

ゲノム研究バイオバンク事業  
「利活用を目的とした日本疾患バイオバンクの運営・管理」  
2020年度 第2回 E L S I 検討委員会  
議事要旨

1. 日時 2020年9月24日（木）14：00～16：30
2. 場所 東京大学医科学研究所1号館 2-3会議室、Zoomによるウェブ会議
3. 出席者（敬称略）
  - （委員）北澤 京子委員、隅藏 康一委員、田村 智英子委員
  - 増井 徹委員、丸山 英二委員、横野 恵委員（五十音順）
  - （慶應義塾大学）小崎 健次郎
  - （理化学研究所）桃沢 幸秀
  - （バイオバンク・ジャパン（BBJ））
    - 古川 洋一（東京大学医科学研究所臨床ゲノム腫瘍学分野）
    - 村上 善則（東京大学医科学研究所人癌病因遺伝子分野）
    - 森崎 隆幸（東京大学医科学研究所人癌病因遺伝子分野）
    - 松田 浩一（東京大学大学院新領域創成科学研究科）
    - 鎌谷 洋一郎（東京大学大学院新領域創成科学研究科）
    - 井上 悠輔（東京大学医科学研究所公共政策研究分野）
    - 永井 亜貴子（東京大学医科学研究所公共政策研究分野）
    - 神原 容子（東京大学医科学研究所公共政策研究分野）
    - 山下 恭司（バイオバンク・ジャパン事務局）
    - 西久保 祐輔（バイオバンク・ジャパン事務局）
    - 小原 好美（バイオバンク・ジャパン事務局）

4. 要旨

（1）BBJ由来のデータの非制限公開について

- ・ 松田から、BBJデータ共有方針について、説明があった。データ公開は、個人別のゲノムデータについては制限公開もしくは制限共有、集計データ、頻度情報、統計値などは非制限公開としている。
- ・ 約18万人のSNPアレイのデータ、そして、ホールゲノムシーケンスについては、1,000人もしくは2,000人、合計で3,000人の一部を制限公開、一部を制限共有で、研究者が使うことができ、20件前後の利用されている。
- ・ 統計値は様々な疾患関連の情報が、非制限公開しており、誰でも使える。NBDGから公開し、ダウンロード数が1万3,000件となった。
- ・ 理研が構築されたJENGERでは、1万5,000件がダウンロードされている。非制限公開の情報は、研究者に利活用されている。

- ・ TOGOVARから、これらのゲノムデータから作成された頻度情報が公開された。GEMジャパンのホールゲノムシーケンス約8,000人のうち3,000人がBBJ由来であり、約9,000万か所のバリエーションの頻度情報が公開されている。他にホールゲノムシーケンス約9,000万か所、SNP18万人、90万か所のSNPの頻度情報も公開されている。NBDCにも18万人の頻度情報を47疾患、男女別のデータを公開予定にした。
- ・ 慶應義塾大学の小崎健次郎氏から、がんの臨床情報データの非制限公開について、説明があった。アメリカのClinVarのデータベースのようなものを日本で作る活動をしている。遺伝性腫瘍の患者に高頻度で認められるバリエーションについて、プライバシーを侵害せずに非制限公開できないか。
- ・ 単一遺伝子病の遺伝子変異は患者に認められ、健常人にはない。多因子遺伝病では、どちらにも特定のバリエーションを認める。遺伝性腫瘍は、変異があっても発症しない人もいる。頻度情報は、バリエーションの意義づけに役立つ。
- ・ バリエーションの定義を参照配列と違う部分は全てとすると、エクソームの解析で、正常と思われるものと病的であるもの、それぞれのデータベースが必要で、病的と思われる変異バリエーションのデータベースを作っている。
- ・ データシェアリングポリシーは、制限共有、制限公開、非制限公開がある。研究であれば、共同研究のチームでの制限共有が倫理上でもしっかりしている。制限公開は共同研究の枠組みは要しないが、データの使用の希望者には提供する。非制限公開では、倫理的な枠組みを、国内のゲノムデータを扱う医療機関が、最初からネットワークを組むのは難しい。データはソフトウェアに組み込んで、非制限公開で、ダウンロードして使えると医療現場で使われるインパクトが桁違いに大きい。
- ・ MGeNDが来年度から厚生労働省本省の事業として、比較的定常的・安定的に使われることが決まった。HGMDが、商用の病的バリエーションデータベースであるが、日本人のデータは薄いのが現状である。どのぐらいの粒度であれば非制限公開が可能か。
- ・ 鎌谷から、ウイルス由来のゲノム配列の非制限公開について、説明があった。理化学研究所の寺尾チームリーダーとニック・パリッシュチームリーダーの代理で説明があった。内在性のHHV6ゲノム、ヘルペスウイルスを解析して論文を作成した。ヘルペスウイルスゲノム配列は、制限公開されているが非制限公開にできるか。ヒトゲノムにあるヘルペスウイルス配列の公開は、海外で積極的に行われ、日本の研究も公開が期待されている。ヒトゲノム上に導入されたヘルペスウイルスの配列の中には、日本人由来の配列も公開されている。

(委員からの意見・応答)

- ・ できるだけ研究成果が広く発信され、活用されるよう公開するのがよいのではないかな。個人情報の保護を担保し、以後の研究に使われるようにしていけばよいのではないかな。全般的に方針には賛成する。
- ・ 非制限公開では、このバリエーションのある人は、乳がん、女性という情報を付け、それ

以上の情報を使いたい場合は制限公開と、バイオバンクから臨床情報を提供する。診断年齢、既往歴はない、乳がんと膵がんの家族歴など、情報を増やすほど個人が特定される。小崎先生の基礎疾患のデータに沿えば、性別、診断年齢、既往歴、家族歴はあってもよい印象はある。

- ・ 今回の議論に基づいて変更案を考えたい。
- ・ 遺伝性腫瘍を扱う立場では、このような情報は役立つ。がん患者は生命保険に任意加入し、手術や入院給付金の請求時に、よくあるがんならば個人を特定できないが、希少がんとわかれば、支払いのトラブルが起きないだろうか。
  - ・ 病歴情報、遺伝情報、併せてこの公開範囲を広げた場合に保険会社で把握された場合、本人が特定されて、保険金請求のときに不利益を被るのではないかという指摘でよいか。
  - ・ 特定はできないと思うが、怪しまれてすぐに給付金が出ないということが起きないだろうか。
- ・ バリエーションデータベースの件だが、候補はデータベースの情報が集約されるべきという中で、どのような位置づけになるのか。
  - ・ 全体は、難病、遺伝性腫瘍も併せてということになる。1つの大きなファイルとして、カタログとして、フィルターに使うイメージである。
- ・ これはいろいろなソースから集めたものを統合し、そのソースになる部分は、ある程度は決まっているのか。
  - ・ 稀少疾患はAMEDの事業で、文献、研究班の成果などが集まっている。ゲノム医療は、最終的にはバリエーションをどう判断するかであり、リソースは大きいほうがよい。
  - ・ ソースがどれくらいのバリエーションがあるかによって、その辺のリスクというのは変わると思った。
- ・ 将来的にソフトウェアに組み込まれた形で使われるようだが、商業的なビジネスとの組み合わせのあたりを教えてほしい。
  - ・ 米国のClinVarは、ウェブブラウザで調べる形と、巨大なファイルデータとして、自分たちのソフトウェアに組み込む二通りの使い方がある。ソフトウェアは、アカデミアが開発したソフトウェアの中に組み込む。医療が整備され、ソフトウェアの質の担保の問題があり、医療機器に準ずる形でソフトウェアが使われることが将来的に予測される。医療機器の診断システムの中に参考資料として、企業が開発するものに部品として使われるが、商業的な活動にはならないと思う。
- ・ 出すデータと利用者の範囲は、ある程度考えざるを得ない。登録や所属機関のドメインでログインするなど、最低限の登録システムがあるとよいのではないか。安全管理措置を設けて、粒度の問題を緩和できないだろうか。
  - ・ 粒度の段階によって制限すべきである。データファイルは、ポータビリティを持って運用されることが大きな境目になる。このデータベースのファイルと、他の

ファイルなどを統合して使われることが国内外で起きる。ダウンロードに制限をする使い方では難しく、広がり狭いと思う。懸念が解かれる粒度で閾値を設けて、公開するほうが医療的なインパクトは強い。

- 一旦ダウンロードしてしまうとトレースできない。利用者の最初の段階以外は、トレーサビリティは確保できない。
  - バリエントを単発の情報として公開するので、他のゲノムに付随しているものではない。
- TOGOVARの場合、nが50以上であれば同定されないのを1つの目安としており、参考になる。年齢を年代、5歳区切りのような工夫は、粒度を細かくせずに済み、利用者にとってメリットになり得る。
- ゲノム医療の分野でAIのソフトウェア機器を開発する際には、診断や治療の方針を示すもとなるデータは、開発者が集めるのか。機器メーカーが集めるのではなく、ゲノム研究あるいはデータベースのデータを利用して機器を作製するのか。ゲノムの場合、必要とするデータが莫大になるので先行する研究で収集されたデータを活用するのか。
  - ClinVarの公開のデータベースなどを新たな機器の機器承認の申請の材料にしてよいというガイドラインが2年ほど前に出た。一般に使われている公共のデータベースを参照にして機器申請をする方向だと理解している。
- 松田から、BBJデータ共有方針の案について、説明があった。何を非制限公開データにするか、現状では、集団としての頻度情報、統計値、解析によって得られるようなもの、個人々別のデータは制限公開としてきた。研究者のニーズ、医療応用を考えると一定の情報は非制限公開にしたほうがよい。個人特定ができない配列情報は、非制限公開としてもよいのではないか。
- 付随する臨床情報について、「登録対象疾患の有無、及び疾患subtypeに関する情報。登録時年齢、性別情報」、「但し非制限公開の対象は、ゲノムデータ及び不随する臨床情報から個人特定が困難であると考えられる場合に限る」としているが、診断時年齢、既往歴、家族歴、部位等の情報を追記する。
- 公開するゲノム情報は、1か所の塩基情報なのでクリアする。不随する臨床情報をどの程度つけてよいのかどうかは論点になる。
- ウイルス配列に臨床情報はない。これが百数十キロ、十数万塩基の情報が個人同定不可な個人識別符号に該当するのか。

(委員からの意見・応答)

- 病的バリエントに関して明確になっている件数はどのくらいか。病的バリエントの同定が手間のかかるものだと思っている。乳がんはBRCA1/2と思うが、病的バリエントについて公開をするのか。

- ・ 乳がんはバイオバンクで7,000人解析した。対照群で11,000人解析をBRCA1/2を含めて11遺伝子解析した場合、全部で大体1,800個のバリエーションがあり、200~300個くらいが病的バリエーションと分かる。登録するデータは、病的バリエーション、病的ではないもの、判断がつかないもので分かれる。臨床情報を参照し、似ているサブタイプから病的と判定するのではないか。
- ・ 稀少疾患でも、ゲノム情報は必要とされている。データがあるというのを個別のバリエーションの情報があるかないか、どういう病気なのかが大切だと理解している。それで診断がつき、その人が不利益を伴うような情報というより、評価をされていないバリエーション、遺伝情報をどのように使っていくのかを研究者や臨床家が使えるデータにはならないのかということが論点ではないか。
- ・ 広い範囲がどの程度か気になる。バリエーションの数はたくさんあり、そこから病的バリエーションを探すのに苦労しているのが実感なので、閾値をどう設定するのか。
- ・ BBJで閾値を定めることが目的ではなくて、疾患研究として対象としている場合には、閾値、重要性、病原性の評価はするが、ほかの疾患や研究者、臨床家がどのように使うかを考えると、情報としては非常に重要である。
- ・ バリエーションの頻度情報も合わせて公開されるのか。
  - ・ このバリエーションは疾患群で何人、コントロールで何人と公開している。
- ・ がん患者で特定のバリエーションが病的かどうか迷う例がある。それが公開されたときに、問い合わせが来ることがある。アメリカではラボ同士で積極的に意見交換が行われ判断を揃える動きになっている。それは誰が対応するのか。
  - ・ イギリスからの問い合わせに対応した。個人から来ることもある。
  - ・ 「個人同定が困難であると考えられる場合に限る」のところで、1万人ぐらいいる乳がんであれば、多少項目数が多くても個人同定には至らないと思うが、疾患数500人ぐらいのがんの場合に項目数を増やすと、個人同定のリスクが高くなる。
- ・ 利用者の登録はしてもよいと思う。トレーサビリティは難しい。入り口を記録しておいてはどうか。非制限公開で全然バリエーションもない計画なのか。
  - ・ ClinVarはユーザー登録なしで使える。臨床の感覚では、誰のアクセスか担保すれば、詳しく出せるのではないか。
  - ・ どのぐらいの情報が1人につくのか。単一遺伝子疾患ならば、1か所病的バリエーションだが、ある程度の閾値を決めて出さないと意味がない。病的バリエーションのBBJのデータが使えるのか。
  - ・ 非常にコモンなバリエーションは載せないで、頻度的には1,000人に1人のようなものとするので、組み合わせても個人同定されないものを予定している。
  - ・ ClinVarのように遺伝カウンセリングをされている先生方が、変異の解釈に多くの情報は要らないが、例えば若年の乳がんの例が他にもあることを理解し、情報共有することを想定している。個人同定されるほどの臨床情報はつけられない。アクセスは登録制なしを第一の方向性に考えている。

- ・ アメリカであれば、遺伝子の情報で社会的な差別を防ぐための法的な仕組みが準備されつつあるが日本ではどうか。
  - ・ 具体的には何も進んではいない。ゲノム医療、ゲノム研究のELSIについて、京大の小杉先生の研究班で議論しようという話はある。
  - ・ 日本版GINAのような感じか。
  - ・ そこまでの話ではない。
  - ・ 議員立法で作ろうという動き3~4年前から、月1回ぐらいの勉強会を開いてやっていたが、まだ成立していない。
- ・ Cowden症候群は少なく、P10の病的バリエーションがある人は有用であり、載せてもらいたい。入口にバリアを設けなくても同意した人が使えるようにし、何かあれば訴えると書き、レアな情報でも出すことはできないだろうか。情報共有は大事で制限したくないが、不利益を被る人がいないようにしたい。
  - ・ GINAは適用範囲が狭い。遺伝情報に基づく差別はよくないというような法律が設けられれば、実効性があると思う。
- ・ データベースの注意書きとして、悪用したら訴えると書くのは無意味か。
  - ・ 不適切な行為であれば、民事責任追及ということ是可以する。例えば保険会社が利用したとして、保険会社の適切なデータ利用なのかそうでないのかについて、通常の保険であれば、遺伝情報に左右されるのは不適切と思うが、がん保険では保険会社も把握できた情報を全く置いておいてということは、バランスが難しい。そういうベースとしては、法律を設けるとするのは望ましい。
  - ・ データを公開するサイトにおいて、社会的に使うのはよいが、個人の同定などはしないように、それを発見した場合にはしかるべきお願いをすることがあると一筆を書くことは、意味がないものか。
  - ・ 無条件の公開でなく、利用者とデータベースを管理している者との契約を結んで利用してもらう。その契約の中に、そのような利用があれば責任を負っていただくことを、契約の中身として定めておくことは十分あり得る。
  - ・ インフォームド・コンセントのフォームにも、情報が不適切に取り扱われた場合には、このような法規制があるとは書けない状況である。しかし、法整備が進む感じも今のところない。アメリカでも、GINAだけではなく、幾つかの州でも包括的な規制をするための法律が導入されている。
- ・ これについて、我々に異論はない。

## (2) 人口動態調査データの提供について

- ・ 松田から、経緯等について説明があった。続けて、森崎より、今後の対応策について説明があった。

### (3) 病院情報の研究利用について

- ・ 松田から、病院情報の研究利用について、説明があった。以前、個人の地域情報を使った研究提案があった。2008年の理化学研究所の研究では、日本人が本土と琉球の大きく2つのクラスターに分かれることが示された。イギリスの研究で、細かい地域情報を使用し、地域間の遺伝情報の差があることが明らかになってきた。
- ・ 遺伝的な背景の違いが、病気のリスクを検討する上でも重要である。遺伝的な背景の違いを考慮しないと、正確なリスク予測はできないことが明らかになってきた。
- ・ 解析に使うデータは、公開済みのSNPアレイのデータ約18万人でこの情報と病院IDに基づいて作られた地域の情報を使いたい。地域情報には、地方別、北海道、東北、関東ぐらいの粒度の情報で研究利用されているが、市区町村レベルでの情報を使いたいという提案である。
- ・ 個人の同定の問題、地域のスティグマに関係するものについては対応し、懸念点について医療機関の先生方にも説明を行った。
- ・ BBJに参加いただいている医療機関のうち、地域別で分かれている徳洲会に相談を進めている。研究計画書の最終案を作り、11月ぐらいに倫理審査への申請で、承認され次第、徳洲会でも検討して倫理審査に進めたい。
- ・ 地域情報は、病院のID番号を生成した病院の所在地に関する情報（市区町村等の分類も含む）とする。利用資格に、死因情報を利用申請できるのは、申請、承認を得た者に限るという文言を追加することで変更を考えている。
- ・ 地域情報データを利用する場合などは、当該地域への不利益がないように、事前にBBJと協議をする。予後情報データの死因情報を用いた成果について論文発表を行う場合は、事前にBBJに報告し、承認を得ることとする。予後情報データの死因情報を利用する場合は、東大との共同研究として行うことを追加している。
- ・ 地域情報のデータのうち、市町村等の分類の利用については、提供元の医療機関の承認の上、東京大学医科学研究所内の利用に限定し、病院情報に関連する市区町村の情報は外部提供を行わない。また、提供元の医療機関の承認を得ることとする。市町村等の分類の利用については、医療機関の承認が必要であること、東大内で行うことをガイドラインに反映させたい。

### (委員からの意見・応答)

- ・ 研究は多様な観点から進めてよいと思う。提供元の医療機関の了承の上とし、BBJ側は、提供元の医療機関の了承が得られれば問題ない。提供元の医療機関が了承しやすくするために、ELSI委員会の了承も得たことも情報として出し、提供元の医療機関が簡便的な検討で意思決定できるようにしておく、先方にもよい。
- ・ 提供元の医療機関の了承の上というのは、地域特性を考えた場合という条件がついているためと理解してよいか。BBJの判断に任せるという承認をもらえば、それでもよい気もする。個別承認か、包括承認のどちらを考えているのか。個別承認の場合に、

地域差を勘案されているのか。

- ・ 基本的には、包括的な承認であればよいと思う。個々の研究案件について、次の研究者の依頼が来たときに、再度説明して承認を取るのは、手続きが煩雑になる。さらに医療機関の業務が増える。全く知らせずに使用しては、問題が生じる可能性があり、そういうレベルで考えている。
- ・ 異論ないということで取り扱いたい。

#### (4) COVID-19に関係するBBJの取り組みについて

- ・ 松田から、COVID-19に関係するBBJの取り組みについて、説明があった。主には2点である。BBJ参加者の後ろ向きの追跡調査がある。BBJの参加者の臨床情報の更新はできない状況だが、一部の疾患等について、追跡調査をする枠組みを作り、昨年度は糖尿病の追跡調査等を行った。現在幾つかの追跡調査が進められている。7月に東大医科研でのIRBの承認と12医療機関に連絡し、半分以上の協力を得られそうである。東大のIRBが承認されたので、各医療機関で承認後に、後ろ向きの調査の体制に進めたい。
- ・ 2つ目が、国の取り組みにBBJも協力する。予算申請中だが、国立国際医療センターでCOVID-19のバイオバンク構築が進められている。国立国際医療センターがハブになる。多数の検体を収集することが見込まれ、BBJはサンプル保管協力をする。
- ・ ゲノム解析等で医科研の先生方も参加予定である。BBJは、倉庫の血清、組織のバンクの余剰スペースの部分にサンプルを受け入れることで国のプロジェクトに協力したい。
- ・ 診療情報の連携について、BBJの参加者の臨床情報の更新が今できないが、クラウドベースで臨床情報を連携できる仕組みができないかを文科省と相談している。
- ・ メディカルコーディネーターが臨床情報を入力して、年に1回データ解析を行っていたが、クラウドベースの場合は、月1ぐらいで電子カルテからクラウド上にデータを上げられる。システムが構築できれば、COVID-19のような案件に対しても、比較的速やかに情報の収集、ゲノム解析に進めることが期待できる。

#### (委員からの意見・応答)

- ・ BBJ登録者の方でCOVID-19にかかった方がいても、協力機関にかかった方でないとキャプチャーできない。大体何%ぐらいの方がキャプチャーできると予想されているのか。
  - ・ 感染者の現状を見ると、国内で3万人程度と考えると、3,000人に1人ぐらいなので、BBJ全体で考えると、100人ぐらいになると思う。一番大きな病院でも、夏時点で、感染者ゼロという返事であり、どの程度有効かわからない。登録者が関連病院を受診しないとその対象者ではなく、十分にカバーできるかどうかは難しい。
  - ・ 感染症の方がPCR検査をするのと並行してゲノム解読もして研究するとよいだろうが、本プロジェクトとは違うと思う。
- ・ 患者をキャッチするのが難しいという問題は、非常に大きな問題である。デジタル化

されたデータが名寄せでつながらない問題がある。この感染症に関してはブレイクスルーするような動きはあるのか。

- ・ レセプトデータと連携ができればあるが、ハードルは高い。医療機関ベースでも、COVID-19にかかるかどうか分からないという状況を踏まえると、網から漏れる可能性はかなりある。
- ・ クラウドの件、アメリカの臨床では、患者さんの実名入りの検査結果がウェブで返ってくる。アメリカはミリアドのシステムを使って結果を見るが、全部IDをもらったオーダーしたドクターが、自分でホームページを開き、患者さんのデータを印刷するので、センシティブな個人個人の遺伝性腫瘍があるかどうか分かるような遺伝子検査結果も全部ウェブで返している。セキュリティの仕組みがどうなっているのかを参考にしたいと思う。
  - ・ 検討しているのは、クラウドというより研究のための外部サーバー利用について、どういう要件が必要かをNBDCなどに問い合わせている。研究利用の場合に必要なセキュリティを参考にしたい。臨床情報で個人名が出てくると、厳格なものが必要になる。
  - ・ FoundationOneは中外製薬が介している。ミリアドジェネティクスのBRACAnalysisは、SRLがシステムで厚労省の審査を通過している。研究と違うかもしれないが、臨床より安全性は高いはずである。
- ・ クラウドの一番下に利用者があるが、BBJの利用者、これは研究利用かと思うが、どういものが該当するのか。
  - ・ BBJに連絡をして、承認を受けて使う方が該当する。クラウド上にある臨床情報のデータベースにアクセスする利用者というイメージである。臨床情報だけ使うか、ゲノム情報、生体試料を使うかというだけの違いで、窓口はBBJの利用申請を経たものを想定している。
  - ・ 両方ともBBJの試料情報の利用者ということか。

#### (5) その他

- ・ 次回のELSI検討委員会は、12月18日（金）を予定している。

(了)