

特集 1 精神疾患の網羅的ゲノム解析～課題と展望

2. 東北メディカル・メガバンク計画における
大規模ゲノム・オミックス解析と疾患発症リスク予測

梅影 創*, 清水 厚志*

抄録：疫学研究を進めていくうえで、バイオバンクとコホート研究の質の向上および参加者（検体）数の拡充と、大規模な情報から目的の疾患に関連する要因を見つけだすための遺伝統計学手法の開発は車の両輪の関係にある。本稿では、疫学研究を推進していくためのバイオバンクとコホート研究の現状に触れ、さらに世界最大規模の出生三世代コホートなどを運用する東北メディカル・メガバンク計画の取り組みを紹介する。次いで、疾患関連解析の最大の難問とされる「失われた遺伝率」の問題について触れ、この問題の克服をめざした手法の一つとして、いわて東北メディカル・メガバンク機構で開発した iwate polygenic model (iPGM) を利用した脳梗塞発症リスク予測について述べる。最後に、コホート研究による精神疾患リスク予測に向けたアプローチについて触れる。

日本生物学的精神医学会誌 32 (2) : 68-74, 2021

Key words : cohort study, biobank, genome-wide association study, polygenic model, precision medicine

はじめに

アメリカの Precision Medicine Initiative に代表されるように、これまでの平均的な患者に向けた医療 (one-size-fits-all 型医療) から個々人の生活習慣、生活環境、遺伝情報などの違いを考慮した個別化医療・個別化予防に関心が注がれるようになった。この目的を実現するためには、バイオバンクやコホート研究の大規模化だけでなく、疾病を分子レベルから理解するために遺伝子発現解析などのオミックス解析情報が必要である。また、コホートの質の観点から家系情報付きコホートの重要性が増している。

東北メディカル・メガバンク (Tohoku Medical Megabank : TMM) 計画は、2011 年の東日本大震災からの復興事業の一つとして活動が始まった比較的若いコホート研究であるが、震災コホートでもあるというユニーク性質を持ち、ゲノム・オミックス解析能力を有する複合バイオバンクである。なかでも「失われた遺伝率」の解明をめざして構築された三世代コホートは、家系情報付きの出生三世代コホートとして世界最大規模を誇っている⁷⁾。

本稿では、疫学研究の概要、ならびに国内外のバイオバンク・コホートの活動状況に触れ、次いで、TMM 計画の概要を紹介する。さらに、現在の遺伝情報を用いた疾患関連解析の最大の問題である「失われた遺伝率」を克服する手法として、いわてメディカル・メガバンク機構生体情報解析部門で開発された iwate polygenic model (iPGM) による脳梗塞発症リスク予測について紹介する。最後に、コホート研究を利用した精神疾患予測へのアプローチについて触れる。

1. 疫学研究およびコホート研究について

疫学研究は、仮説の設定を目的とする記述疫学、仮説の検証を目的とする分析疫学、仮説の証明を目的とする実験疫学に大別され、コホート研究はこの中の分析疫学に分類される (図 1)。分析疫学の研究手法は、時間経過を考慮する縦断研究と考慮しない横断研究 (断面研究) に分けられ、コホート研究は前者の縦断研究に分類される。さらにこの縦断研究は大きく分けて前向きコホート研究、後ろ向きコ

Large-scale genomic and omics analysis and risk prediction of disease incidence in Tohoku Medical Megabank Project

*岩手医科大学いわて東北メディカル・メガバンク機構生体情報解析部門 (〒028-3694 岩手県紫波郡矢巾町医大通 1-1-1) So Umekage, Atsushi Shimizu : Division of Biomedical Information Analysis, Iwate Tohoku Medical Megabank Organization, Iwate Medical University. 1-1-1, Idaidori, Shiwagun Yahabacho, Iwate 028-3694, Japan

【清水 厚志 E-mail : ashimizu@iwate-med.ac.jp】

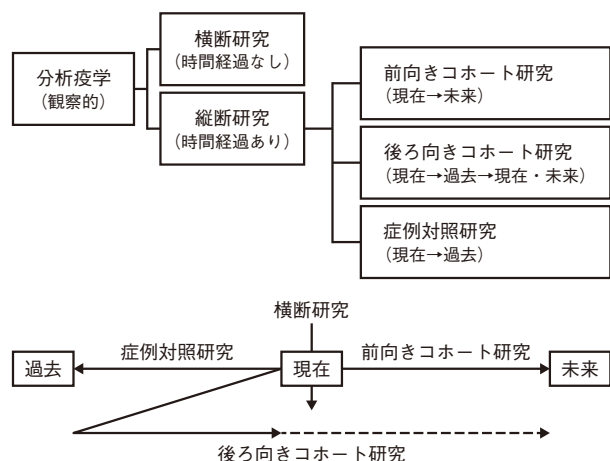


図1 分析疫学におけるコホート研究の位置付け

コホート研究、症例対照研究の3つに分類される。前向きコホート研究は、対象コホートが疾病に罹患する前に特定の要因に曝露している群としていない群と分け、疾病の罹患状況などを追跡調査する研究手法である。後ろ向きコホート研究は、すでに特定の要因に曝露した集団を対象に、過去の曝露状況を一度振り返り、現在または将来の疾患発症状況を検証する手法である。どちらもコホート研究の経時的な観察という点で等しい。一方、症例対照研究は、例えば疾患患者集団と一般健常者集団が別々に集められ、それぞれ過去に遡って曝露要因が検証された後、これらの集団間で比較検証される手法であり、コホート研究とは集団の扱い方が異なる。

2. 国内外の主なバイオバンクとコホート研究

前述の疫学研究を推進するために、古くからコホート研究やバイオバンクの構築が世界中で行われている。The International Hundred Thousand Plus Cohort Consortium (IHCC) の参加コホートリストから主なバイオバンクとコホート研究のリストを表1に示した。国内のバイオバンクとコホート研究については、国立研究開発法人日本医療研究開発機構のウェブサイト上で詳細にまとめられている (https://www.amed.go.jp/site/biobank/itiran_index.html)。前向きコホート研究については、All of Us research program¹⁴⁾ にみられるように大規模化が主流になっているが、わが国においてもTMM計画、日本多施設共同コホート研究 (Japan Multi-Institutional Collaborative Cohort Study: J-MICC STUDY)、多目的コホート研究 (JPHC study)、鶴岡メタボロームコホート研究、愛知県がんセンター病院疫学研究から構成される総計約39.6万人

規模のゲノムコホート連携が構築目目となっている。また、「失われた遺伝率」(後述)の解決を目的とした家系情報付き前向きコホート研究については、アイスランドのdeCODEジェネティクス、オランダのLifelines(三世代コホート)、イギリスのAvon Longitudinal Study of Parents and Children (ALSPAC)(出生コホート)、日本のTMM計画の三世代コホートなど世界に数えるほどである。

3. TMM計画について

TMM計画は、2011年3月11日に発生した東日本大震災により被災された住民の方々へのコホート調査を通じた長期健康支援、および個別化医療・個別化予防実現のための研究拠点の形成の2つの目標をめざした震災復興事業として計画された。TMM計画の第1段階(2011年～2016年)では地震や津波の影響を受けた地域を中心とした8万人規模調査と宮城県全域および岩手県の一部を対象とした7万人規模の三世代(生まれてくる児の世代、児の両親、児の祖父母の世代を加えた三世代)コホート調査の計15万人規模で進められた。これらの調査での健康調査、血液・生化学検査、生理機能検査にゲノム情報が加えられ、健常人集団から構成される前向きコホートとして整備された。さらに、上記コホートのゲノム解析データ、エピゲノムやメタボロームなどのオミックス解析情報(jMorp¹⁷⁾に収載)が合わせて蓄積され、「試料」と「情報」を保有する複合型のバイオバンク(TMMバイオバンク)が整備された。第二段階(2017年～2020年)では、公的なバイオバンクとしての成熟とコホート調査の追跡調査並びに2次調査(2回目の健康調査)が計画され、被災地の健康復興と個別化医療・予防の実現に向けた基盤整備が進められている。本計画の推進のため、東北大学に東北メディカル・メガバンク機構(Tohoku University Tohoku Medical Megabank Organization: ToMMo)が、岩手医科大学にいわて東北メディカル・メガバンク機構(Iwate Medical University Iwate Tohoku Medical Megabank Organization: IMM)が設置され、ToMMoは宮城県住民を、IMMは岩手県住民を主に担当し、それぞれが有機的に連携しながら本計画が進められている。分譲情報は、TMM計画の統合データベースであるdbTMMにカタログとして整備され、国内の研究者に対する提供体制が整えられている。

表1 国外の主なバイオバンクとコホート研究

バイオバンク名, コホート研究名	国, 地域	開始年	終了年	参加者	目標数
23andMe	USA	2007	継続中	6,800,000	
Norwegian Family Based Life Course Study	Norway	1960	2011	5,266,270	
China PEACE (Patient-centered Evaluative Assessment of Cardiac Events) Million Persons Project	China	2014	継続中	2,000,000	4,000,000
Million Women Study	England, Scotland	1996	2001	1,320,000	
Cancer Prevention Study II (CPS-II)	USA	1982	1983	1,185,106	
Finnish Maternity Cohort Serum Bank	Finland	1983	2016	950,000	
Korea Biobank Project	Republic of Korea		継続中	830,000	
Apolipoprotein MORTality RISk study (AMORIS)	Sweden	1985	1996	812,073	
Million Veteran Program	USA	2011	継続中	785,000	1,000,000
EPIC (European Prospective Investigation into Cancer, Chronic Diseases, Nutrition and Lifestyle)	UK, Italy, France, Germany, Norway, Netherlands, Denmark, Spain, Greece, Sweden	1992	2023	521,000	
Newfoundland 100K Genome Project / Sequence Bio	Canada, Province of Newfoundland and Labrador	2018	継続中	520,000	
China Kadoorie Biobank	China	2004	2008	512,891	
UK Biobank	England, Scotland, Wales	2006	2010	502,713	500,000
Children's Hospital of Philadelphia (CHOP) Biorepository	USA, Europe, South America, Canada, Saudi Arabia, Australia	2006	継続中	500,000	1,000,000
SAPRIN (South African Population Research Infrastructure Network)	South Africa	1993	継続中	350,000	700,000
Canadian Partnership for Tomorrow's Health (CanPath)	Canada	2008	継続中	330,000	
Norwegian Mother and Child Cohort Study (MoBa)	Norway	1999	2008	284,000	
Netherlands Twin Registry	Netherlands	1987	継続中	275,000	
45 and Up Study	Australia	2006	2009	267,153	
Geisinger Cohort - MyCode Community Health Initiative	USA	2007	継続中	252,160	250,000
Trndelag Health Study (HUNT)	Norway	1984	2019	248,000	
BioVU Vanderbilt	USA	2007	継続中	244,000	
U.S. Precision Medicine Initiative / All of Us	USA	2018	継続中	191,105	1,000,000
Lifelines	Netherlands	2006	2013	167,729	
ALSPAC (The Avon Longitudinal Study of Parents and Children)	UK	1990	継続中	14,901	

4. ゲノムワイド関連研究

近年のコホート研究は、ゲノム情報をフルに活用したゲノムワイド関連研究 (genome wide association study : GWAS) などの分子疫学的手法を取り入れたゲノムコホート研究が主流となっている。GWAS は、疾患情報などの質的形質 (qualitative trait) だけでなく、身長・体重などの量的形質 (quantitative trait) に影響する遺伝的要因の網羅的な検索に用いられる遺伝統計学の一手法であり、2002 年に日本の理化学研究所のグループが心筋梗塞に関連する一塩基多型 (single nucleotide polymorphism : SNPs) の同定手法として先鞭をつけた¹²⁾。欧州バイオインフォマティクス研究所 (EMBL-European

Bioinformatics Institute : EBI) とアメリカ国立ヒトゲノム研究所 (National human genome research institute : NHGRI) との共同による GWAS 関連論文や解析結果の網羅的キュレーションサイトである the NHGRI-EBI Catalog of human genome-wide association studies, 通称 GWAS カタログ (<https://www.ebi.ac.uk/gwas/>) では、2020 年 10 月の段階で 4,761 報の論文と 213,519 件の GWAS 実施例が納められている。

5. Polygenic model

GWAS が開発された当初は、GWAS は遺伝率を説明するためのツールとして期待されていた。双子

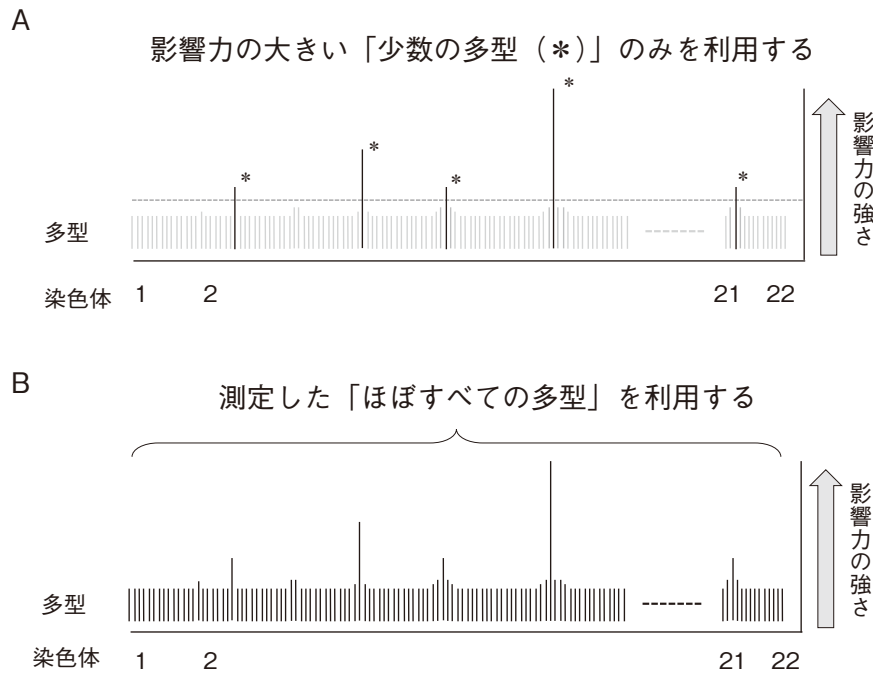


図2 従来のGWASとiPGMの違い

A: 通常のGWASのイメージ, B: iPGMのイメージ, をそれぞれ示した。マイクロアレイで測定された多型情報を1番染色体から順番に(左から右へ)並べ、影響力の強さを縦棒の長さで表現した。

研究により遺伝率が80%であると推定された身長(量的形質)に影響するSNPsの探索にGWASが用いられ^{3, 8, 19)}2010年までに約40のSNPsが同定されているものの、同定されたゲノム多型のみでは遺伝率の5%程度しか説明できないといった「失われた遺伝率」の問題が顕在化することとなった⁹⁾。この問題を説明するモデルの一つとして、GWASで用いられる影響量の大きい少数の多型だけでなく(図2A)、弱い影響を持つ多数の遺伝的要因も含めて遺伝率を説明しようとするpolygenic model¹⁾が注目されるようになった。

6. iPGMによる脳梗塞発症予測

IMM生体情報解析部門は、polygenic modelとして測定した多型のほぼすべて(約357,000個の遺伝子座)を用いる新規のリスク計算手法としてiPGMを考案した(図2B)。本手法に基づき、多因子疾患として知られる脳梗塞をモデルに、バイオバンクジャパン(BioBank Japan:BBJ)の脳梗塞患者(13,214人)と国内の複数のコホートから選抜された対称群(26,470人)の検体から脳梗塞発症リスク予測法が開発された⁵⁾。この予測手法の精度検証として以下のように研究が進められた。まず、症例対象研究での検証として久山町研究の脳梗塞患者と対象検体

(各1,097検体)を用いた検証が行われ、次いで、コホートでの検証として、J-MICC studyとJPHC studyから提供された脳梗塞患者と対称群(各336検体)を用いた検証がそれぞれ行われた。これらの検証から、遺伝的高リスク群の脳梗塞発症オッズ比は、遺伝的低リスク群の1.8~2.0倍に上昇することが確認され、iPGMによる脳梗塞発症リスク予測が可能であることが確かめられた⁵⁾。さらに、この予測法を全国平均とほぼ同じ年齢・職業分布の久山町研究¹¹⁾の検体(既往歴や非同意者を除いた3,038人の10年間にわたる追跡調査)に適用し、前向きコホートにおける発症予測精度の検証が行われた。リスク因子を脳梗塞の発症にかかわるとされる高血圧、糖尿病、喫煙に限定し、リスク因子の個数(0, 1, 2以上)で層別化した脳梗塞発症のハザード比層別化解析が行われた結果、リスク因子が2個以上である場合の遺伝的スコアレベルが高値群の脳梗塞発症のハザード比は、低値群の3.03倍高いことが示された(図3)¹¹⁾。また、遺伝的スコアレベルが低値群であっても、リスク因子が2個以上になるとハザード比が1.62倍に上昇すること、遺伝的スコアレベルが高値群であっても、リスク因子の個数が0個であれば低値群の1.18倍にまで抑えられることも示され、生活習慣に気を配る(リスク因子を減らす)ことによって脳梗塞発症リスクを抑える

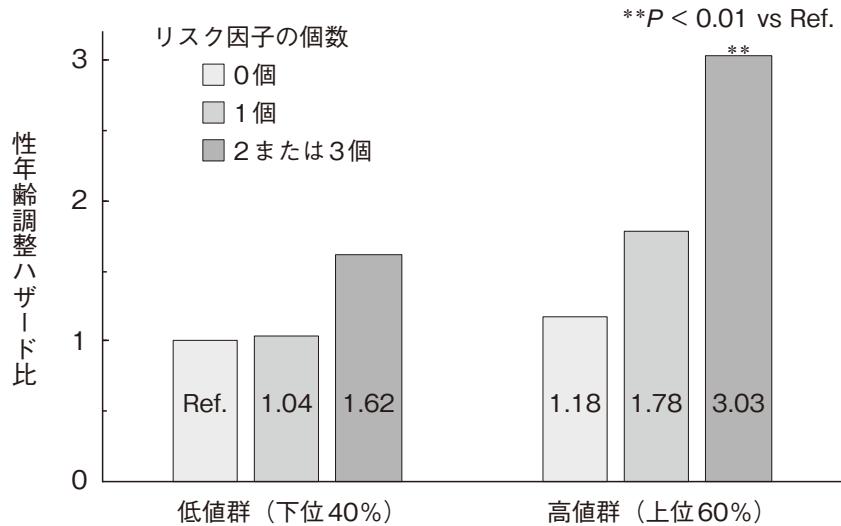


図3 脳梗塞発症のハザード比層別化解析 (文献4より改変)

遺伝的リスクスコアレベルが下位40%の低値群(左), 上位60%の高値群(右)に分け, それぞれのリスク因子(高血圧, 糖尿病, 喫煙)の個数に応じた性年齢調整ハザード比を示した。図中のアスタリスク(**)は, コントロール(Ref.)とのP値が0.01未満であることを示す。

ことが可能であることも示唆された⁴⁾。iPGMは, 本稿で取り上げた脳梗塞だけでなく, 他の多因子疾患にも適用可能であるため, 個別化医療・予防の実現に向けた先導的な手法になることが期待される。

7. TMM計画での「こころの健康」調査と精神疾患関連遺伝子同定アプローチについて

TMM計画では, 被災地コホートであるという観点から, 「こころの健康」に関連した調査として心理的苦痛(K6)⁶⁾, 抑うつ症状(The Center for epidemiologic studies depression scale: CES-D)¹³⁾, アテネ不眠尺度(Athens insomnia scale: AIS)¹⁶⁾の3つの調査も実施されている。TMM計画において, 成人であれば精神神経疾患(心的外傷後ストレス障害[Post traumatic stress disorder: PTSD], うつ病), 小児であれば自閉スペクトラム症が優先的な解析対象として定められている。

精神疾患は多因子疾患とされ, 遺伝率が高いとされる自閉スペクトラム症(遺伝率90%以上)²⁾や統合失調症(遺伝率80%)¹⁵⁾などについてGWAS等が精力的に行われている。なかでもPsychiatric Genomic Consortiumによる統合失調症に対するGWASでは108個の疾患関連サイトが報告され, その中でもドーパミン受容体(dopamine receptor D2: DRD2)とグルタミン酸神経伝達に関与する遺伝子との関連は, 病態生理学的仮説と一致すると

いった成果が得られている¹⁵⁾。一方, 同定されたSNPsのエフェクトサイズが非常に小さいことや患者間の異質性, あるいは診断の不安定性などの理由から, 患者群全体の疾患リスクを説明することは困難と考えられている。近年, 自閉スペクトラム症に関するGWASにおいて, 自閉スペクトラム症と診断された集団を複数の診断ツール¹⁰⁾あるいは機械学習の手法²⁰⁾を用いてサブグループに分け, グループ別にGWASを行うことで新規の遺伝子変異を同定する手法が報告されている。

このように, 精神疾患に代表される遺伝率が高く, 異質性が高い疾患においては, 前述のアンケートの利用, 医師による診断, 機械学習によるクラスタリングなどを考慮した層別化を実施し, GWASやiPGMなどのpolygenic modelに基づいた解析を行うことで, これまで拾いあげることのできなかったSNPsの同定につながるものと期待される。また, 精神疾患に関する研究ではないが, SNPs情報だけでは統計的な検出力に問題があるという観点から, 全ゲノム情報を解析に用いることで「失われた遺伝率」の問題を解決しようとする研究¹⁸⁾も進められており, 精神疾患を含む多因子疾患研究への応用も期待される。さらに, 遺伝率の観点から, 世界最大規模の家系情報付き出生コホートであるTMM計画の三世代コホートは, 精神疾患を含む多因子疾患リスク予測研究において「失われた遺伝率」を埋める重要な役割を担うものと期待される。

おわりに

本稿では、TMM 計画の概要ならびに疫学研究の最大の難問とされる「失われた遺伝率」を説明する仮説として polygenic model に触れ、この概念に基づいて開発された iPGM による脳梗塞発症リスク予測について紹介した。iPGM は、影響力の高い SNPs のみを用いる GWAS とは異なり、測定するすべての SNPs を用いるため、これまでの GWAS の手法では見逃されてきたような影響力の低い SNPs も拾いあげてモデル構築することが可能である。また、iPGM 法はマイクロアレイを利用した解析手法であるため、全ゲノム解析に比べ低コスト（サンプル当たり 1 万円未満）にリスク予想ができることも個別化医療・予防への普及に大きく貢献するものと期待される。また、本稿の終わりに述べたように、多因子疾患とされる精神疾患に影響を及ぼす SNPs の探索についても適切なクラスタリングを行うことができれば、新たな SNPs の発見につながるものと期待される。また、遺伝率の観点から TMM 計画の家系情報付き三世代コホートは、精神疾患を含む多因子疾患関連遺伝子の探索に大きく貢献していくものと期待される。

本論文に開示すべき利益相反は存在しない。

文 献

- 1) Fisher RA (1918) The correlation between relatives on the supposition of mendelian inheritance. *Trans R Soc Edin*, 52 : 399-433.
- 2) Grove J, Ripke S, Als TD, et al (2019) Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet*, 51 (3) : 431-444.
- 3) Gudbjartsson DF, Walters GB, Thorleifsson G, et al (2008) Many sequence variants affecting diversity of adult human height. *Nat Genet*, 40 (5) : 609-615.
- 4) Hachiya T, Hata J, Hirakawa Y, et al (2020) Genome-wide polygenic score and the risk of ischemic stroke in a prospective cohort : The Hisayama study. *Stroke*, 51 (3) : 759-765.
- 5) Hachiya T, Kamatani Y, Takahashi A, et al (2017) Genetic predisposition to ischemic stroke : A polygenic risk score. *Stroke*, 48 (2) : 253-258.
- 6) Kessler RC, Barker PR, Colpe LJ, et al (2003) Screening for serious mental illness in the general population. *Arch Gen Psychiatry*, 60 (2) : 184-189.
- 7) Kuriyama S, Metoki H, Kikuya M, et al (2020) Cohort profile : Tohoku Medical Megabank project Birth and Three-Generation cohort study (TMM BirThree Cohort Study) : rationale, progress and perspective. *Int J Epidemiol*, 49 (1) : 18-19m.
- 8) Lettre G, Jackson AU, Gieger C, et al (2008) Identification of ten loci associated with height highlights new biological pathways in human growth. *Nat Genet*, 40 (5) : 584-591.
- 9) Maher B (2008) Personal genomes : The case of the missing heritability. *Nature*, 456 (7218) : 18-21.
- 10) Narita A, Nagai M, Mizuno S, et al (2020) Clustering by phenotype and genome-wide association study in autism. *Transl Psychiatry*, 10 (1) : 290-301.
- 11) Ninomiya T (2018) Japanese legacy cohort studies : The Hisayama Study. *J Epidemiol*, 28 (11) : 444-451.
- 12) Ozaki K, Ohnishi Y, Iida A, et al (2002) Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. *Nat Genet*, 32 (4) : 650-654.
- 13) Radloff LS (1977) The CES-D scale : A self-report depression scale for research in the general population. *Appl Psychol Meas*, 1 : 385-401.
- 14) Sankar P and Parker L (2017) The precision medicine initiative's All of Us Research Program : an agenda for research on its ethical, legal, and social issues. *Genet Med*, 19 : 743-750.
- 15) Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014) Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, 511 (7510) : 421-427.
- 16) Soldatos CR, Dikeos DG and Paparrigopoulos TJ (2000) Athens Insomnia Scale : Validation of an instrument based on ICD-10 criteria. *J Psychosom Res*, 48 (6) : 555-560.
- 17) Tadaka S, Saigusa D, Motoike IN, et al (2018) jMorp : Japanese multi omics reference panel. *Nucleic Acids Res*, 46 (D1) : D551-D557.
- 18) Wainschtein P, Jain D, Yengo L, et al (2019) Recovery of trait heritability from whole genome sequence data. *bioRxiv*, 588020.
- 19) Weedon MN, Lango H, Lindgren CM, et al (2008) Genome-wide association analysis identifies 20 loci that influence adult height. *Nat Genet*, 40 (5) : 575-583.
- 20) Yousaf A, Waltes R, Haslinger D, et al (2020) Quantitative genome-wide association study of six phenotypic subdomains identifies novel genome-wide sig-

nificant variants in autism spectrum disorder. *Transl*

Psychiatry, 10 (1) : 215–225.

■ ABSTRACT

Large-scale genomic and omics analysis and risk prediction of disease incidence in Tohoku Medical Megabank Project

So Umekage, Atsushi Shimizu

Division of Biomedical Information Analysis, Iwate Tohoku Medical Megabank Organization, Iwate Medical University

In advancing epidemiological study, the improvement of both cohort study and biobank is essential. Also, the development of statistical genetics approaches to identify disease risk factors is indispensable. In this article, we will review the recent biobank and cohort study to promote epidemiological research, and introduce the Tohoku Medical Megabank Project, which includes the largest three-generation birth cohort in the world. Next, we describe the “missing heritability”, which is one of the most difficult issue in cohort research, and show one of the hypotheses, the polygenic model, to overcome this. In addition, we introduce the iWate PolyGenic Model (iPGM) based on the polygenic model to predict genetic predisposition to Ischemic stroke. Finally, we discuss effective approaches toward prediction of multifunctional disorders such as psychiatric disorders through cohort study.

(Japanese Journal of Biological Psychiatry 32 (2) : 68–74, 2021)
