# 日本医療研究開発機構　　創薬基盤推進研究事業
# 事後評価報告書

## I 基本情報

研究開発課題名：ナショナルセンター・バイオバンクネットワークを基盤とする疾患別情報統合データベースを活用した産学官連携による創薬開発研究
Industry-academic-government collaboration study for drug development utilizing Integrated Disease-specific Information Database based on National Center Biobank Network

研究開発実施期間：令和 2 年 12 月 17 日～令和 7 年 3 月 31 日

研究開発代表者　氏名：後藤　雄一
Yuichi Goto

研究開発代表者　所属機関・部署・役職：
国立研究開発法人国立国際医療研究センター・NCBN 中央バイオバンク・バイオバンクアドバイザー
National Center for Global Health and Medicine, National Center Biobank Network (NCBN), Central Biobank, Biobank Advisor

## II 研究開発の概要

ナショナルセンター・バイオバンクネットワーク（NCBN）が有する、専門性の高い臨床情報と科学的に高品位といえる試料を用いて、参画企業 7 社の要望から精神・神経疾患、免疫炎症疾患を対象疾患として選定し、産官学が連携して創薬開発研究を行った。精神・神経疾患においては、脳の病態をよりよく反映しているとされる脳脊髄液（CSF）を用いたオミクス解析を中核にした解析、免疫炎症疾患においては病状や薬剤反応に対応した時系列を加味した横断的な試料・情報を用いたオミクス解析を実施した。

本研究事業には、測定データや解析データを疾患別情報統合データベース（Integrated Disease-specific Information Database: IDID）に集約する非競争的フェーズ、および、参画企業等が独自の薬剤開発へと進展させる競争的フェーズを含んでいる。本事業で得られた非競争的データを、事業終了後 3 年間の独占アクセス期間を経て、最終的には広く公開することで我が国の創薬開発に大いに資することができる。

7 つの製薬企業と 7 つのアカデミア研究機関が協働した本事業では、「代表者協議会」、「運営事務局」等を組織し、企業間、企業-アカデミア間の連絡・調整作業を行い、円滑な運営を行った。「データ利用審査委員会」を 3 年度目から設置してデータ利用を審査・推進した。「知的財産委員会」を適宜開催して学会・論文発表の公表に向けた事前審査等を実施した。「契約委員会」を適宜開始して共同研究契約に基づく事業活動を実施した。具体的な成果は、以下の 6 項目が挙げられる。

1）疾患別情報統合データベースの構築と運用

多様な試料を用いたオミクスのデータを国立国際医療研究センター（NCGM）内のオブジェクトストレージサーバーへ検疫プロセスを経て格納した。GAPFREE4 では外部機関として企業個社の研究者が、NCGM 所内に設置

するNCBNのデータサーバーや計算サーバーへとアクセスすることとなるため、これに特化した「セキュリティガイドライン」を策定した。一部は同センター内における高速計算サーバーによる一次的解析を可能とした。これにより作業自由度が大幅に向上した。データ利用申請・登録申請に対するプロセスを迅速に進めるために、アカデミア、企業の研究グループより予め研究者リストを提出して作成しておくことで、その照合プロセスの多くを自動化した。データ利用審査委員会の承認を経て大多数の研究者が運用希望したのは、データを持ち出して利用する方法であり、PBMCのマスサイトメトリーデータ、プロテオーム測定データ、自己抗体測定データ、RNAseq解析データ、および空間的遺伝子発現解析データ等が対象となった。また、全ゲノム解析データへのアクセスは、NCGM内での解析のみに限定された使用法としている。本事業終了後の3年間の独占アクセス期間は、参画企業とNCGMがその運営費用を賄い、さらなるデータ利用とその運営体制維持を図る。3年後のデータ公開に向けた活動もこの体制で行う。

2) 精神疾患

統合失調症148、健常対照220、品質管理検体105のCSF試料について、SOMAscanによるプロテオーム解析、Human Metabolome Technologiesによるメタボローム解析、国立医薬品食品衛生研究所によるリピドーム解析を実施した。また同時に採取した血漿についても、上記の解析に加え網羅的miRNA解析を行い、さらに全ゲノム解析（AMEDコントロール群ゲノムで未解析であった症例）を実施した。特にCSFプロテオーム解析では、測定可能であった7596分子の中7283分子がCV<15%（中央値4.9%）と高精度のデータを得ることができた。リピドーム解析でも、GAPFREE1時に比べてCSF測定性能が顕著に向上し395分子（GAPFREE1では133分子)、血漿では812分子が測定でき、また得られたデータの85%は再現性が確認できた。

統合失調症148例および、健常対照106例（年齢マッチング）を用いて解析を行ったところ、群間の平均値に有意差が認められる分子（2268分子がFDR q<0.05）が確認された。また、統合失調症において、異常値を示す集団(10%以上)が認められる分子も多数見出された（異常高値を示す症例が10%以上ある分子が910、異常低値を示す症例が10%以上ある分子が2258など)。

臨床情報（発症年齢、罹患年数、PANSS、BACS）を基にWGCNA（Weighted Gene Coexpression Network Analysis）解析を行い、類似した発現変動を示すタンパク群をクラスターとして特定した。各クラスターの生理機能やパスウェイ情報と関連付け、その変動パターンと臨床情報との相関を解析することで、特定の臨床症状と関連するタンパク群を抽出した。

3) 神経変性疾患

精神疾患と連携して解析手法を統一し、多系統萎縮症(MSA)54例、パーキンソン病(PD) 67例、脊髄小脳変性症(SCD) 90例、筋萎縮性側索硬化症(ALS) 50例、大脳皮質基底核変性症(CBS) 36例、進行性核上性麻痺(PSP) 36例、コントロール 220例において、髄液・血漿プロテオーム測定、血漿miRNA-Seq解析、髄液・血漿親水性メタボローム測定、髄液・血漿疎水性メタボローム解析(国立衛研)を完了した。これらの多層オミクス解析データを臨床情報とともにIDIDに格納した。当初の目標を全て達成した。

髄液プロテオーム解析により、分子AがMSA特異的に高値を呈することを見出した。ROC解析では臨床的に類似しており早期の鑑別が困難なMSA小脳型(MSA-C)とSCD、MSAパーキンソン型（MSA-P）とPDにおいて、各々AUC 0.970・感度90.3%・特異度100%およびAUC 0.860・感度80.0%・特異度91.5%と高い鑑別能を示した。特に発症3年以内の症例に限定すると、MSA-C vs SCD: AUC = 0.993、MSA-P vs PD : AUC = 0.921と極めて高い鑑別能を示し、分子Aは、MSAの早期鑑別診断マーカーとして極めて有用であることも示された。

ALS 79件、コントロール 6件のロングリード解析を完了してIDIDに格納した。ゲノム構造異常、リピート異常伸長を優先的に解析し、候補遺伝子を2個発見した。そのうちのひとつ、遺伝子BにALSに特異的な3kbのレトロトランスポゾンの挿入を同定した。この挿入配列は公的データベースに登録されていない配列であった。企業の承諾を得て外部研究機関との共同研究（個社研究）の準備を進めている。

PSP と PD、CBS と PD、PSP と CBS と鑑別が困難な疾患同士の多層オミクス解析による鑑別を試みた。髄液プロテオーム解析においては、いずれの組み合わせも有意な差異を認めなかった。一方、髄液疎水性メタボローム解析では、PSP と CBS で有意差を認めるのは 36 分子であり、分子 C 及び分子 D が PSP のみで、あるいは PSP の方がより大きく減少していた。PSP と CBS を鑑別する有力なバイオマーカー候補であると考えられる。

4) 免疫炎症疾患

薬剤応答性や疾患特異性、疾患活動性変化を捉えられる対象患者の選択条件を合わせるために企業との協議を重ね、合意した 5 つの自己免疫疾患（全身性エリテマトーデス：SLE、関節リウマチ：RA、強皮症：SSc、シェーグレン症候群：SS、多発性筋炎／皮膚筋炎：PM/DM）患者から前向きに血液等を採取し、各種オミクス測定（CyTOF、全ゲノム解析、DNA メチル化解析、遺伝子発現量解析（バルク PBMC、血球分画)、プロテオーム解析、自己抗体解析）を実施した。

さらに、IDID に格納された測定データを用いて、疾患群間の比較解析を実施した。各患者及び健常人（RA:51, SLE:45, SS:59, SSc:31, PM/DM:18, コントロール:16 例の計 220 例）の Day1 の PBMC の CyTOF 解析データを各疾患群間で比較解析した結果、末梢血中の各種免疫細胞の存在比に疾患間で差異が認められた。また、PBMC から 6 分画した細胞ごとの RNA-seq データ（RA:19, SLE:15, SS:21, SSc:6, PM/DM:9,コントロール:10 例の計 80 例, Day1 検体）を疾患群間で比較解析した結果、いずれの疾患でも発現遺伝子に差異が認められたが、疾患によって変動遺伝子数に幅があった。さらに患者血漿（RA:28, SLE:24, SS:25, SSc:22, PM/DM:15, コントロール:17 例の計 131 例）を用いた自己抗体測定データについて比較解析した結果、健常人と疾患間のほか、各疾患間でも同様に自己抗体抗原タンパクに差異が認められた。

データの精度向上および層別化マーカー候補の同定加速化のため、質の高いバイオバンク試料を活用し、プロテオミクス及び自己抗体の解析を追加した。NCGM バイオバンクが保有する自己免疫疾患患者血漿 468 検体のプロテオミクス及び 198 検体の自己抗体プロファイリングを前向き採取した患者検体とともに実施した。さらに、自己免疫疾患で重要視される病態局所での変化に対する理解を深めるため、生検組織を用いた空間的遺伝子発現解析（Visium）を実施した。

5) オミクスデータ解析

IDID への入力情報の整理のため、疾患ワーキンググループ（WG)ごとにオミクス測定結果、臨床情報項目、非競争的フェーズの解析結果等の情報を IDID へ格納する際に、ディレクトリーを作成した。利用しやすいように常に最新のディレクトリー表を研究者間で共有した。また令和 4 年度から PRISM 研究との連携を図り、医薬基盤・健康・栄養研究所の研究者が分担研究者として参加し、AI を用いた層別バイオマーカー解析を精神・神経疾患で実施し、結果を IDID に格納した。

6) 企業との研究協力、共同研究の状況

運営事務局、疾患ＷＧ、契約委員会・知的財産委員会等を組織し、アカデミアと企業からの参加者が構成員となり活動した。疾患ＷＧにおいては、解析する対象者の選定と解析手法の検討を行い、企業からのニーズに最大限応えた。一方で、企業間でも意見調整を行うことで、スムーズな事業運営を行うことができた。知的財産委員会においては、アカデミア研究機関研究者が学会発表、論文発表を行う際に、事前の承認ルールを作成して対応した。特許出願についてもルールを定めた。

The National Center Biobank Network (NCBN) utilized its specialized clinical information and high-quality samples to select mental and neurological disorders, and immune inflammatory diseases as target diseases based on the requests from seven participating companies. A collaborative research project for drug discovery was conducted through the collaboration of industry, academia, and government. For mental and neurological disorders, the research will focus on omics analysis using cerebrospinal fluid (CSF), which is considered to better reflect brain pathology. For immune inflammatory diseases, the research will conduct omics analysis using time-series and cross-sectional samples and information corresponding to disease status and drug responses.

1) Construction and operation of a disease-specific information integration database

Since researchers from individual companies will access the NCBN data server and computing server installed at the NCGM as external organizations in GAPFREE4, we established specific "security guidelines" for this purpose. After approval by the Data Utilization Review Committee, the majority of researchers requested to take out and use the data at their own companies. Access to whole genome analysis data is limited to analysis within NCGM. Following the conclusion of this project, a three-year exclusive access period will be established, during which participating companies and NCGM will cover operational costs to facilitate further data utilization and maintain the operational framework.

2) Mental disorders

Proteome analysis using SOMAscan, metabolome analysis using Human Metabolome Technologies, and lipidome analysis conducted by the National Institute of Health Sciences were performed on CSF samples from 148 patients with schizophrenia, 220 healthy controls, and 105 quality control samples. Additionally, plasma samples collected simultaneously were subjected to comprehensive miRNA analysis, and whole-genome analysis (for cases where the genome had not been analyzed in the AMED control group). In particular, CSF proteomics analysis yielded high-precision data, with 7,283 out of 7,596 detectable molecules having a CV <15% (median 4.9%). Lipidomics analysis also showed a significant improvement in CSF measurement performance compared to GAPFREE1, with 395 molecules (133 molecules in GAPFREE1) measured in CSF and 812 molecules in plasma, and 85% of the obtained data showed reproducibility.

3) Neurodegenerative diseases

By unifying analytical methods in collaboration with mental disorders, we analyzed 54 cases of multiple system atrophy (MSA), 67 cases of Parkinson's disease (PD), 90 cases of spinocerebellar degeneration (SCD), 50 cases of amyotrophic lateral sclerosis (ALS), 36 cases of corticobasal degeneration (CBS), progressive supranuclear palsy (PSP) 36 cases, and controls 220 cases. These multi-omics analysis data were stored in IDID along with clinical information. All initial objectives were achieved.

4) Immune-mediated inflammatory diseases

After extensive discussions with companies to align the selection criteria for patients who can be captured for drug response, disease specificity, and changes in disease activity, five immune-mediated inflammatory diseases were agreed upon: systemic lupus erythematosus (SLE), rheumatoid

arthritis (RA), systemic sclerosis (SSc), Sjögren's syndrome (SS), and polymyositis / dermatomyositis (PM/DM). Prospective samples were collected from patients with these diseases, and various omics measurements, including CyTOF, were conducted. Furthermore, comparative analyses between disease groups were performed using the measurement data stored in IDID.

5) Omics data analysis

To organize input information for IDID, directories were created to store omics measurement results, clinical information items, and analysis results from non-competitive phases for each disease working group (WG) when storing information in IDID. Additionally, starting in fiscal year 2022, collaboration with the PRISM study was initiated, with researchers from the National Institute of Biomedical Innovation, Health and Nutrition participating as collaborative researchers to conduct AI-based stratified biomarker analysis for mental and neurological disorders, with results stored in IDID.

6) Status of research collaboration and joint research with companies

An organizational structure was established, including the Operations Secretariat, Disease Working Groups, Contract Committee, and Intellectual Property Committee, with members comprising participants from academia and industry. The Disease Working Groups selected subjects for analysis and discussed analytical methods to maximize responsiveness to corporate needs. The Intellectual Property Committee established pre-approval rules for academic researchers to present at conferences or publish papers, and also established rules for patent applications.