

SARS-CoV-2 におけるゲノム疫学データインフラの抱える課題：パンデミックへの対応戦略の刷新（2020）¹

要約作成者：吉田隼大²

本報告書の概要

本報告書は、2019 年末から流行しているコロナウイルス感染症（COVID-19）の対策の一つとして大きな注目を集めているゲノム疫学（genomic epidemiology）に関して、技術の現状を説明したうえで、現在の感染症対策に不足している点や将来の発展的利用に必要な課題を指摘し、課題解決のための提言を行うものである。本報告書は、アメリカの学術機関である米国科学アカデミー（The National Academies of Sciences, Engineering, and Medicine）が設立した委員会によって執筆された。ゲノム疫学とは、ウイルスのゲノム情報を用いて感染症の感染メカニズムなどを解明する研究・学問分野を指す言葉である（「序論」中の定義を参照のこと）。後述されるように、感染症対策に有益な情報をもたらすためには、ゲノムに関するデータは他の臨床データや疫学的データと統合され運用される必要があり、本報告書もそうした諸データの統合運用に関する記述に紙幅の多くを割いている。

本報告書は五章構成である。第一章で序論、第二章で SARS やエボラウイルスなど過去の感染症事例におけるゲノム疫学の使用事例を報告したのち、第三章では現在行われているゲノム疫学の取り組みの諸事例が紹介される。第四章では有益なデータの統合とはどのようなものであり、その実現にはどのような取り組みが必要かについての考察が行われる。最終第五章では、ガバナンスという観点から、主にゲノムデータと統合される臨床データや疫学的データの開示・利用に関する国内および国際的な法的枠組みについての紹介と検討が行われる。なお、三章～五章の各末部には、委員会による「提言」が簡潔にまとめられているため、参考にされたい。

以下は本報告書の要約である。

¹ National Academies of Sciences, Engineering, and Medicine 2020, *Genomic Epidemiology Data Infrastructure Needs for SARS-CoV-2: Modernizing Pandemic Response Strategies* (2020). Washington, DC: The National Academies Press.

² 京都大学文学研究科 倫理学専修 修士一回生

全体の抜粋要約（段落前に付した番号が報告書の章番号に対応する）

ゲノム疫学とは、感染症ウイルスのゲノム情報を利用して、感染メカニズムの特定や感染拡大の防止に努める学問・研究分野だ。たとえば感染症ウイルスは自己複製の過程で頻繁に変異を起こすため、そうした変異を辿ることで、感染経路や感染症の発生源の特定が可能になる。2019 年末から世界中で流行している新型コロナウイルス感染症（COVID-19）への対応にも、この技術は有効である。現在、アメリカでもゲノム疫学によるコロナウイルス研究が行われているが、それらを支える体制や仕組み作りは決して十分ではない。

（3）まず、ウイルスのゲノム情報を収集する仕組みが整っていない。ゲノム情報の解析や提供は各地の諸機関で自発的かつ散発的に行われているのが現状だ。米国全土で様々な種類の人間集団からウイルスゲノムの収集が行えるように、国や州、群などあらゆるレベルでの幅広い機関において検査と解析の能力を充実させる必要がある。また、それらによって得られた情報を公開された形で広く保管し、かつ国外の機関とも情報共有しなくてはならない。

（4）次に、ゲノム疫学においては、ウイルスのゲノム情報を単体で用いるのではなく、感染症対応で従来用いられてきた疫学的データや臨床データと組み合わせて運用することが必須である。そのためには、それらデータの収集・分析を行う様々な諸機関の間で、データの統合・共有のための協力関係やネットワークの構築、システムづくりが不可欠だ。データの統合・共有そのものが過度な負担とならないように、それらを支援する機関の設立や資金提供、および分析にかかわる人員の養成なども重要だ。加えて、データの共通規格化や共有が許されるデータの内容・条件の明確化も、セキュリティやプライバシーへの配慮とともに進められる必要がある。また、これらの必要な取り組みに関しては、米国ではインフルエンザウイルス等の従来ある感染症に関するシステムが既に存在しているため、それらを発展・拡充させることが有効と思われ、新たな制度作りの場合にも参考になるだろう。

（5）最後に、法制度の観点から言えば、ウイルスのゲノム情報共有に関する法律が存在しない一方で、疫学的データや臨床データの共有・開示を規制する法律や規則は存在している。これらはゲノム疫学にとって効率的なデータ利用の障壁となり得るため、強力な権限を備えた指導者や機関が法改正や特例的な手続きを行うことによって、そうした問題を取り除く必要がある。また、現在の法律の内容が現場の機関や研究者に周知されていないため、彼らが必要以上に慎重になることによって効率的なデータ利用が妨げられているという現状もある。このため、特定の種類のデータを扱う際にどのような手続きが必要となるのかについて、法の規定と運用に関する明確なガイダンスを行うことも必要だ。

凡例：

特殊な用語や注意すべき表現については、元記事における英語表記を（）で示した。また、要約文中の重要な箇所は下線を引いて示した。

第一章 序論

2019年12月、中国武漢で従来確認されていない重篤な肺炎の症例が初めて見つかり、その原因が、2002年に発生した重症急性呼吸器症候群（SARS）コロナウイルスに関係した新しいコロナウイルスであると特定された。その後六か月の内に、SARS-CoV-2と命名された新型コロナウイルスは世界中に広がり、少なくとも一千万人に感染し、推計五十万人の死者を出した。社会的・経済的にも甚大な被害をもたらしたコロナウイルス感染症 2019（COVID-19）に対しては、本報告書執筆時点（2020年7月）で有効なワクチンが開発されていない。

COVID-19を含めた、複数の感染症の集団発生³は、パンデミックへの備えや対応の手法を開発する必要があることを示唆している。DNA解析やゲノム学、疫学、そしてビッグデータ分析といった新技術が、症状・無症状を問わず感染拡大のネットワークを追跡したり、感染拡大が起こる場所やリスクのある集団を特定したりするにあたって、革新をもたらしている。更にそれによって、ウイルス感染を防いだり遅らせたりすることで、社会的・経済的な混乱を緩和し、疾病率や致死率を下げることができる。それ以外にも、これら新たな病原体の追跡手法は、公衆衛生的な介入戦略や医療的な対抗措置、疾病管理を実行するための重要な時間的猶予を与えてくれる。これらの技術を最大限有効に活用できるように必要な準備や整備を、あらゆる側面から進めていくことが求められる。

コロナウイルスの進化と SARS-CoV-2

コロナウイルスは SARS、MERS、そして今回の SARS-CoV-2 と、21世紀に入ってからすでに三度の大規模な感染を引き起こしており、その事実は、脅威となるウイルスの変異などを知らせることのできる、しっかりしたウイルスのゲノム進化の測定法を作り出すことが極めて大切であることを示している。

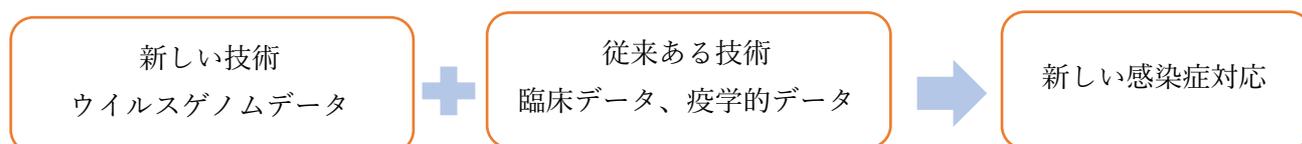
コロナウイルスは大規模な RNA ゲノム（28-32kb）を持っており、そのうち 20kb が自己複製機構に使われる。その機構を通して、ウイルスはその構成要素である構造たんぱく質、膜たんぱく質、エンベロープたんぱく質、スパイクたんぱく質、ヌクレオカプシドを複製する。SARS-CoV-2 は人間の上皮細胞にある受容体であるアンジオテンシン変換酵素 2（ACE

³ 本稿では、“outbreak”を「集団発生」と訳す。主に感染症の突発的な発生や感染拡大を指す言葉だが、主に医療機関など特定の空間内で短時間に多数の症例が発生する際に用いられることが多い。厚生労働省の「医療機関における院内感染対策について」では「一定期間内に、同一病棟や同一医療機関といった一定の場所で発生した院内感染の集積が通常よりも高い状態のことであること」と定義される。また、「北大病院感染対策マニュアル（第6版）」においては「『一定期間内(time), 特定の地域(place), 特定の集団(person)で予想されるより多く感染症が発生すること』あるいは、『公衆衛生上重要な感染症（新興感染症など本来あってはならない感染症）が発生すること』とされる。

2) にウイルスのスパイクたんぱく質が結合することで、人間に感染する。このスパイクたんぱく質には受容体を結合させる領域があり、それが抗体の攻撃対象やワクチン開発の目標になる。コロナウイルスを含む RNA ウイルスは RNA 依存性 RNA ポリメラーゼと呼ばれる酵素を利用してゲノム複製を行っており、この酵素がもつ複製のエラーを起こし易いという性質が、RNA ウイルスの変異の発生率が高いことの主要因になっている。比較的大規模なゲノムサイズを持っているコロナウイルスは、遺伝子の複製エラーを修正する仕組みを進化させることで変異率を他の RNA ウイルスよりも低く抑えているが、遺伝子の組み換えや挿入、削除といった現象によってもゲノムには変化がもたらされ、変異が生まれる。こうした機構によって、ウイルスは宿主の免疫や抗ウイルス薬に素早く対応し続けることが可能になっているのだ。それゆえ、RNA ウイルスのゲノム変異を監視することは、ウイルスの発生予測や病気の深刻さの予測、薬やワクチンの効果測定、宿主の集団における感染ネットワークの追跡などにとって重要である。あらゆるゲノム疫学研究にとって重要なことは、ウイルスの進化に影響する基本要因と選択圧とを理解することであり、SARS-CoV-2 における遺伝子の多様性がどのように発展してきたかを観察することは、抗ウイルス薬やワクチンなどの開発や病気の診断法の改善につながり得るのである。

SARS-CoV-2 の解明におけるゲノム学の可能性

ゲノム解析技術のもつ速度、粒度、扱いやすさ、持ち運びしやすさ⁴の進歩のお陰で、感染症の集団発生時におけるゲノム監視技術のもつ可能性は飛躍的に高まった。そして、それらの効果は特に、ゲノム解析が臨床的なデータや集団に基づいた疫学的データと組み合わせられた時に発揮される。これらゲノム解析技術の到来以前は、集団発生への対応に用いられる情報は、疫学的なデータだけに限られていたのだが、今日ではそれら種々の分野から得られるデータを組み合わせることで、以前より精確かつ的確な対応が可能になったのだ。



このウイルスゲノム解析こそが、「ゲノム疫学」という新しい分野の要石である。ウイルスゲノムの情報は、大規模感染のあらゆる段階で有用である。まず初期段階においては、感

⁴ ゲノム解析に必要な設備などが十分でない地域での集団発生対応が求められる場合などに効果を発揮する、「ポータブル型 DNA 解析 (portable sequencing)」と呼ばれる技術が登場している。これは小型化された各種医療ユニットを現地に展開し、その場で迅速に病原体の採取やゲノム解析、他のデータベースとの情報の統合や応答が行えるようにするものだ。詳細に関しては、以下の論文等を参照のこと。

Jennifer L. Gardy and Nicholas J. Lamon 2018 “Towards a genomics-informed, real-time, global pathogen surveillance system”, *Nature Review Genetics*, vol 19, 2018, pp. 9-20
URL : <https://www.nature.com/articles/nrg.2017.88>

染した組織の DNA 解析によって、病原体の遺伝子を特定することができ、宿主とウイルスの発生源を特定することも可能である。さらに次の段階では、そのデータを他のデータと組み合わせることで、感染拡大の時間的、空間的なパターンや、感染が連鎖する経路を描写することが可能となる。さらに、ウイルス感染の特徴を理解することは、より効果的な介入と対抗策の開発に役立つだろう。加えて、集団発生が起こっていない時期でも、ゲノム疫学はウイルスの進化や感染の動向を追跡することに役立つとされる。

ゲノム疫学の定義：

ゲノム疫学とは、感染症の感染（transmission）と疫学（epidemiology）を理解するために病原体のゲノム解析技術を用いることとして定義される。

さらに、ウイルスの具体的な種々の特徴に影響を与える変異がどんなものかについての知識が前もって得られている場合には、全ゲノム解析 (WGS) という手法が効果的である。WGS は集団レベルにおいては、疫学的データと組み合わせられることで、病原体の変異を、さらに詳細な解像度で感染のパターンを明らかにすることが出来る。加えて、病原体の WGS は、人間・動物・環境の健康を一つのものとして考える One Health アプローチ の枠組みの中に位置づけられる。殆どの感染症は動物由来であり、多彩な生物学的状況の中で人間に感染するため、これら三つの領域におけるゲノム疫学の利用が、これら病気の監視、予防、管理に役立つと考えられている。

第二章 COVID-19 以前の感染症集団発生におけるゲノム疫学の適用例

ゲノムデータ、臨床データ、疫学的データを組み合わせるゲノム疫学の利用は 21 世紀の COVID-19 以前の集団発生の際にも行われており、この手法の有する価値と、将来における感染症集団発生への対応の改善の可能性を示してくれている。

・ SARS-CoV

最初に広範なゲノム解析が行われ活用された集団発生は、SARS コロナウイルスが引き起こした 2003 年のケース (SARS-CoV) だ。このとき、ゲノムの解析と疫学的データの組み合わせにより、動物性の感染源から人間集団に複数の独立した感染が起こったことが明らかにされ、また、人間への感染経路となる動物 (ジャコウネコ、タヌキ) も特定された。患者についての他のデータが不十分なために、異種間での感染や人から人への感染に伴う変異や病原性の増大などは、具体的にはほんの一部しか特定できなかった。このケースにおけるゲノム解析は、人間への感染源の特定と、人間への感染を可能にした変異の特定を可能にしたと言える。

・ エボラウイルス

西アフリカで起こったエボラウイルスの集団発生においては、持ち運び可能なゲノム解析プラットフォーム (ポータブル型 DNA シーケンサ) とリアルタイムでの解析を活用して行われた大規模な WGS によって、感染の動態や様子が明らかになった。 ウイルスゲノム解析は、集団発生の起源はある単一の (ウイルスの人間集団への) 流入であると確定するのに用いられたし、さらにゲノムデータと疫学的データの解析によって母乳を介した感染があること、無症状の回復者であっても回復後長期間が経っても性行為による感染が起こり得ること、などが明らかにされた。これら感染メカニズムの発見により、疾病から回復した男性に向けての対策と指導で、回復後もエボラウイルス RNA の有無を確かめるために精液検査を繰り返し受けることを推奨するようになった。また、集団発生期間中も、ウイルスの変異を特定したり、集団発生終息後もウイルスの地理的な感染ルートを遡って特定し、将来の準備計画に役立てられるようにしたりしている。 エボラウイルスの事例では、エピソード期間中のリアルタイムでの応答として、統合されたゲノム疫学の役割の重要性が明らかになったのだ。

・ ジカウイルス

2015 年から 16 年にかけてアメリカ大陸で流行したジカウイルスの場合には、アメリカで確認されているウイルスの諸系統は仏領ポリネシアに共通の祖先をもつものであり、それらウイルスは当初の予測より一年以上前にブラジルで流行していただろうということが、これらの技術により明らかになった。加えて、アメリカでの鎮静後の 2017 年に、キューバ

で集団発生が起こっていたことが、キューバからの渡航者を対象にしたゲノム監視によって明らかになった。さらにゲノム研究により、この集団発生はその前年に起こった他国からの流入が原因であり、キューバ当局による蚊への積極的な対抗策が奏功したために感染拡大が遅れていたことも明らかになった。ゲノムデータは集団発生への対応策や、渡航者への注意喚起と行動監視による緩和策をとることを可能にする。

・季節性インフルエンザ

季節性インフルエンザの感染追跡を行う、既に存在している国内外のネットワークは、ゲノム疫学の発展に寄与してきたし、SARS-CoV-2の対応に直接応用可能である。 季節性インフルエンザはコンスタントに進化するので、ワクチンの開発と改良にはゲノムの監視が不可欠だ。アメリカ疾病予防管理センター（CDC）のFluViewは、長期の、様々な集団にわたるインフルエンザウイルスの拡大を追跡する監視ツールであり、これは様々なレベルの健康機関から多様なデータを集めている。また、インフルエンザウイルスのゲノム情報のさらに広範な共有を可能にすべく科学者らによって設立されたネットワークに、全インフルエンザデータの共有のための国際機関（GISAID）がある。 GISAIDは、COVID-19においても同様の取り組みを行っている非常に重要な機関である。

・おたふくかぜ

おたふくかぜはアメリカでは1967年のワクチン導入以来殆ど流行が起こっていないが、2016年から7年にかけて複数の大学での感染を中心に数百の症例が報告された。ゲノム解析の結果、単一の系統のウイルスが2006年から、主に米国内部のみで流行していたことが分かり、さらに詳細な解析によって、大学キャンパスへの複数のウイルスの持ち込みが、小規模なクラスターを発生させていたことがわかった。この研究の意義は、ウイルスゲノムのデータが「接触関係（contact links. 寮など）」や「活動の繋がり（activity links. スポーツチームやクラブ活動など）」のような臨床情報とどのように組み合わせられることで、集団発生のリスクを同定し、食い止めることにつながるかということを示した点にある。

・抗生物質耐性菌

抗生物質耐性菌（抗生物質に対する耐性を備えた細菌のこと）の感染拡大が起こった場合、抗生物質の開発速度はそうした細菌の進化に追い付かないので、その感染は出来る限り食い止めねばならない。そのためには、地域や病院レベルで患者の移動記録とゲノムデータを統合するシステムが開発・運用される必要がある。たとえば、米国の病院で感染を拡大したある抗生物質耐性菌に対して、ある機関が疫学的データとゲノム解析を組み合わせ、最初に起こった感染を特定して具体的な監視と臨床実践の方針を明確にし、それが集団発生の防止に役立ったという事例がある。こうした分野においては、地域の医療機関単位で様々なデータを共有し役立てる制度的枠組みやプログラムの作成、整備が行われており、実際に成果を

上げている。

・食物由来の病原性細菌の集団発生

ゲノム疫学の有益性が最も顕著なのは、食物由来の病原性細菌の集団発生においてである。この分野でゲノム疫学は、食品の安全管理に役立っている。たとえば、国家の監視ネットワークにおいてシガトキシンという毒を生成する菌の感染が発覚した、24の州からの56の患者に対する調査（特に食事に関するアンケート調査）の結果、彼らが病気の発症の一週間前にある共通した食材を摂取していることが分かった。すなわち、国内のある生産者から得られた小麦から患者に検出されたものと同じ菌が検出され、ゲノム解析の結果、患者と食材に関係があることが明らかになったのだ。ゲノムの解析データの情報共有の容易さと情報量の多さ、精確性のお陰で、地域レベル、州レベル、そして国家レベルでのより多岐にわたる共同が促進されているのである。

最良の実践（ベスト・プラクティス）と将来の成功に向けた鍵となるもの

これら数々の事例が示すように、他分野からの様々な種類のデータを組み合わせて用いることで、多様なレベルでの保健機関や官・民の研究機関、大学病院などを含む様々な機関の間の連絡が促進され、ある感染症集団発生の疫学的な状況分析をより豊かできめこまかなものにすることが可能になる。

このような諸機関に跨る協力というものは、ゲノム疫学のような多分野にわたるアプローチの将来における成功可能性の鍵となるものだ。その典型例の一つが、上述したような食物由来の病原性細菌の研究、対応である。

しかし、既に多くの進歩があったとはいえ、公衆衛生の研究機関によって用いられる解析の手法における一貫性や標準化を保証するためには、さらに多くのことが成し遂げられねばならない。加えて、リアルタイムでのデータ収集活動の継続的な発展や、予測に用いられるべき機械学習技術の導入なども、次の課題となっている。

集団発生に対応するための、時宜に適い、かつ効果的な実践的手法としてのゲノム疫学の適用範囲を広げるためには、公共機関・民間団体の双方においてゲノムの解読、解析の技術的な能力を確立することが不可欠である。とりわけ、感染症集団発生でより大きな被害を受けやすい資源の限られた地域における共同や支援などを含め、様々な取り組みが現在行われている。

第三章 SARS-CoV-2 に関連した現在のゲノム疫学的取り組み

SARS-CoV-2 集団発生の中である現在、この感染症に関連した様々なデータが、様々なネットワークで収集、共有されている。データは世界中で加速度的に生み出されている一方で、それらデータを素早く共有するネットワークは、地域、国家、世界レベルで現在構築されつつあるところである。この章では、SARS-CoV-2 データの情報源となる機関の一部を概観し、それら情報源が持つ限界をはっきりさせ、SARS-CoV-2 のために諸データを統合・分析する取り組みに光を当てる。

現在運用されている SARS-CoV-2 の諸情報源

・米国疾病予防管理センターによる公衆衛生における緊急対応、疫学、監視を目的とした SARS-CoV-2 解析 (U.S. Centers for Disease Control and Prevention SARS-CoV-2 Sequencing for Public Health Emergency Response, Epidemiology, and Surveillance, SPHERES⁵)

SPHERES は、国家レベルでの大規模なゲノム解析の取り組みを促進するために 2020 年 5 月に開始された。その目的は、臨床や公衆衛生の研究機関から得られる高品質なウイルスゲノム解析データの生成と共有を加速させ、また、統合的な分析のために用いられるその他情報の収集を進めるにあたっての基準を設けることである。今では全五十州に最新のゲノム解析を行う機能が備わっているが、ゲノム研究の規模を拡大する取り組みは、人員の数や諸団体との協力が限られているせいで、うまく進んでいない。現在の取り組みは季節性インフルエンザを想定した枠組みの中で行われており、規模の拡充が必要である。SPHERES が目的とするところは、特に

- ・高い水準でのモニタリングの維持
- ・国家レベル、地域レベルでのベースライン設定のための解析データの利用
- ・持続可能な、長期にわたるデータ収集の実行
- ・ウイルスゲノムの解析データを利用しての接触追跡の取り組みの支援
- ・ウイルスの遺伝子多様性の長期的な監視
- ・官・民のパートナーシップモデルを通じた新たな共同と技術革新の促進 である。

SPHERES は現在資金面でも不十分な状況にあるし、国内での地域によって偏った試験・解析の実践状況のせいで、得られたサンプルも不十分である。その結果、地理的な観点あるいはウイルスゲノムの多様性の観点から、そのサンプルは集団に関しを代表性を持つものとなっていない。

また、さらなる限界として、多くの研究機関が臨床データを提供していないため、

⁵ 米国疾病予防管理センターによる SPHERES の紹介ページ URL：
<https://www.cdc.gov/coronavirus/2019-ncov/covid-data/spheres.html>

SPHERES は臨床データの収集や連結は行えないということがある。

・**国立生物工学情報センター (The National Center for Biotechnology Information, NCBI⁶)**

NCBI は国立衛生研究所 (NIH) に属し、全てのゲノム解析データの主要な保管所として機能している。ゲノム学者たちは NIH のデータ共有ポリシーに則って、NCBI のデータベースに SARS-CoV-2 のゲノム情報などを預けたり公開したりしている。NCBI の方も、SARS-CoV-2 に関連する情報などを提供することが出来る。ウイルスゲノムを取得する方法にはいくつかの異なるものが確立されているのだが、ある種の基準というものが定まっている。この SARS-CoV-2 に関する諸研究機関からのサンプル提供やデータ提出によって、標準化の動きが進むだろう。また NCBI は様々な情報源からの細菌病原体のゲノム配列情報を統合するプログラムも実行しており、これはゲノム解析を用いたウイルスの発見、追跡を目的とする同様なモデルへのヒントになり得る。

・**インフルエンザデータの共有のための国際機関 (GISAID⁷)**

前述した GISAID はインフルエンザウイルスに関するデータ共有の主要なプラットフォームであり、SARS-CoV-2 にも対応するようになっている。GISAID は地理的データ、ゲノムデータ、臨床データ、疫学的データを収集して、集団発生状況下でウイルスがどのように進化し広がっていくのかを理解することに役立っている。ほとんどすべての主要な SARS-CoV-2 研究が GISAID を利用しているほどに、重要な機関である。

・**Nextstrain⁸**

Nextstrain は、多様な感染症の病原体から得られたゲノムデータを利用して、疫学的調査や集団発生への対応を支援するように作られたプラットフォームであり、世界中からの公に利用可能なゲノム配列情報の仲介役となっている。また、様々なスケールでデータを研究するのに利用可能な分析ツールや視覚化ツールも提供してくれる。Nextstrain は、世界中に広まっている SARS-CoV-2 の主要な系統や新たに誕生してくる系統を特定したりするのに貢献しているし、それが提供するツールによって、ゲノムの変異や感染率などをはじめとする様々な研究が可能になっている。

・**米国立 COVID コホート共同研究 (National COVID Cohort Collaborative, N3C⁹)**

米国衛生研究所 (NIH) が後援している N3C は、患者レベルでの COVID-19 の臨床データの集中型ポータルであり、臨床医や研究者、ヘルスケアの専門家が用いる手法やツールの

⁶ ホームページ URL : <https://www.ncbi.nlm.nih.gov>

⁷ ホームページ URL : <https://www.gisaid.org>

⁸ ホームページ URL : <https://nextstrain.org>

⁹ ホームページ URL : <https://ncats.nih.gov/n3c>

展開や評価を行うプラットフォームでもある。N3C は COVID-19 の臨床データの分析の効率やアクセスのしやすさを改善し、COVID-19 の分析能力を拡張し、共同的なデータ共有の新たなアプローチを示すために設立された。その共同研究は現在、COVID-19 に感染した患者についてのデータを蓄積しており、それらデータは、暗号化された識別子 (identifier) を用いて外部の他のデータに安全に結合されるものとなっている。

地域レベルで機能している機関の例

・ブロードインスティテュート (Broad Institute¹⁰)

マサチューセッツ州ボストンにあるブロードインスティテュートは遺伝子研究を牽引してきた機関であり、現在そこに所属しているグループはボストン地域における COVID-19 の流入と拡散について研究を行っている。その研究によると、ボストン地区への国内、外国両方からの 30 以上の流入やそれに続く大規模感染事例（集合住宅での無症状感染や、薬剤師の集会など）の発生が、マサチューセッツ州における集団発生を引き起こしたという。ブロードインスティテュートは 2016 年から州の公衆衛生研究所やアメリカ疾病予防管理センターと提携しており、そうしたプログラムは、州の公衆衛生機関同士の国家レベルにおける共同体制のモデルになるかもしれない。

・チャン・ザッカーバーグ・バイオハブ (Chan Zuckerberg Biohub¹¹)

カリフォルニア州において、カリフォルニア大学サンフランシスコ校およびカリフォルニア州政府の公衆衛生部門などと提携したパートナーシップのもとで、COVID-19 に関連した次のような取り組みを実施している。

- ・カリフォルニア全土の郡および診療所に COVID-19 の検査を提供する研究機関である CLIAHUB
- ・郡の公衆衛生担当部局による利用のため、データの視覚化や疫学的データの取り込みを行うシステムである COVID-Tracker の Nextstrain への実装
- ・カリフォルニアにおける州や郡の公衆衛生担当部局や学術機関、商業目的の研究機関やそのほかアカデミックな研究機関の共同体制である COVIDNet。これを通して、パートナー（サンプル提供者）はバイオハブや諸研究機関にサンプルを送り、ゲノム解析に役立てる。

¹⁰ ホームページ URL : <https://www.broadinstitute.org>

¹¹ ホームページ URL : <https://www.czbiohub.org> なおこの機関は、Facebook の創業者マーク・ザッカーバーグとその妻プリシラ・チャンが運営するチャリティー団体チャン・ザッカーバーグ・イニシアティブによって設立された。

・ワズワース・センター (Wadsworth Center¹²)

ニューヨーク州政府の保健部門に属する公衆衛生機関であるワズワース・センターは、SARS-CoV-2 検査の緊急使用権限¹³を受理した最初の研究機関で、SARS-CoV-2 の検査任務に堪えられるよう能力を拡充した。当センターはニューヨーク市を含むニューヨーク州における感染の研究と対応に当たっている。

その他諸機関

英国において類似の任務を果たしている機関に COVID-19 ゲノム英国連合 (Genomics UK Consortium) があり、多くの機関や団体と協力しながらウイルスのサンプル収集、ゲノム解析に取り組み、さらに諸医療機関や英国政府へのデータ提供、公衆衛生的な介入や政策策定の支援などを行っている。また、GISAID へのデータ提供も継続的に行っており、Nextstrain で閲覧することができる。

また、国際的なレベルでのゲノムデータ共有に関わる機関として、ゲノミクスおよび衛生のための国際連合 (The Global Alliance for Genomics and Health) などもあり、各国の各機関が提携している。

プライベートセクター、アカデミア、公衆衛生担当機関の間でのパートナーシップ諸事例

ワシントン州における、Microsoft 社のデータサイエンスチームとワシントン州の衛生局による、データ収集・管理システムの構築を目的とした提携の例がある。また、カリフォルニア州においては、Amazon Web Services 社とカリフォルニア大学サンフランシスコ校による、ベイエリアにおいて採取されたサンプルのゲノム解析のための提携などがある。

SARS-CoV-2 ゲノムデータと臨床データ・疫学的データを統合する現在の取り組み

本節では、これら取り組みのうち、特に有益と思われる武漢、ニューヨーク市、アイスランドの事例を紹介する。

・武漢

2019 年後半に武漢中央病院で、肺に異常があるが通常の呼吸器に作用する既知の病原体の検査は全て陰性という男性が見つかり、病原体を特定するためゲノム解析が行われた。その結果、以前にコウモリで確認されていた SARS コロナウイルスに類似した新たなゲノム

¹² ホームページ URL : <https://www.wadsworth.org>

¹³ Emergency Use Authorization. 緊急事態宣言下において、米国食品医薬品局が未承認薬などの使用を許可したり既承認薬の適応を拡大したりする法的制度。

が発見された。それと並行して、未確認の重篤な肺炎の症状を呈した複数の患者が武漢市内の別の病院の集中治療室に運ばれていた。彼らの多くが地域の同じ海鮮市場で勤務していた、という行動記録も明らかにされて感染経路が示唆された。この集団発生は、SARS 感染と類似した疫学的特徴を有していたため、研究者たちはまずそれがコロナウイルスであることを検査し、かつそれを確定させた。その後完全な SARS-CoV-2 ゲノムを特定した。さらに、彼らは SARS-CoV-2 のゲノム情報を、米国立生物工学情報センターおよび米国立医学図書館のデータベース、GISAID に提供したため、その後の COVID-19 の新たな症例を検査する際に、その情報を参照することが可能になった。

・ニューヨーク市

同市では 2020 年初めから感染が増え始め、3 月と 4 月に 172,000 の症例と 13,000 の死亡例が報告されるまでになった。研究者たちは IRB の承認と患者の同意などを得た上で、3 月初めの数週間で陽性とされた 84 の患者から臨床サンプルを得、それらの全ゲノム配列を解析し、GISAID に既に保存されていた別の SARS-CoV-2 のゲノム情報と比較した。その結果、ニューヨーク市に複数の独立した感染流入が起こったことが分かった。同市の症例の 87% は、当時欧州で支配的であったある単系統と同定され、欧州からの渡航が原因とされた。また、その他分析によって、1 月末から 2 月半ばにかけて、経路不明の国際的な感染の期間があったことが推定された他、ワシントン州から得られた系統に合致する症例も見つかり、これは国内での同市への感染流入を示唆している。これら 84 人の患者は同市の幅広い地区と近隣の諸都市に居住しており、地域内に広範な未確認の感染が起こっていることが示唆された。

・アイスランド

今回のパンデミックへのアイスランドの対応は、それが即座にゲノム疫学的な技術に影響を与えて革新的な手法を発達させた、という点で特筆に値する。同国は人口に比して巨大な数の観光客を毎年受け入れているため、外国からの感染症流入の脅威は大きい。そのため 25 年にわたって、集団レベルでのゲノム解析を行って病気のリスク要因を発見する取り組みを続けてきている。今事例では、症状と渡航履歴に基づいて感染のリスクが高いとされた 9,199 人への検査と、13,000 人の住民に対する集団レベルでのランダムな感染検査を行い、前者の 13% および後者の 0.6-0.8% が陽性となった。陽性者から採取されたウイルスゲノムのデータは GISAID にも提供され、流入や感染拡大の経路が明らかにされた。米国に比べて格段に小規模な国での事例だが、大いに参考になると思われる。

結語

上記の通り様々な試みがなされているとはいえ、現在のアメリカ合衆国では、SARS-CoV-

2のゲノム解析データの現在の情報源や、これらデータを関連する臨床・疫学的データと統合する努力は、典型的には受動的、反動的であり、整備されておらず、十分な資金が与えられていない。その結果として、現在利用可能なデータは多くの重要な人間集団の特徴を表現できておらず、バイアスのかかったもので、ウイルスの進化と変異、また、ゲノム配列の変異体と毒性、病原体、臨床的帰結や対抗策の有効性との関係に関する多くの切迫した問いに答えられない。それゆえ、SARS-CoV-2 ウイルスの解析データおよび関連する必要なデータは集められていない。

提言 1 :

アメリカ合衆国保健福祉省は、合衆国全土で、SARS-CoV-2 の代表的で質の高い完全なゲノム配列データが得られることを保証すべきである。また将来においても、SARS-CoV-2 以外の発生したパンデミックの病原体について、同じことが保証されるべきだ。それによって、ウイルスゲノムデータがゲノム監視 (genomic surveillance) に不可欠な役割を果たすことが可能となる。

- ・病原体は人種や民族、ジェンダーや年齢、地域、その他様々な人口統計学的特徴 (居住形態や臨床所見や帰結、遺伝性といったもの) のような要素の広範な多様性を表す個人から採取されねばならない。
- ・試験を行う多くの地理的に分散した場所でゲノム解析の能力を充実させ、支援する必要がある。その対象には公衆衛生の研究所や学術・医療機関が含まれる。
- ・全土から集められた SARS-CoV-2 の臨床サンプルは現行の原則にもとづいて収集され解析されるべきであり、ベースラインとなるデータを与え、リアルタイムに近い感染追跡を可能にすることが求められる。
- ・ゲノム配列情報は公的にアクセス可能なデータベース (GISAID に連携した国立生物工学情報センターなど) のもとで、開かれた仕方で共有されるべきだ。

第四章 ウイルスゲノムと臨床データおよび疫学的データを組み合わせ、ゲノムの追跡を行うための枠組み

この章では、SARS-CoV-2 やその他の病原体の感染や進化、病気の治療、感染予防についてウイルスゲノムのデータが役立つものになるためには、どんなタイプの臨床データ、疫学的データと組み合わせる必要があるかについて述べる。

ウイルスの感染や進化、そして臨床疾患についての諸考察

・包括的なデータ収集に関する考察

まず、臨床データや疫学的データの収集は非常に重要だ。たとえば、サンプルが得られた時間的、地理的情報を活用することで、病原体の拡大の様子や拡大経路が判明したり、クラスターやスーパースプレッダーの予測・特定などが出来たりする。また、実際の病気の症状や重篤度などの情報によって、ウイルスの進化が明らかになったりする。そのほか、抗ウイルス薬の投与やワクチン接種などに関する情報も、SARS-CoV-2 がそうした対抗策から回避するようにして進化していく様子を解明するのに役立つ。

一方で、得られる疫学的なサンプルが、適切な集団を代表しているということが保証されている必要がある。この問題を解決するためには、単に地理的な観点からの考察だけではなく、ジェンダーや人種、民族集団、住環境、職種などについての考察も不可欠である。様々な立場の人から意見を集めたり、資源の限られた地域（部族の領域や極めて小規模な病院など）の人々がきちんと反映されることを積極的に支援するチームを設立したりすることが必要だ。

・感染

人口動態的なデータをゲノムに関する情報と組み合わせることで、感染の発生源を探ることが出来る。たとえば、ワクチンが利用可能になった後は、仮に新しい感染事例が見つかった場合に、それがまだ大規模感染が続いている外国から持ち込まれたに過ぎないものなのか、それとも国内での感染による（ワクチンによる集団免疫が不十分であることを示す）ものなのか、ということが確定できる。あるいは、どのような場所で、またどのような経路をたどってクラスターや感染拡大が起こっているか、といったことについても知ることができる。

・ウイルスの進化と選択圧¹⁴のもたらす影響

これらの情報やデータは、ウイルスの進化の様子を解明したり、将来どのように進化して

¹⁴ Selective pressures. ある生物のそれぞれの変異に対し、環境的な要因によってもたらされる生存率の差。

いくつかを予測したりすることにもつながる。

たとえば、SARS-CoV-2 の進化を過去の SARS-CoV や MERS-CoV などと比較すると、興味深い相違が見つかる。まず、SARS-CoV-2 は単系統で発生したのに対し、過去の二つでは独立した変異からの複数の流入が起こっている。また、SARS-CoV-2 は症状が顕在化する前から高い感染力を持つのにに対し、過去の二つでは感染が起こるのは主に発症後である。そのため、SARS-CoV-2 は無症状感染の事例が非常に多いという特徴をもつ。こうした違いはどういった遺伝的差異によってもたらされているのか、より強い毒性を持つものはどのような過程で生まれるのか、毒性と感染力にはどのような関係があるのかといった、様々な問いがある。

SARS-CoV-2 も、進化していく中で、様々にその危険性を変えることがあるかもしれない。多様性を増すということ自体が、自然選択による進化の加速をもたらす可能性があるのだ。ウイルスは毒性が高い方にも低い方にも進化する可能性があるので、変異率が高い＝進化が速いということは、それ自体では脅威とならないだろう。しかし、ウイルスは世代交代が速いため、希少な変異が短時間の内に広範な影響を及ぼす可能性はある。

感染力に影響するスパイクたんぱく質の変異については、複数の報告されているが、1月に中国と欧州で特定された D614G 変異体が世界中で広まっている。他にも複数のスパイク変異が報告されているが、いずれも感染は比較的小規模にとどまっている。これらが人間の細胞の受容体とのかかわりにどのような変化をもたらすかはいまだ不明であり、さらなる研究が必要である。

様々な選択圧が、ウイルスを進化させ、その感染の（生物種としての、ウイルスから見た）成功率に影響し得る。たとえば、ソーシャルディスタンスがウイルス分子のエアロゾル中の滞留しやすさに正の選択をもたらすと、ウイルスの滞留しやすさが増してその結果感染の機会が増えると考えられる。その一方で、空気中に滞留するウイルスは自己複製が出来ないため、ウイルスの滞留しやすさと自己複製の速度はトレードオフの関係にあるとされるが、インフルエンザウイルスにおけるようにこのトレードオフ関係が感染した人間の体内でのウイルス量の低下をもたらすかは定かではない。また、抗ウイルス薬やワクチンといった選択圧も、ウイルス RNA のそれぞれ特定の領域に変化をもたらし得る。これらが将来どのように働くかを予測し、リアルタイムでの監視と評価を行うためには、やはりウイルスゲノムのデータと患者にかかわる他の情報との連結が欠かせないのだ。

・臨床疾患

現在までに、ゲノム配列のある変異がある特定の臨床現場における帰結（clinical outcomes. 臨床転帰）やウイルスのもつ特徴の変化をもたらしたりしたという事例は報告されていないが、ウイルスの系統と臨床的な諸特徴とを関連付けて特定しておくことは、ウイルス研究や将来において特別な介入を必要とする患者に益するかもしれない。ゲノムのデータと臨床現場から得られるあらゆるデータを統合することで、疾患の重篤度やウイルス量の変

化に関連する変異や、その他ウイルスがもたらす様々な症状に関連する変異を特定することが可能になる。

データ統合を支援する機会

データは単に統合するだけでは不十分であり、時宜に適いかつ有用な仕方で用いられねばならない。それには、膨大なデータの中から最も重要な部分を判断、抽出して、多くの当事者に利用しやすいように提供することが大切であり、また、そのデータ管理の仕方を様々な状況や情報源に対応できるように柔軟で拡張可能なものとしておくことが大切である。

現在、多くの機関から集められる多種のデータの管理や収集、整理を行ったりする機関は存在しない。こうしたデータ統合を支援するインフラの建設には、諸機関の共同や相互運用性、柔軟性、プライバシーにかかわる問題が付きまとう。たとえば、データは多くの形式で用いられるであろうという点から相互運用性が重要になるし、データの共有と公開にはプライバシーの問題がある。

諸機関の協力・参加を促すためには、データ統合の結果得られる知見や利益がそれら機関にきちんと還元されるような仕組み作りが必要になるだろう。

また、諸機関に対し、システムへのデータ提出がしやすくなるような基準となるデータの統一した形式を作ることも大切だ。これら形式は情報源やデータの種類に応じてある程度種類に幅を持たせるべきである。

インフラストラクチャーに関するニーズ

以前のデータ統合の試みは、ほとんどが小規模のものだった。SARS-CoV-2 や将来の集団発生に対応するためには、これらは国家規模でのインフラに拡大されねばならない。疫学的監視を主任務とする米国疾病予防管理センターには、州レベルの保健部門との繋がりや他国および地球規模でのネットワークとの繋がりもあるが、まだインフラが十分に整っているとはいいがたく、同センターがその資金提供に責任を負うところの国内の諸研究機関の多くは資源不足の状態にある。また、こうした十分なデータ統合を可能にするためには、同センターや他の地域レベルでのデータ統合ネットワークで進行中の試みから様々な知見を得ることが大切である。

・詳細な臨床データとの連結を可能にする、国立 COVID コホート共同研究 (N3C) 内の独立部門のモデル

N3C では、データセキュリティの観点から、SARS-CoV-2 に関するデータの収集を、他から独立した部門の中で行っている。N3C はこの SARS-CoV-2 のプロジェクトの中だけで用いられる識別子を用いてデータの収集・管理をしており、N3C 内のデータとの統合のた

めには、ウイルスゲノム情報あるいはこれら記録へのリンクなどが（公開されたデータベースへのアクセスとは別個に）N3C に保管される必要があるかもしれない。

ただ、N3C に保管されているデータは、個人の再特定化を避ける観点から種々の制約を持つ形になっているので、接触追跡のような重要な疫学的調査活動は支援されていない。しかしそれでも、N3C に SARS-CoV-2 のゲノムデータを含めることで、臨床現場で重要なような情報の蓄積が可能になるかもしれない。

・ SARS-CoV-2 データの統合のためのインフルエンザ用のインフラストラクチャーの利用

SARS-CoV-2 に関するデータ統合には、インフルエンザウイルスのゲノム変異を追跡するためのシステムを活用することが可能かもしれない。現在、委員会が SARS-CoV-2 のゲノムデータ使用に関して設定している目標と同じことをインフルエンザウイルスに対して行っているシステムが既に存在しているのだ。こうしたシステムには、全 50 州やアメリカのその他自治州にあるおよそ 100 の公衆衛生機関および 300 の臨床研究機関が既に参加しており、インフルエンザウイルスの動向監視に有益なデータの提供・収集・分析などを行っている。また、患者からの臨床データを収集するネットワークも充実している。インフルエンザウイルスの患者の情報から得られた、リスクを高くする医療条件（心血管疾患、慢性肺疾患、免疫力の低下した状態、肥満や妊娠など）は COVID-19 に対しても類似している。

・ 診療情報の共有（Health Information Exchanges）

現在のアメリカでは国家全土で病院や臨床医による電子的な医療記録の仕組みが一般化し、また、診療情報の交換や共有を行うネットワークも多くの医療機関に導入されている。加えて、電子的な診療情報（電子カルテ）の共有を臨床医療目的で行う制度も設立された。

その後も、国家規模での保健情報の共有を支援するネットワークの設立へ向けた取り組みや、官民のパートナーシップなども開始されている。2020 年 4 月 1 日には保健福祉省の保健情報技術国家調整官事務局が、医療機関や電子医療記録の提供者、情報取引機関に情報の共有を妨げる行為を禁じる規則を制定し、これは同年 11 月 2 日に発効予定である。この国家規模でのネットワークはまだ実現されていないが、公衆衛生担当部局や研究者たちに SARS-CoV-2 や COVID-19 についての有益な情報を生むものとなるだろう。ただ、このように相互運用性を備えた保健記録のシステムを実装することは、個人の健康に関するプライベートな情報が広く共有されることを意味し、これは個人のプライバシー権の侵害のリスクだけでなく、公共の信用を失わせ、結果として将来の公衆衛生に重要なデータ収集の精度を低下させるおそれもある。プライバシー保護のための様々な取り組みとシステムの更新が必要になると思われる。

・ 参加型サーベイランス（Participatory Surveillance）

参加型サーベイランスとは、個人個人が病気の症状を自発的にクラウド型のシステムに

報告することで、共同体レベルの監視を実現するシステムだが、既に複数例が世界中で利用されている。これらは、民間の衛生機関や研究機関へ提供された疫学的データを、従来の歩哨型サーベイランス¹⁵より多く収集する。たとえば、インフルエンザ流行の場合は、感染者の多くが毎年医療ケアを受ける訳ではないので、多くの自己報告システムによって情報が収集されている。ボストン小児医院の Flu Near You システムはその一例であり、ウイルス研究で集められた疫学的データと、インフルエンザなどの疾患の診断検査を組み合わせたものだ。インフルエンザの恐れがある症状を申告した個人には、自宅での検査キットが与えられ、サンプルが採取・収集された後、検査の結果がシステムに保管された症状のデータと照合されるという仕組みになっている。この自宅キットを用いたやり方は今回のパンデミックにおけるウイルスゲノム解析にも利用され得る。

COVID-19 パンデミックでも、多くの共同体に根差した監視システムが登場してきている。たとえば Flu Near You システムは SARS-CoV-2 に対応できる COVID Near You システムに発展させられた。これら既存の COVID-19 症状申告、接触追跡のシステムに、ゲノム解析の手法を加えられる可能性がある。

・パートナーシップ、組織間の協力、組織のもつ能力に関する諸考察

様々なセクターやレベルでの研究機関同士のパートナーシップを促進することも、国家規模でのデータ統合を実現するために大変重要になるだろう。こうした試みは、人口集団全体を表現できるように多様な種類の研究機関の間で行われなければならない。研究機関に加えて病院や臨床機関とのパートナーシップも重要な情報源になるし、それらは特に非都市圏やアクセスの限られた地域で力を発揮する。担当機関の能力が不十分であったりサンプル活用に様々な障壁があったりして中央の機関による検査が行われられないような状況下で、パートナーシップによってそうした地域のデータを得て、活用できるようになることが期待される。

また、作られたデータの現場での使用可能性や現場の機関がもつ能力に関する諸考察なども、パートナーシップの重要な諸側面である。たとえば、諸機関への密接な支援や協力、能力の構築などは、感染症集団発生時に現場の諸機関で経験されているストレスを軽減するのに有効である。これらの機関の多くには、煩雑な書類手続きや時代遅れの通信技術のせいでより広範なデータベースに報告されていない多くの患者についての有益なデータがあるので、そうしたストレスを緩和し、より簡便なデータの収集、分析、共有を可能にする手法を用いることが極めて大切である。また、パートナーシップは、そうしたデータがそのエンドユーザー（現地の諸医療機関など）に益する仕方で運用されることを保障することにもつながるだろう。たとえば、ゲノムデータには様々な利用可能性があるが、臨床データや疫学的データは特に患者のケアや公衆衛生的な帰結に重要であろう。

¹⁵ 患者の報告に頼るのではなく、監視に当たる人員やシステムなどが積極的に情報を採取するシステム。

・ゲノム疫学を目的としたスタッフの人数・能力の拡張

テクノロジーが発達しても、多分野に精通し高度に訓練された人材によるチームは、データ統合を考えるうえで依然として重要であり続けている。2016年から、ブロードインスティテュートはマサチューセッツ州の公衆衛生機関及び疾病予防管理センターと、地域レベル、州レベルでの研究員に対する訓練員の養成プログラムを通じたゲノム解析能力の向上計画を共同で進めている。こうしたプログラムも、国家レベルでの同様の施策の参考になり得るだろう。

結語

上記したような様々なデータの統合や運用に関するインフラを整備する試みは、地域レベルなどで、あるいは限定的には行われているものの、国家規模で、また集中的に実現された例は未だない。今回の COVID-19 でそうしたインフラの重要性が露になった。現在あるシステムを参考に、国家レベルでの有益に働くデータインフラを計画、実現し、また、十分な資源やスタッフを供給することが極めて重要だ。

提言 2 :

アメリカ合衆国保健福祉省は、ゲノムデータと臨床データ、疫学的データ、その他 SARS-CoV-2 集団発生を含む公衆衛生的な対応に必要な複数の情報源から得られた関連する諸データを、精密、効率的、安全に結合させる能力を備えた、現行のプログラムインフラの上に構築される国家のデータインフラシステムを開発、資金提供すべきだ。また、そのようなシステムは以下のようなことを行うべきだ。

- ・ゲノムデータと臨床データ、疫学的データやその他の関連する諸データを、定期的にデータ収集を行う研究機関にとって過度な負担にならないような仕方で結合することを可能にすること。
- ・データがシステム内で用いられ共有されている際に、個人がもつ、個人を特定可能にするような情報が開示されないままにされることを保証するようなデータ共有の実用的な方法を案出し促進すること。
- ・データの標準化、相互運用可能性、柔軟性、実践的な連結を目指すという姿勢に基づくこと。それらには、将来導入されるかもしれない全国患者識別子¹⁶についての考察も含まれる。
- ・そのようなシステムを創出するために必要なデータだけでなく、そうしたデータの収集と

¹⁶ National Patient Identifier. アメリカの全国民に医療目的で各人に固有の番号を割り振り、それによって全ての保健機関・衛生機関で患者のデータを共通して管理するもの。プライバシーへの懸念等から未だ導入には至っていないが、導入の是非について様々な議論がなされている。

解析を支援する機構への資金提供——それらにはデータサイエンスと感染症疫学との中間領域における「データラングリング¹⁷」についての制度化された教育の促進も含まれる——も考慮すること。

- ・データが実際どのように使われるかについての予測などを含む再評価を毎年定期的に行って、改善に必要な課題を特定したり、手続きの改善（季節性インフルエンザの年間リスクを評価している既存の合衆国のインフラにもとづけば、診療情報の共有のしやすさと対象範囲、およびその他の諸テーマを改善するといった取り組みがありうる）を促進したりすること。また、データ提供を行う私的なヘルスケアシステムや、データ収集を行う州や地域の公的保健機関等の、様々なスケールでの支援を行うことが出来る機関を、現在進行中のシステム開発と評価に含めることを保証すること。

¹⁷ Data wrangling. データの分析を可能にするため、データの整理や加工を行うこと。データ分析過程の多くを占める重要な作業。

第5章 ガバナンスと規制に関する考察

現在米国では、ウイルスゲノムのデータやウイルスのサンプルの共有を規制、命令する連邦法や州法は存在しないため、それらの共有は自発的かつ自由に行われている。一方、臨床データや疫学的データに関しては、それらを規制する連邦法及び州法が存在している。本章では、米国と他国におけるこれらデータのガバナンスと規制に関する考察を扱う。

米国は連邦国家であり、公衆衛生に関する法的権限は主に州政府が保有しているので、現在連邦法は臨床・疫学的データの共有を要求してはいない。現在すべての州や領土が定期的に米国疾病予防管理センターとデータの共有を行っているが、これらは自発的に行われたものであり、しかもウイルスのゲノムデータではなく疫学的データのみであることが多い。現状のように、データ収集と分析にかかるコストに対する資金援助や、データ共有に関する明確な規制やインフラが欠けていては、データ共有は不完全なままである。

また、プライバシー関連法がデータ共有の障壁となるとしばしばいわれるが、連邦法では、公衆衛生と研究目的でのデータ共有は許可されている。法の誤解と混乱、法的アドバイスを求めることへの躊躇がむしろデータ共有の障壁となっている。連邦法が何を許可しているかについての明確なガイドラインが必要だ。

また、それに加え、様々な障壁を取り除くためには、国家レベルでの指導力とガバナンスが必要だ。強力な権限を持った機関が主導して、効率的なデータ共有実現のための制度作りが進められねばならない。

連邦制がもたらす諸障壁と諸機会

COVID-19 のようなパンデミックによる、国家規模の州を跨ぐ脅威に対応するには、連邦制という制度自体が持つ様々な障壁があるが、ウイルスのゲノムデータの共有の必要性は、各州の純粋に自発的な取り組みから、連邦政府権限による共有に移行することで、そうした障壁を取り除く機会となるかもしれない。

国際的なデータの共有についての諸障壁

国際的な公衆衛生的脅威に対抗するために諸国に課される責務を規定した、法的拘束力のある条約として、国際保健法規（International Health Regulations, IHR）がある。IHR の下では、諸国に対し、「症例の定義、研究結果、リスクの発生源と種類、症例数と死亡数、疾患の拡大に影響する諸条件」を含む、国際的な関心の対象となる潜在的な公衆衛生的緊急事態についての「時宜に適い、正確で、十分詳細な公衆衛生情報」を WHO と共有する責務がある。明言はされていないが、そこに挙げられている「公衆衛生情報」には、ウイルスゲノムの情報が含まれると解釈され得る。少なくとも COVID-19 への初期の対応で示された

ように、国際的な関心の対象となる潜在的な公衆衛生的緊急事態を引き起こし得る病原体のゲノム情報を共有することが期待されていると言える。

現在 IHR はウイルスゲノムの情報の共有への明示的な責務を定めていないが、今回のパンデミックを受けて改正が為され、そうした条項が追加される可能性もある。加えて、他の国際会合でも、将来におけるゲノム情報の共有が含意されているし、多くの（30 以上）国では既にデータ共有に関する法制化が進んでいることを考えれば、国際的に同様の制度作りが進められる可能性はかなり高いだろう。

また、IHR においては、あらゆる情報を機密かつ匿名にすることが求められる。あるリスクの「評価と管理」に不可欠な場合に限り、例外的に個人データの開示を許容しているが、その場合はそれらデータが公正かつ合法的、目的に適った仕方で扱われることが保障されねばならないとされる。また、それらデータは必要以上に長期間保存されてもいけない。そして IHR は、プライバシー権や差別されない権利などの人権に適った仕方で運用されねばならない。

推測される、また明確にある国内の法的障壁

連邦法では、ウイルスのゲノム情報の共有に対する責務は無いのに対し、臨床データや疫学的データの共有に関する国内法において、様々な法的障壁が推測されるし、また明確に存在している。これらの法の適用と潜在的な利害関係者との関係が明らかにされなくてはならない。以下では、データ共有に関する障壁としてよく紹介される二つの連邦法（医療保険の相互運用性と説明責任に関する法律、およびコモン・ルール）を例示し、そのウイルス情報の共有についての適用範囲を明確にする。

医療保険の相互運用性と説明責任に関する法律（THE HEALTH INSURANCE PORTABILITY AND ACCOUNTABILITY ACT, HIPAA）

HIPAA は本報告書で論じられている様々な研究機関のほとんどすべてにかかわるものである。（HIPAA の諸規則の一部である）HIPAA プライバシー規則は、特定可能な個人にかかわる保健情報（protected health information, PHI）の利用と開示のための規則を定めている¹⁸。これに対し HIPAA セキュリティ規則は、あらゆる電子的な PHI について、安全な形で保管、通信されることを要求するものだ。

・公衆衛生機関への PHI の開示

まず、HIPAA の対象となる公衆衛生機関への PHI 開示は、データの対象の同意や許可な

¹⁸ HIPAA は生体試料そのものを扱っていないが、生体試料にかかわるあらゆる PHI を規制し得る

どを求めずとも許されることが明文化されている。また、商業目的の関係者が公衆衛生機関に PHI を開示する際も、その提携協定が許す限りであれば認められることが通常である。加えて、保健福祉省の公民権局が先日、商業関係者に公衆衛生目的の COVID-19 関連情報の開示を、大幅に認める執行裁量権¹⁹を行使し、これは商業関係者による公共機関へのデータ収集支援を後押しするものになった。

PHI の利用、開示は、HIPAA プライバシー規則が定める「最小限度の必要」基準を満たさねばならない。すなわち、諸機関は「データ利用、開示、要求で意図されている目的を達成するのに必要な最小量の PHI のみを」利用、開示、要求するために「合理的な努力」をすることが要求されている。この基準を満たすものを具体的に定義することが諸機関には求められているが、より上位の機関の一元的な要求に依拠することも認められているので、COVID-19 目的のための諸機関のデータ開示は可能だろう。

特定可能な PHI の開示は HIPAA で認められているが、それでも多くの機関は、セキュリティ上の配慮から（あるいは単に HIPAA の誤った理解のせいで）個人の特定を可能にするような要素を廃した情報を開示することを望んでおり、より特定を難しくした PHI も使用される。情報を受け取る側の機関は、許容される情報利用、開示を規定し、個人の再特定や対象となった個人への接触を禁じた情報利用協定に従わねばならない。

また、HIPAA プライバシー規則に則って「匿名化 (de-identification)」されたデータは HIPAA の対象から外れ、無制限の利用、開示が可能となるため、諸機関はしばしばデータ開示の際に、（許可されているため必要ない筈の公衆衛生機関への開示の場合にさえ）HIPAA の匿名化基準を満たそうとする傾向がある²⁰。

・法により要求される PHI の開示

また、HIPAA は、PHI の開示が他の州法、連邦法によって要求される場合には開示を認めることを明文化している。たとえば、ある州が SARS-CoV-2 関連の情報開示を命令すれば、対象の諸機関はそれらに合致する限りにおいて開示が許される。なお「最小限度の必要」基準はこうした場合の開示には適用されない。

・研究 (research) 目的での PHI の利用と開示

HIPAA は「研究」目的での PHI 利用と開示も許容している。「研究」とは、「研究開発や試験、評価を含む、一般化可能な知見に寄与したりそれらを発達させたりするために行われ

¹⁹ Enforcement discretion. 警察官や各種行政機関等の法執行機関が持つ、法を執行したいと思う対象を選ぶことが出来る権利。

²⁰ HIPAA は匿名化の方法として、18 のカテゴリーから成る個人を識別する要素を除去することを要求する「セーフハーバー方式」と、統計学の専門家に要求してデータの受け取り手が対象を再特定する可能性が低いことを確認してもらう「統計学/専門家方式」を定めている。

る体系的な調査」と定義されている。ただ、研究目的での PHI 開示には通常、データの対象となる個人による前もっての承認が必要だが、以下の三条件全てが満たされる場合には、プライバシー委員会や治験審査委員会によってその必要は放棄 (waive)、変更 (alter) されることが可能だ。

- ・ PHI の利用・開示に、個人のプライバシーへの最小限度のリスクしかない場合
- ・ その必要が放棄、変更されない限りその研究が実行不可能である場合
- ・ PHI へアクセスし、利用しない限りその研究が実行不可能である場合

・ 公衆衛生と研究の区別

HIPAA は「研究」のための利用の定義をしている一方で、何が「公衆衛生」のための利用に当たるのかを定義していないので、COVID-19 関連のデータ開示の際にどちらの規則に従うべきかという混乱をもたらし得る。しかし、HIPAA のプライバシー、セキュリティ規則に監督権を持つ公民権局が、いずれの目的であれ、公衆衛生機関への開示はプライバシー規則のもとで許容されるというガイダンスを示している。このため、HIPAA の研究条項は公衆衛生機関以外の機関にかかわる「一般化可能な知見」のための COVID-19 関連情報の利用や開示だけを統括するものと言えるだろう。

同意なしに行ってよいもの

- ・ **公衆衛生機関への全ての**情報開示
- ・ 「匿名化」された情報開示
- ・ 他の法により要求される情報開示
- ・ 公衆衛生機関以外の機関への研究目的での情報開示のうち、三条件に基づいて同意の必要が放棄された場合

同意が必要なもの

- ・ 公衆衛生機関**以外の機関への** (研究目的の) 情報開示

コモン・ルール (COMMON RULE)

コモン・ルールとして知られる、人間を対象とした研究を統括する連邦規則は、個人特定可能な生体試料やデータを用いる研究に適用されるものであり、連邦政府による資金提供を受けた人間の研究を対象とする²¹。コモン・ルールは研究 (HIPAA と同様の仕方で定義されている) のみに適用され、公衆衛生目的の利用と開示には適用されない。

コモン・ルールでは、特定可能な生体試料やデータを用いた研究は治験審査委員会に承認されねばならない。一般的に言えば、研究に先立ってのインフォームドコンセントの取得が

²¹ 連邦政府が資金源でない研究でも、それを行うヘルスケア機関が連邦政府を資金源とする (他の) 研究を行っていて、かつコモン・ルールに従うことに合意している場合は対象となる。

必要だが、これもまた HIPAA と同じではないが類似した基準を全て満たす場合に放棄、変更され得る。その基準とは、HIPAA の三条件に次の二つを加えたものだ。

- ・ (インフォームドコンセントの必要の) 放棄が、研究参加者の権利と福利に悪影響を及ぼさない場合
- ・ 研究後いつでも適正な時に、研究参加者が関連情報を提供される場合

また、もともと臨床目的や公衆衛生目的などの非研究目的で収集されたデータや検体を研究目的に二次使用することは、コモン・ルールの対象とならないという点も重要である。SARS-CoV-2 の場合、その研究に使われるデータは多くがこうしたデータであることが多いので、この免責事項が、SARS-CoV-2 や COVID-19 調査に有益に働くだらう。また、HIPAA の統括下にある情報を使用する研究もコモン・ルールの免責対象である。

また、先日コモン・ルールは新たな免責事項を含むように改定された。それは、新たな合意取得の方法である包括的同意 (broad consent) のもとでの特定可能な生体試料や情報のデータ作成を可能にするものだ。包括的同意のもとで作成されたデータは、その後の研究使用の際に治験審査委員会のレビューを受ける必要はあるが、当該データ作成過程への参加者に再度合意を取る必要はない。データ作成の際に満たすべき条件などは様々あるが、一度それらに則ってデータが保管されれば、その後の研究は治験審査委員会の認める範囲内においてコモン・ルールの免責対象になるため、研究にかかる手間は削減される。

また、そもそも特定可能でないデータや生体試料は、コモン・ルールのもとでは、人間を対象とした研究とはみなされず、その対象外となる。コモン・ルールは HIPAA の匿名化基準を採用していないので、データや生体試料がいつ特定不可能となってコモン・ルールの対象から外れるのかという問題があるが、この場合、研究者は治験審査委員会の判断を仰ぐのが普通になるだろう。

・ 非典型的な情報源と州法

その他の情報源となる機関や研究も、ゲノム情報ではないにせよ、何らかの有益な情報を提供するかもしれない。それらがどれほど情報を共有できるかは、それらを拘束するプライバシーポリシーや規則に依存するだろう。州法も、そうした非典型的なデータソースの共有可能性にかかわりうる²²。

ガバナンスに関する諸提言

提言 3 :

²² 州の保健関連のプライバシー法は、通常は精神衛生や人間のゲノム、薬物乱用治療、生殖と性に関わる健康データなど、よりセンシティブかつ COVID-19 研究にはさほど重要ではないと思われる情報に関わるものであることが多い。

アメリカ合衆国保健福祉省は、重大な公衆衛生的問題と基礎科学的問題に取り組むにあたって、SARS-CoV-2 ゲノム配列情報の使用にとつての、効果的で持続可能で科学的知見にもとづいた指導・統治の機構を確立すべきだ。また、国家的な戦略を発展させ、その戦略の実行の成功に必要な資金提供も保証すべきだ。

- ・この機構の確立に取り組む指導者は、鍵となる諸問題を特定して優先的に検討し、適切な集団を代表するデータを生み出し、また、データ共有の障壁となるものを退けることが十分可能な権限と責任を備えていなければならない。

- ・臨床データや疫学的データに連結させられた SARS-CoV-2 ゲノム配列情報に関する国家戦略は、目標や優先順位、それらを達成する道筋を明確化するように策定されるべきだ。

- ・広く関連する領域の専門家からなる委員会が、臨床データ・疫学的データに連結させられた SARS-CoV-2 ゲノム配列情報についての国家戦略を監督し提言する広い権限と、関連した研究のための利用可能なデータの提供とともに設立されるべきだ。

結語

連邦法は上記したような様々な場合においてデータや生体試料の開示を（しばしば個人への合意を取る必要すらなく）許可しているにもかかわらず、法についての誤解や混乱、法に抵触することを恐れるが故の保守的な法解釈こそが、データ共有に向けての真の障害となっている。当委員会は、高度にセンシティブなデータを提供した責任を問われることを恐れて特定可能なデータの収集に躊躇しているという証言や、共通の患者識別子が無かったり匿名化の手法がデータ連結の妨げとなったりするせいで諸データ間の連結が難しくなっている、という証言を得た。

連邦法は保健福祉省に、長官により緊急事態が発令された際は HIPAA の一部を放棄する権限を与えているが、放棄が許可されている条項は臨床的なケアの提供に大きな影響を与えるため、条項の放棄というやり方では有効なデータ共有を実現することができないだろう。このため、保健福祉省は代わりに、執行裁量権を選択的に行使してさらに大規模なデータ共有を可能にすることで、パンデミックに対抗しようとしている。例えば、同省内の公民権局が、商業関係者に公衆衛生機関との PHI の共有を認めるガイダンスを行った（p. 25 参照）。

このように HIPAA やコモン・ルールの適用に関して保健福祉省から詳細なガイダンスを行うことで、法運用についての不明確さを減らし、データや生体試料の共有を劇的に改善することが可能かもしれない。そうしたガイダンスは理想的には、詳細で、データや生体試料の共有の明確な方法を、分かり易い例を用いながら説明し、常に迅速にアップデートされかつ新たな問題・関心に応答できるものであるべきだ。

保健福祉省は臨床データや生体試料の広範な共有を達成するために、より幅広く執行裁量権を行使することが可能であるが、そのような（ある種非常事態的な措置としての）執行

裁量権は、その発動と終了に関して明確な基準を持たねばならない。

データが公共の信用を得られる仕方で扱われることを保障するためには、公衆衛生諸機関は、プライバシー、情報の機密、安全性、非差別、公民権に関する法律などを遵守しようとせねばならず、また、データ収集にあたって公的な透明性を保持していなければならない。