている。

## 謝辞

本ガイダンスの検討過程では、「ゲノム医療実現推進プラットフォーム事業 先端ゲノム研究開発 (GRIFIN)」各課題採択機関の研究者の方々から、アンケートの回答やヒアリング、Eメールによる貴重なご意見を頂戴しました。また、先導的 ELSI プログラムのプログラムスーパーバイザー、プログラムオフィサー、加藤和人研究開発代表と研究開発分担・研究参加メンバー、および基盤研究事業部バイオバンク課の方々からは、本研究の進行や方向性についてたくさんの助言をいただきました。世界経済フォーラム第四次産業革命日本センターの藤田卓仙氏からは、法律の専門家の視点で、本ガイダンスの内容について助言をいただきました。皆様には、ご多用のなか、快く本研究へのご協力をいただき、深く感謝いたします。

## 執筆者

### 先導的 ELSI 研究プログラム オミクス研究班

山本 奈津子（大阪大学 データビリティフロンティア機構）

荻島 創一（東北大学 東北メディカル・メガバンク機構）

川嶋 実苗（科学技術振興機構 バイオサイエンスデータベースセンター）

片山 俊明（ライフサイエンス統合データベースセンター）

岡田 随象（大阪大学大学院 医学系研究科）