

2021年12月27日

理研 IMS におけるヒトゲノム情報データの共有に関する基本的な考え方について

理研 IMS ヒトゲノム情報に関する委員会

山本 一彦  
菅野 純夫  
油谷 浩幸  
松田 浩一  
ピエロ・カルニンチ  
古関 明彦  
中川 英刀  
鈴木 治和  
大野 博司

理化学研究所生命医科学研究センター（理研 IMS）は、その前身である理研ゲノム医科学研究センター（2008-2013）および理研遺伝子多型研究センター（2000-2008）の頃から、現AMEDゲノム医療実現バイオバンク利活用プログラム（前 文科省委託事業オーダーメイド医療実現化プロジェクト）「バイオバンク・ジャパン（BBJ）」と協力し、そのゲノム解析に寄与してきた。理研で解析し長年にわたって蓄積してきたデータについて、どのように公開して社会へ貢献していくのか、その方向性と透明性が問われつつある。そこで、理研 IMS が生み出してきたゲノムデータについて、外部識者を含めて方向性を議論し、ゲノムデータの扱いに対する理研 IMS のあり方を検討するために、本委員会が設置された。委員会で行われた議論を踏まえ、ヒトゲノム情報データの共有に関する基本的な考え方を以下に示す。

#### 1. データ公開の現状と今後について

バイオバンクジャパン（BBJ）の公開すべきゲノムデータについては、東京大学医科学研究所にて、その枠組みを策定している。よって、理研 IMS としては、医科学研究所の方向性に沿って、BBJ 及び AMED が定めるデータ公開方針に従ってデータ公開\*することとする。

BBJ 以外のヒトゲノムデータに関しても、いろいろな議論があるものの、研究リソースとなる大規模なゲノム研究プロジェクトでは概ね論文発表時もしくはデータ生成後2年以内に公開\*することが望ましいと国際的に考えられている。理研 IMS としてもこの方向性に対する検討と実行が求められる。

ただし、BBJ の枠組みの中でも、データを産出した機関にデータ公開\*の判断が任されており、理研 IMS が産出したデータの中で現在非公開となっているものがある。特に、生データ (idat ファイルなど) や imputation ファイルがこれに含まれる。これらのデータについて、委員会としては、論文発表から何年も経過している場合には、請求があれば、BBJ が定めるデータ利用者のセキュリティガイドラインに則ったデータ取り扱いの手続きをとった上で、公開\*すべきである。

\*本文章において、データ公開とは AMED 「ゲノム医療実現のためのデータシェアリングポリシー」(<https://www.amed.go.jp/content/000060867.pdf>)で定義される制限共有、制限公開、非制限公開のいずれかを指す。

## 2. クラウドの利用に関して

クラウドの利用は、増大するデータに対応し最新データや解析技術を共有するために、必須となりつつある。海外では、Genomics England や UK biobank などが、ヒトデータと計算環境をクラウド上のプラットフォームで利用者に提供している。国内では、BBJ が第五期に向けてクラウド利用の検討を開始しているものの、現状ではまず第一歩として機関外サーバの使用を認める方向である。複数のゲノムセンターで構築されるアカデミッククラウドでのデータセンター「CANNDs」を AMED が構築、国立情報学研究所が学認クラウド導入サービスを提供するなど、徐々に動いているものの、セキュリティ面、運用面での安定性をどのように担保するかなど、国内でのクラウド利用にはまだしばらく議論があろう。また、国外とのデータのやり取りについても、問題が提示されている。

ゲノム科学の将来として、常にアップデートしたデータを提供し、そのデータを用いた解析を行う計算環境も利用者に提供するようなクラウドプラットフォームが構築され、多様な利用者がそのプラットフォームを利用しながら速やかに研究を推進し、社会や医療に貢献していくことが想定される。理研 IMS としても、個人情報の秘匿性とサイエンスにおける Openness という双方の重要性を認識しつつ、BBJ や他機関と協力しながら、人類の健康、疾患克服に貢献するためにどのようなデータ公開が最適か、検討していくことが求められる。

## 3. その他

上記二点を念頭におき実現させていくため、理研 IMS 内に委員会等を設置し、具体的対応を進めることとする。