

バイオバンク（UK Biobank）利用の現状

医薬産業政策研究所 主任研究員 岡田法大

製薬産業とバイオバンクの連携

ゲノム情報をはじめとするオミックスデータと臨床情報における表現型との紐づけを可能とする大規模データとして、バイオバンク研究において集積された情報を創薬へと利用する動きが活発になっている。製薬産業とバイオバンクの連携は2010年代後半から増加してきており、本邦においても、2020年に東北メディカル・メガバンク機構と日本製薬工業協会の連携が開始され、2021年3

月には東北メディカル・メガバンク機構と製薬企業5社による「全ゲノム情報と医療・健康情報の統合解析コンソーシアム¹⁾」が開始された(表1)。製薬企業と連携を行うバイオバンク研究は、遺伝子多型をゲノム全域に渡って網羅的に検索する手法であるゲノムワイド関連解析 (genome-wide association study : GWAS) を積極的に実施している英国の UK Biobank や、フィンランドの FinnGen 等が中心となっており、製薬企業からは

表1 バイオバンクと製薬企業の連携の例（各社プレスリリースから抜粋）

年	製薬企業	連携先バイオバンク／連携内容
2017	GlaxoSmithKline／Regeneron	UK Biobank（英国） 50万人のエクソームシーケンスを行うイニシアチブを発表 データの公開に先駆けて一定期間の独占使用権を合意 後に、AbbVie、Alynlam Pharmaceuticals、AstraZeneca、Bristol-Myers Squibb、Biogen、Pfizer、Regeneron、武田薬品工業もエクソームシーケンスを行うコンソーシアムを別途設立
2017	Abbvie／AstraZeneca／Biogen／Celgene／Genentech／Merck & Co／Pfizer	FinnGen（フィンランド） バイオバンク、大学、医療機関等と製薬会社の産学連携の研究組織に参画 後に、GlaxoSmithKline、Sanofi も参加
2019	Amgen／AstraZeneca／GlaxoSmithKline／Johnson & Johnson	UK Biobank（英国） 50万人の全ゲノムシーケンスを行うイニシアチブを発表 データの公開に先駆けて一定期間の独占使用権を合意
2020	Novartis／Bayer／Roche	Finnish Biobank Cooperative（フィンランド） パートナーシップ契約を締結 2020年にNovartis、2021年にBayer、2022年にRocheが契約
2020	Amgen／AstraZeneca／Biogen／Bristol Myers Squibb／Genentech／GlaxoSmithKline／Janssen Pharmaceutical／Pfizer／Regeneron／武田薬品工業	UK Biobank（英国） コンソーシアムが5万3,000人の血中タンパク質を測定する研究の開始を発表 データの公開に先駆けて一定期間の独占使用権を合意
2021	Abbvie／Pfizer／Biogen	UK Biobank（英国） 30万人の全エクソームシーケンスを利用した疾患とバリエーションの関連性を含むデータベースを公開 マサチューセッツ工科大学とハーバード大学の研究機関との協働
2021	エーザイ／小野薬品工業／第一三共／武田薬品工業／Janssen Pharmaceutical	東北メディカル・メガバンク機構（日本） 10万人の全ゲノムシーケンスを行うコンソーシアムの設立を発表
2022	第一三共／武田薬品工業	東北メディカル・メガバンク機構（日本） ウェアラブルデバイスを利用した共同研究の開始を発表

出所：製薬企業又はバイオバンクが発表したプレスリリースから主要な動向を抽出

1) 東北メディカル・メガバンク機構、プレスリリース、<https://www.megabank.tohoku.ac.jp/news/45027>

ゲノム情報に関連する大規模データの需要が高いことが連携先のバイオバンクの特徴や、連携の内容から窺うことが出来る。バイオバンク研究の実施数は世界的に増加してきているものの、製薬企業と連携を行うバイオバンクは現時点では一部のバイオバンクに限定されており、各バイオバンクの試料・情報の商用利用に関する方針や、国外からの情報へのアクセスに関する方針の違いが影響していると考えられる。製薬企業は疾患バイオバンクだけではなく、一般集団バイオバンクとの連携も進めていることから、特定疾患に関する情報に限らず、一般集団の情報も必要としていることが分かる。今後、バイオバンクの大規模なデータは創薬研究において重要な情報源となる可能性があり、製薬産業からも、学術機関や医療機関が中心となり実施されているバイオバンクの試料・情報を用いた研究の実施体制や、蓄積された試料・情報の有用性を精査し、研究への参画方法の検討を行っていく必要がある。本稿では、それらの検討の参考になる情報に関して、公開されている学術論文等を基に、UK Biobank の現状を中心に調査を行った。

データアクセスの方針

バイオバンクに試料を提供する研究参加者は、試料提供時に自身の生体情報が研究に利用されることに関する同意を行うことが一般的であるが、利用可能となる研究の範囲や利用者の制限は、各バイオバンクで異なる方針がとられており、情報へのアクセスが可能となる研究者を自国内の研究者や公的研究機関に所属する研究者に限定しているバイオバンクも多く存在する²⁾。最も多くの研究成果の創出に貢献している UK Biobank は、当初より蓄積した試料・情報を可能な限り広く利用できるようなオープンアクセスリソースを目指して設立されており、適切な利用目的の申請と試料

移転の契約を行うことで、学術利用、政府活動、慈善活動、商用利用等の目的を問わず英国外からも利用できるように申請を広く受け付けている³⁾。対照的に、米国の一般集団バイオバンクである All of Us では、現時点では営利団体や国外の研究者に対してデータへのアクセスを許可しておらず、類似した試料・情報を集積しているバイオバンクの間でも情報の共有方針の違いが存在する⁴⁾。これらの利用方針の違いは、同意取得文書における試料・情報の提供範囲に関する記載内容や、自国内におけるゲノム研究の国際競争力確保の目的、大規模なゲノム情報の共有プラットフォームの構築の可否、個人情報保護への懸念等の様々な要因に起因する。UK Biobank では、同意取得文書に国外の研究者や営利企業も、試料・情報が利用可能となる旨が明記されており、製薬企業と連携してオミックスの解析を大規模に実施している。また、それらの大規模データへのアクセスもクラウド上に構築された専用の解析環境から可能となっており、情報の量とアクセスの容易さも利用頻度の増加につながっていると考えられる。試料・情報へのアクセス範囲を拡大することには、前述のように、情報保護等の懸念点も存在しているものの、UK Biobank の試料が他のバイオバンクと比較しても多数の学術論文の創出に寄与している現状を見ると、研究の実施という側面では、最も有用な生体情報の情報源として大きく貢献していることが分かる（図1）。

情報へのアクセスの方針は、最終的な成果物が掲載される学術論文の著者情報から見ても、研究の実施体制に大きな影響を与えていることが分かる。学術論文を多く公表している UK Biobank、China Kadoorie Biobank、東北メディカル・メガバンク機構の名称がタイトル、アブストラクト、キーワードに含まれる原著論文に関して、学術文献のデータベースである Web of Science を利用し

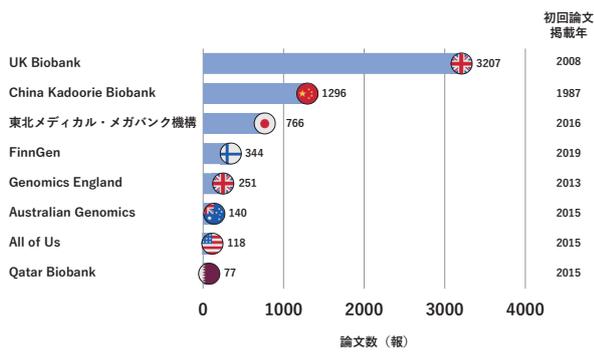
2) 健康・医療戦略推進本部、第2回 ゲノム医療協議会 参考資料3、

https://www.kantei.go.jp/jp/singi/kenkouiryuu/genome/genome_dai2/gijisidai.html

3) Conroy, Megan, et al. The advantages of UK Biobank's open-access strategy for health research. *Journal of Internal Medicine*, 2019, 286.4: 389-397.

4) All of Us Research Hub, Institutional Agreements, <https://www.researchallofus.org/institutional-agreements>

図1 主要なバイオバンクの論文数



出所：論文数（学会発表、知財申請等を除く）を公開している主要なバイオバンクを対象に、各バイオバンクのホームページの情報から作成。

ホームページへの掲載基準や情報の更新頻度は各バイオバンクによって異なる。

2023年1月12日時点の情報

て調査を行った。UK Biobank関連の論文では、最も多くの論文に著者として関与している国は英国であるものの、バイオバンクの運営拠点である英国の研究者が、論文著者に含まれていない論文も半数程度存在する。一方で、北京大学とオックスフォード大学の共同研究である中国の China Kadoorie Biobank 関連の論文では、国外からのデータアクセスを受け付ける前に中国の研究機関からの優先的な利用期間を設けているため、中国又は英国の研究者が大半の論文の著者に含まれ、本邦の東北メディカル・メガバンク機構においても国外の研究者のデータへのアクセス許可を行っ

ていないことから、関連する全ての論文において日本の研究者が著者に含まれており、バイオバンクの試料・情報利用に関する方針の違いが反映されている⁵⁾。各バイオバンクがホームページで報告している論文数と Web of Science で検索した際の論文数は、UK Biobank 以外のバイオバンクでは大きく異なっていたが、最も大きい要因は、発行年が比較的早く、少数の症例の情報を用いて実施された研究において、バイオバンクの名称が文中に用いられていない論文が多く、Web of Science において抽出できなかったことが挙げられる。その他にも、Web of Science の検索条件での取得漏れや各バイオバンクでの掲載基準の違い、情報反映の即時性の違い等の要因で全ての論文を網羅することが不可能であったが、各バイオバンクの大規模データを利用した主要な論文の大部分は、今回の抽出条件で集計の対象となっているため、各バイオバンクの論文著者国の傾向を示すための参考値として提示する（表2）。

UK Biobank の成果物

UK Biobank ではオープンアクセスの原則の中で、研究で得られた知見を他の研究でも利用可能とするために研究結果の公表を求めている。本稿では世界中の研究者が試料・情報へのアクセスを行い、多様な研究が実施されている UK Biobank の成果物からバイオバンク研究の現状を把握する

表2 主要なバイオバンク関連の論文著者国（参考）

UK Biobank (4,188報)			China Kadoorie Biobank (245報)			東北メディカル・メガバンク機構 (77報)		
順位	論文著者国	論文数	順位	論文著者国	論文数	順位	論文著者国	論文数
1	英国	2,062 (49%)	1	中国	241 (98%)	1	日本	77 (100%)
2	米国	1,706 (41%)	2	英国	217 (89%)	2	オランダ	1 (1%)
3	中国	905 (22%)	3	米国	48 (20%)	2	スウェーデン	1 (1%)
4	オーストラリア	510 (12%)	4	オーストラリア	18 (7%)	2	米国	1 (1%)
5	スウェーデン	404 (10%)	5	デンマーク	4 (2%)			
			5	日本	4 (2%)			

出所：Web of Science® クラリベイトをもとに作成

2023年1月12日時点の情報

集計対象国の研究者が一人でも論文の著者に含まれる場合に、当該国の論文として集計

バイオバンクの名称がタイトル、アブストラクト、キーワードに含まれる論文を抽出しているため全ての関連論文を網羅している数値ではない。

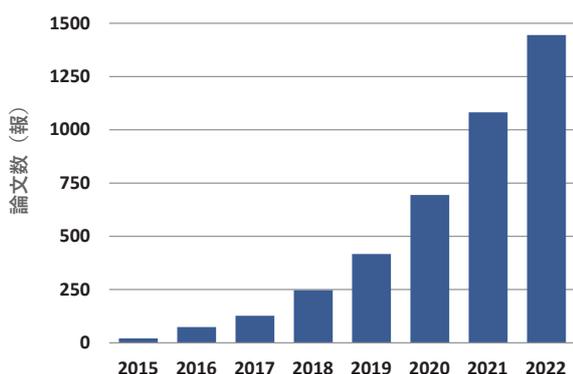
5) 東北メディカル・メガバンク機構、バイオバンク試料・情報関連ウェブサイト、
<http://www.dist.megabank.tohoku.ac.jp/faq>

ために、Web of Scienceに登録されている文献を用いて調査を行った。調査はタイトル、アブストラクト、キーワードに“UK Biobank”又は“United Kingdom Biobank”が含まれる原著論文を対象とした。Web of Scienceに登録されている対象論文数は4,188報となり、調査同日の2023年1月12日時点でUK Biobankのホームページに掲載されている3,207報よりも多い数となっている。前述のようにホームページへの反映には論文公開から時間を要することや、Web of Scienceでの抽出にはレビュー論文等が含まれることが要因であるが、大半がUK Biobankの情報を利用した研究の報告であることを確認し、UK Biobankに関連する論文として調査対象とした。今回の調査で、論文の著者国の集計を行う際には、集計対象国の研究者が

一人でも論文の著者に含まれる場合に、当該国の論文として集計を行った（例えば、英国と米国の著者2名で執筆された論文は英国と米国それぞれの国で1報として扱い集計を行った）。

UK Biobankに関連する論文数の経時推移を確認すると、論文数は年々増加しており、バイオバンクでのデータの蓄積と共に多くの論文が公表されていることが確認できる（図2）⁶⁾。著者の所属機関の国籍を見ても、英国だけではなく、米国や中国を中心として、様々な国からUK Biobank関連の論文が発表されており、自国外の研究者に対しても試料と情報を広く公開しているオープンアクセスの戦略が、生体情報を用いた研究の発展にも大きく貢献していることが分かる（図3）。

図2 UK Biobankに関連する論文数の推移



出所：Web of Science®クラリベイトをもとに作成
2023年1月12日時点の情報

－国際連携

臨床情報を用いた研究では、複数の国の研究機関が連携して実施される研究が増加しており、バイオバンクをはじめとするレジストリデータを用いた研究も例外ではない。UK Biobankにおいては、バイオバンクと試料・情報を利用する研究者との共同研究という形ではなく、アクセスの許可という形で情報を共有しているため、UK Biobankの関係者を著者として含めない形式で論文の投稿が可能となっている。本稿では、バイオバンク研究の実施国の動向を調査するために、UK Biobankに関連する論文の著者の所属機関の国籍から、国境を跨いだ共同研究について集計を行った（図4）。当該

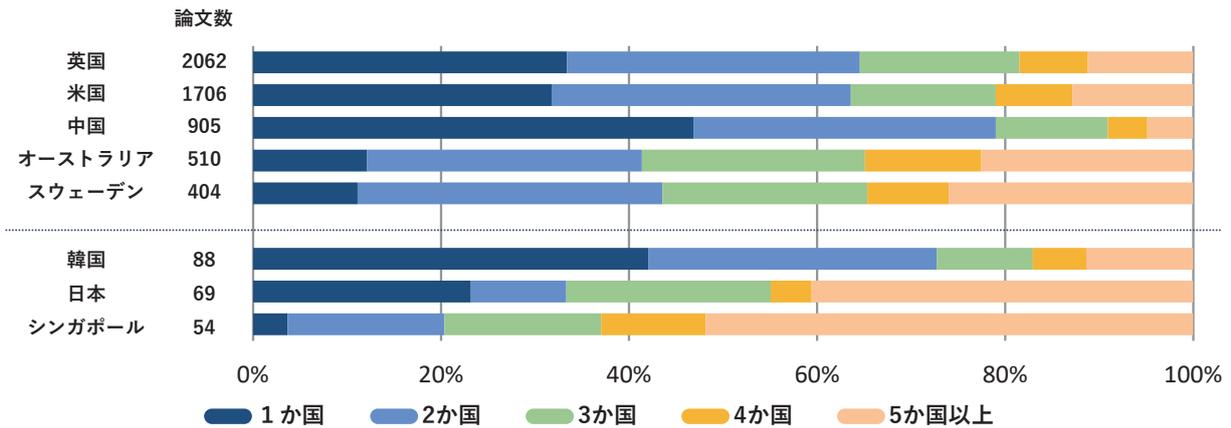
図3 UK Biobankに関連する論文の著者国



出所：Web of Science®クラリベイトをもとに作成
2023年1月12日時点の情報

6) UK Biobank, Past data releases, <https://www.ukbiobank.ac.uk/enable-your-research/about-our-data/past-data-releases>

図4 UK Biobank に関連する論文の共著分析



出所：Web of Science®クラリベイトをもとに作成
論文数上位5か国と、中国を除くアジアにおける論文数上位3か国の情報を掲載
2023年1月12日時点の情報

国の研究機関に所属する研究者が一人でも含まれている論文の数を分母として、各論文の著者の所属機関の国数の分布を確認した。論文数が上位である英国と米国では、単独国、2か国、3か国以上の著者によって執筆された論文がそれぞれ約3割程度の分布となり、類似した国際連携の傾向が確認できた。一方で、論文数が上位の国においても、単独国の著者によって執筆された論文が4割を超える中国や、単独国の著者によって執筆された論文が約1割程度に留まるオーストラリアやスウェーデン等、研究の成果創出に向けた国際連携の動向には各国の特徴が存在している。アジアを見てもその特徴は顕著であり、韓国は中国同様に自

国単独での研究が4割を超える一方で、シンガポールは複数国で実施されている研究に参画した際の論文が過半数を占めていることが分かる。各国の共著国の内訳を確認しても、地理的な特徴やクラスターは確認されなかった（表3）。

－製薬産業との連携

UK Biobank に関連する論文の著者の所属機関は、英国と米国の学術機関が論文数の上位を占めており、製薬産業を含む民間企業が著者に入る論文の割合は少ないが、2022年の製薬企業が著者として関与する論文数は3年前の2019年の論文数から約3倍となる39報に増え、徐々に増加傾向とな

表3 UK Biobank に関連する論文の共著国

英国			米国			中国		
順位	共著国	割合	順位	共著国	割合	順位	共著国	割合
1	米国	31%	1	英国	37%	1	米国	32%
2	オーストラリア	16%	2	中国	17%	2	英国	20%
3	オランダ	12%	3	オーストラリア オランダ	13%	3	スウェーデン	9%
韓国			日本			シンガポール		
順位	共著国	割合	順位	共著国	割合	順位	共著国	割合
1	米国	41%	1	英国	55%	1	米国	78%
2	英国	23%	2	米国	54%	2	英国	61%
3	オーストラリア ドイツ	11%	3	オランダ	33%	3	ドイツ	48%

出所：Web of Science®クラリベイトをもとに作成
割合算出の分母は各国の著者が一人以上含まれる全ての論文数とした。
2023年1月12日時点の情報

っている。製薬企業の中で最も多くの論文の著者として関与しているのはRegeneron Pharmaceuticals社（米国）であり、同社は、UK Biobankと連携し、UK Biobankに蓄積された試料のエクソームシーケンスを自社の研究所で実施している。Regeneron社が関与する論文数は、本邦で最も多くの研究成果を報告している理化学研究所の論文数を上回る論文数となっている。大手製薬企業25社⁷⁾にRegeneron社を加えた26社が著者として関与している論文は、全体の4,188報中126報で全体の約3%であった。

－研究内容

UK Biobankに関連する論文の研究内容の傾向を調査するために論文のタイトルの解析を行った。単語単位のbi-gram（隣接する二つの単語からなる文字列）を用いて論文タイトルに使われている表現の頻度を集計した。上位から“Mendelian Randomization”が555報で約13%の論文のタイトルに含まれており、続いて“Randomization Study”、“Cohort Study”と研究や解析の手法が上位に並んだ(表4)。観察研究であるバイオバンクの性質上、介入によるランダム化が出来ない中で、“Randomization Study”という表現が上位に挙がってきた理由は、“Mendelian Randomiza-

tion”と共に、“Mendelian Randomization Study”の研究手法が論文のタイトルに多用されていることが理由である。メンデルランダム化(Mendelian Randomization)は、ゲノム情報を操作変数として扱い、観察研究における交絡を軽減する研究手法であり、近年増加している研究手法である（詳細は後述）。その他の研究手法として“Polygenic Risk Score (PRS)”や“Genetic Risk Score (GRS)”も上位に現れており、ゲノム解析に関する論文が多く公表されていることが分かる。疾患領域に関してもCOVID-19、心血管疾患、2型糖尿病、がんを表す表現が上位に並び、バイオバンクの情報を用いて多く研究されている疾患領域も読み取ることが出来る。

UK Biobank の試料

UK Biobankでは健康状態や、身体測定、血液、尿、唾液等のサンプルと検査結果の提供に加え、X線吸収測定法(DXA法)による骨密度測定結果やMRI等の画像情報、加速度センサーの情報が提供されており、多様なドメインの情報が利用可能となっている(表5)⁸⁾。UK Biobankのホームページで公開されている2021年に報告された論文931報の中からWeb of Scienceに登録されている859報を対象に、研究対象とした疾患領域と、UK

表4 UK Biobankに関連する論文のタイトルに用いられた表現

順位	bi-gram	論文数	順位	bi-gram	論文数
1	Mendelian Randomization	555 (13%)	11	Type 2	129 (3%)
2	Randomization Study	358 (9%)	12	2 Diabetes	129 (3%)
3	Cohort Study	309 (7%)	13	Physical Activity	124 (3%)
4	Genome Wide	294 (7%)	14	Cross Sectional	117 (3%)
5	Wide Association	239 (6%)	15	Risk Factors	107 (3%)
6	Prospective Cohort	181 (4%)	16	Population Based	100 (2%)
7	Association Study	152 (4%)	17	Genetic Risk	97 (2%)
8	Covid 19	142 (3%)	18	Mendelian Randomisation	95 (2%)
9	Polygenic Risk	140 (3%)	19	Cancer Risk	85 (2%)
10	Cardiovascular Disease	135 (3%)	20	Risk Scores	75 (2%)

出所：Web of Science®クラリベイトをもとに作成

ゲノム関連の研究・解析手法を緑色、疾患関連を橙色で示した。ストップワード又は“Biobank”という単語を含む表現は集計対象外とした。

2023年1月12日時点の情報

7) DATA BOOK 2022 (日本製薬工業協会) の世界の大手製薬企業25社に準ずる。

8) Littlejohns, Thomas J., et al. UK Biobank: opportunities for cardiovascular research. *European heart journal*, 2019, 40.14: 1158-1166.

表5 UK Biobank で集積されている主な試料・情報（一部）

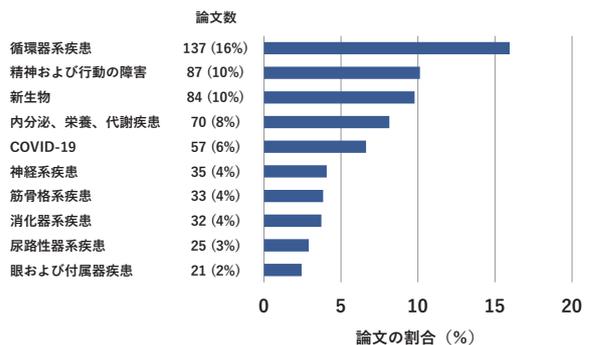
データ	内容	データ	内容
アンケート (一部)	- 基礎 民族、教育、雇用、世帯情報等 - 生活 喫煙、アルコール、身体活動、食事、睡眠等 - 出生時情報 出身地、出生時体重、母乳育児、母親の喫煙等 - 家族 家族の病歴、父母の年齢、父母の死亡時年齢等 - 病歴 病状、投薬、手術、がん検診、疼痛等 - 認知機能 記憶力、反応時間、数値記憶等 - 聴力検査	身体測定	血圧、脈拍、動脈硬化 身長、ウエスト、BMI 握力、肺活量、骨密度、運動時心拍数 視力、眼圧、網膜画像
検体回収	血液、尿、唾液	ゲノム	SNP タイピング エクソーム配列 全ゲノム配列
		生化学検査	アルブミン、クレアチニン、HbA1c 等 血漿メタボローム
		食事	過去24時間の飲食品の消費（複数回測定）
		画像	頭部、心臓、腹部のMRI 頸動脈超音波 全身DXA スキャン
		加速度 センサー	手首加速度センサーを7日間装着

出所：Littlejohns, Thomas J., et al. UK Biobank: opportunities for cardiovascular research. *European heart journal*, 2019, 40.14: 1158-1166.

Biobank の特徴的な試料・情報であるゲノム情報、画像情報、加速度センサーの使用状況について論文のアブストラクトとメソッドの記載内容から調査を行った。身体測定と生化学検査の結果も利用頻度の多い情報であったが、利用される検査項目が研究ごとに多岐にわたるため、今回は調査の対象とはしなかった。

はじめに、研究対象の疾患領域の分布を確認するために、各論文に対してICD-10に準拠した疾病分類表の大分類を一つ割り当てた。複数疾患を対象とした研究や、解析手法の開発等に関する研究で特定の疾患に紐づかない論文に関しては、割り当てを行わなかった。2021年に報告された論文で、最も多く研究された疾患領域は循環器系の疾患であった(図5)。これは前項の論文タイトルの分析においても上位に“Cardiovascular Disease”が含まれていたこととも一致しており、心血管疾患の研究はバイオバンク研究における最も大きな研究対象の一つとなっている。その他では、うつ病や認知症が含まれる精神および行動の障害、各種がんに関連する疾患が含まれる新生物、糖尿病が含まれる内分泌・栄養・代謝疾患、COVID-19に関連した研究が多く報告されており、これらも論文タイトルの分析の傾向と概ね一致する。単一の

図5 UK Biobank に関連する論文の対象疾患 (上位10分類)



出所：医薬産業政策研究所にて作成
図の体裁維持のため、表記疾病分類の大分類の名称を一部省略している。
COVID-19はエマージェンシーコードが利用されたため別区分での集計とした。

分類が割り当てることができない研究も多く存在しており、その中にはゲノム情報や頭部MRIの解析手法を提案する論文等が多く存在した。

UK Biobank において、現時点で公開されているゲノム関連の情報は、大きく分類してマイクロアレイを用いたSNPタイピング、全エクソーム配列、全ゲノム配列の三種類に分類される⁹⁾。公開されているサンプル数はそれぞれ、SNPタイピングが50万、全エクソーム配列が47万、全ゲノム配

9) UK Biobank, Genetic data, <https://www.ukbiobank.ac.uk/enable-your-research/about-our-data/genetic-data>

列が20万となっている。2021年に報告された論文859報の中でゲノム情報が利用された研究は64% (547報) であり、論文タイトルの分析結果の傾向と同様に、ゲノム関連の研究が過半数を占めていた。全エクソーム配列と全ゲノム配列のデータは2021年の途中で公開されたデータであり、利用された研究はゲノム情報を利用した論文の9% (47報) であったが、今後もサンプル数が増加することがUK Biobankからも公表されており、今後これらの情報を利用した論文数の増加が予想される。論文タイトルの分析で最上位となっていたメンデルランダム化を用いて解析している論文は、ゲノム情報を利用した論文の34% (187報) であり、GWASで検出されたバリエーション等の情報を応用した研究も数多く実施されていることが、この調査でも確認することができた。

画像情報に関しては、頭部と心臓、腹部のMRI画像、DXA法による全身スキャン、頸動脈超音波の画像、及び眼底画像が提供されている¹⁰⁾。2021年に報告された論文の中で画像情報を利用した論文は16% (140報) であり、最も高頻度で利用された情報は、頭部のMRI画像に由来する情報の9% (78報) であった (図6)。頭部のMRI画像が使用された疾患領域は、精神および行動の障害が最多で、28% (22報) であった。画像情報の中には、実際の画像だけではなく、画像から導出される数値情報が算出されており、その数値も併せて公開されている画像も存在する。今回は画像自体を利用した論文に加えて、導出された数値情報を利用

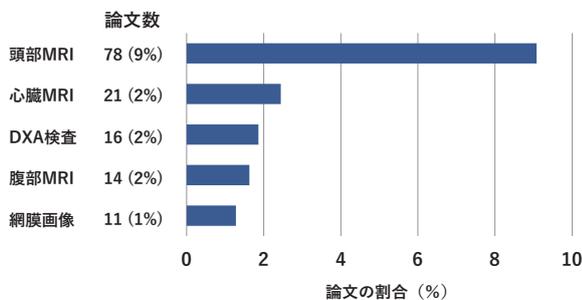
した論文も集計に含めており、頭部MRIの画像に関しては、脳容積や灰白質の容積等の様々な数値情報が併せて公開されているため、これらの情報を利用した論文も多く存在する。このように過去の研究で得られた情報を二次利用できる環境を構築することも、情報の加工が容易となるため、利用頻度を高める一因となっていると考えられる。

加速度センサーの情報が利用された論文は2% (20報) であった。加速度センサーは主に、日中の活動量の推定、又は睡眠時間の推定のいずれかを行う研究で用いられていた。UK Biobankでは類似した情報として、MET (Metabolic Equivalent Task) scoreの算出のために、1週間の活動内容 (激しい運動、ウォーキング、テレビの視聴、パソコンの使用等) それぞれに費やした時間や、睡眠時間等の情報をアンケートにて収集しており、そちらの情報を利用している研究が加速度センサーの情報を用いる研究と比較して多数存在した。加速度センサーの情報は情報の加工方法自体が研究の途上であり、研究で共変量の一部として使う場合等では現時点では採用されづらい傾向にあると考えられる。

大規模データを用いたゲノム研究の実施国

多様な試料と情報を蓄積しているUK Biobankの研究成果から見ても、ゲノム情報を用いた研究がバイオバンクの研究において最も大きな研究分野であることが分かった。今回の調査において、ゲノム情報と疾患等の臨床情報の関連を調査するような、バイオバンクの試料・情報を用いたGWASにおいては、遺伝子による疾患リスクの推定にPRS又はGRSが中心的に使用され、観察研究における因果関係を推定するために、GWASで検出されたバリエーションをメンデルランダム化に用いた研究が多く実施されていることが確認された。それぞれの解析手法の簡単な説明を、東北メディカル・メガバンク機構のサイトを参照して以下に記す。

図6 UK Biobankに関連する論文の画像利用頻度



出所：医薬産業政策研究所にて作成

10) UK Biobank, Imaging data, <https://www.ukbiobank.ac.uk/enable-your-research/about-our-data/imaging-data>

- PRS (Polygenic Risk Score)、GRS (Genetic risk score)¹¹⁾

GWAS等で、疾患との関連が示唆された一塩基多型等のバリエーションについて、高リスク多型がもたらす推定効果量と、各個人が持つ高リスク多型の数の積をすべて足し合わせて得られる数値。個人ごとに算出され、このスコアに基づいて、様々な疾患における遺伝的な発症リスクの高低を定量的に評価できる。疾患の有無だけでなく、血圧や血糖値などの連続的な測定値についても用いられる。PRSとGRSは同様の解析手法であるが、PRSでは疾患とより関連が弱いバリエーションもスコアの算出に含めるものを指す場合もある。

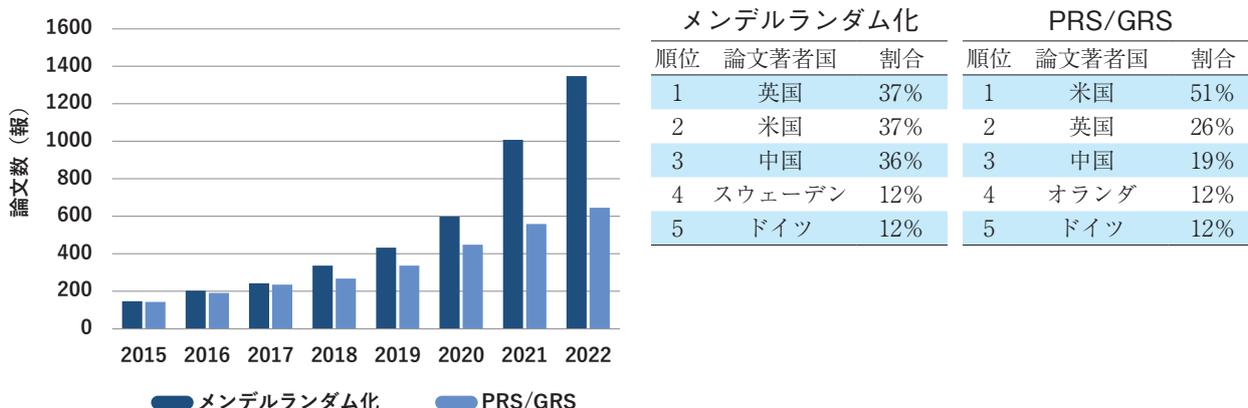
- メンデルランダム化¹²⁾

一塩基多型等のバリエーションがランダムに分配されるという法則を利用し、バリエーションを操作変数として利用して、観察研究において生じる交絡を軽減する手法。東北メディカル・メガバンク機構の研究では、BMIの値と大腸がんのリスクの関連性を確認する際に、生活習慣等の交絡因子の影響を受ける実際のBMIの値ではなく、ゲノム情報から予測されたBMIの値を用いて、BMIの高い集団と低い集団での大腸がんのリスクを比較した。予測されたBMIの値を解析に利用することにより

集団間の背景因子が均等になり、従来の観察研究と比較して交絡の影響を軽減することが出来る。

前項までの集計ではUK Biobankに関連する研究のみに焦点を当てていた。UK Biobankの試料・情報を利用した研究の実施国は、バイオバンク研究参加者の人種の分布を踏まえ、欧州からの利用頻度が増加している可能性がある。そこで、バイオバンクのゲノム情報を用いた研究において、現在最も多用されているメンデルランダム化とPRS/GRSに注目して、大規模データを用いたゲノム研究を実施している国について調査を行った。Web of Scienceを利用してこれらの研究に関連する(タイトル、アブストラクト、キーワードに単語が含まれる)原著論文を確認すると、メンデルランダム化(検索語:“Mendelian Randomization” OR “Mendelian Randomisation”)とPRS/GRS(検索語:“Polygenic Risk Score” OR “Genetic Risk Score”)に関連する論文はそれぞれ論文数を伸ばしており、近年注目されている研究手法であることが確認できる(図7)。それぞれの研究手法を利用した論文の著者の所属機関の国籍の傾向は概ねUK Biobankに関連する研究と同様であったが、PRS/GRSを利用した研究では、米国の論文数が他国を圧倒しており、研究内容においては

図7 ゲノム研究(メンデルランダム化、PRS/GRS)の論文数推移



出所: Web of Science® クラリベイトをもとに作成
2023年1月12日時点の情報

11) 東北メディカル・メガバンク機構、プレスリリース、<https://www.megabank.tohoku.ac.jp/news/46976>

12) 東北メディカル・メガバンク機構、プレスリリース、<https://www.megabank.tohoku.ac.jp/news/44297>

国ごとの特徴がみられた。この傾向はUK Biobankに関連する研究の中でも同様であり、PRS/GRSを利用した研究は米国で多数実施されていた。

最後に、いずれかの研究手法を用いた研究の論文から国際連携の現状を確認する。複数国の機関の著者が含まれる論文において、同一論文内で著者となっている国の組み合わせを確認し、共著論文数の上位20番目までの組み合わせを用いて、連携の状態を図示した(図8)。大規模データを用いたゲノム研究は、米国、英国、中国が主導しているものの、国際連携の中心となっているのは米国と英国であり、中国は他国との連携が比較的少なく、国内で研究を遂行している傾向にある。続いてオランダ、ドイツ、スウェーデンが続き、それぞれの国の間でも多くの共同研究が実施されていることが確認できた。科学技術・学術政策研究所が公表している国・地域別論文発表数¹³⁾においても、基礎生命科学と臨床医学の分野では、米国、中国、英国が上位3か国であり、今回の調査と類似した結果となっているが、今回調査を行った領域に限定すると、4位以下の国の論文数との隔たりが大きいことがわかる。また、スウェーデン、デンマーク、フィンランドといった北欧諸国がラ

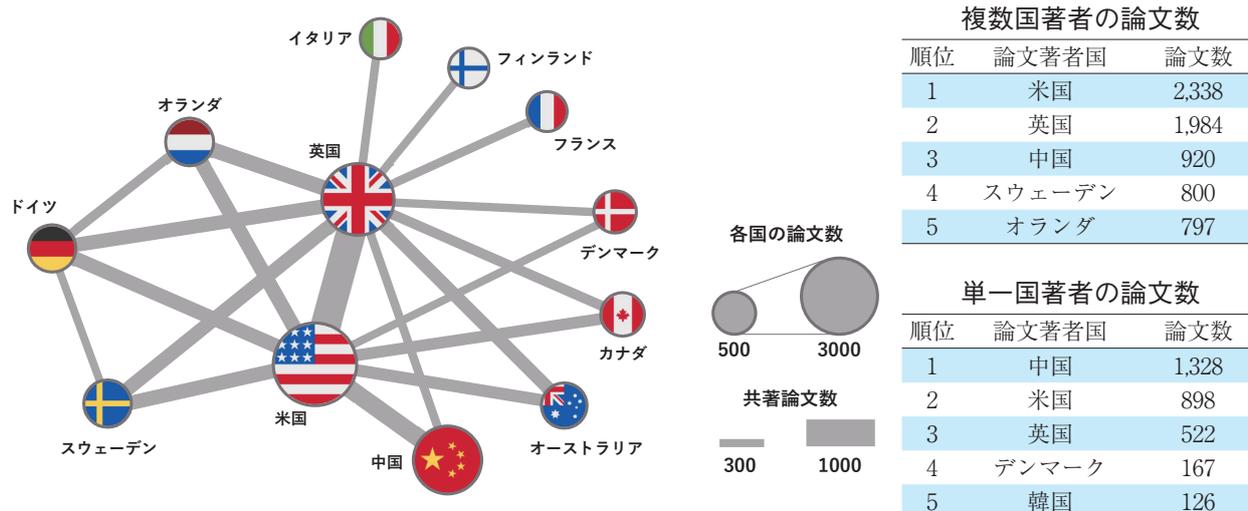
ンキングの上位にくることも、基礎生命科学と臨床医学全体のランキングとは少し異なる特徴といえる。

まとめ

1990年代後半から開始され、各国で進められてきたバイオバンク研究は、近年データの集積が進んできたことによって、生体情報の大規模データとして、蓄積された情報を利用した多くの成果が報告されてきている。バイオバンクの利用方針によって、製薬企業からの利用や海外からの情報アクセスへの対応は現時点で異なっており、個人情報保護や大規模データの共有方法等のデータ管理に関する諸問題に関しては、情報の更なる有効活用のために、バイオバンクと製薬産業が協力して検討していくことが望まれる。

欧米の複数の製薬企業はバイオバンクとの共同研究によって、大量のゲノム情報を創薬に利用できる基盤を作っており、本邦の製薬企業も欧米に遅れをとらないためにバイオバンクとの連携を開始している。一方で、本邦においては現在利用可能な最も大規模なバイオバンクである UK Biobankを利用した研究や、大規模なゲノム情報を利

図8 大規模データを用いたゲノム研究の連携



出所：Web of Science® クラリベイトをもとに作成
2023年1月12日時点の情報

13) 科学技術・学術政策研究所、科学研究のベンチマーキング2021、
<https://www.nistep.go.jp/research/science-and-technology-indicators-and-scientometrics/benchmark>

用した研究に関する論文数は他の国から遅れをとっている。本邦では自国内のバイオバンクの整備が進んできているため、人種差等の影響を踏まえ、国内の情報を多く利用している可能性も存在するが、他国でも類似した状況の国は存在している。バイオバンクの情報が利用可能となってきたからまだ日が浅いことも踏まえると、科学技術・学術政策研究所が公表した、論文生産において日本のポジションが近年低下してきているという報告とも傾向が一致しており、大規模データを用いたゲノム研究の分野も例外ではなく、研究力の低下が生じている懸念がある¹³⁾。本邦では、大学や病院が主導し、自組織で収集した試料や情報を基にしたバイオバンク研究は増加しており¹⁴⁾、データ集積の観点では他国の状況に追随しているにもかか

わらず、論文数が少なくなっているのは、それらの情報を利用して成果に繋げる研究において、他国と比較して遅れをとっている可能性が高い。研究成果を多く公表している、米国、英国、中国の国内における、UK Biobank 関連の論文数の上位3組織は全て大学となっていたが、本邦では理化学研究所が最も多い論文数となっており、次点以降には大学が並ぶものの、上位国と若干異なる傾向がみられている。他国の傾向をみると、論文数の増加のためには、大学を巻き込んだ研究力の向上が必須であり、本邦の創薬分野での国際競争力を維持させるためにも、今後、製薬企業もアカデミアとの連携をより深め、生命科学分野の大規模データを扱う人材の育成に力を入れていく必要があるだろう。

14) 日本医療研究開発機構、バイオバンク利活用ハンドブック第3版、<https://www.amed.go.jp/content/000085128.pdf>