



東北大学



東北メディカル・メガバンク機構  
TOHOKU MEDICAL MEGABANK ORGANIZATION



2016年4月23日

東北大学 東北メディカル・メガバンク機構

**世界初、大規模コホートの健康調査情報・ゲノム解析情報等を統合する  
東北メディカル・メガバンク 統合データベース「dbTMM」を開発  
～1,070人分の統合された情報を制限付公開開始～**

＜発表のポイント＞

- ・健康調査及び全ゲノム配列情報を含む生体試料の解析情報を世界で初めて統合したデータベースを開発
- ・宮城県の地域住民コホート参加者1,070人分のデータ、制限付公開を開始
- ・高速かつ高度な検索を可能とする機能を開発
- ・ゲノム医療実現化の推進におけるプラットフォームとして期待

東北大学東北メディカル・メガバンク機構では、東北メディカル・メガバンク統合データベース「dbTMM」を開発しました。これは、大規模ゲノムコホート調査<sup>\*1</sup>由来の、健康調査及び全ゲノム配列情報を含む生体試料の解析情報（基本属性情報、調査票情報、生理学検査情報、検体検査情報、ゲノム・オミックス情報、診療情報、MRI画像情報等）を世界で初めて統合するものです。

「dbTMM」により、宮城県の地域住民コホート参加者1,070人分の基本属性情報、健康調査情報、全ゲノム配列情報を最初のリリースとして制限付公開を開始しました。「dbTMM」は、定められた登録・審査の手続きを経て、全国の研究者に利用されデータシェアリングの基盤となります。

「dbTMM」は、全ゲノム配列情報等に非常に強力な機能である、“大規模データ向け高速検索”や、“検索後層別化集団の統計学的自動特徴付け”等の新たな機能を備えています。精確な層別化と健康調査及び解析情報を統合したフェノタイプング<sup>\*2</sup>の提供が可能になり、新たな医療としてのゲノム医療実現化の推進のプラットフォームとして期待されます。

**【概要】**

東北大学東北メディカル・メガバンク機構（以下、ToMMo）は、宮城・岩手両県で岩手医科大学いわて東北メディカル・メガバンク機構とともに実施している、コホート調査の参加者の健康調査及び解析情報（基本属性情報、調査票情報、生理学検査情報、検体検査情報、診療情報、MRI画像情報、ゲノム・オミックス情報等）を網羅する統合データベース「dbTMM」（The database of Tohoku Medical Megabank）を開発することに成功しました。これら多様なデータを統合して格納する世界初のデータベースです。このデータベースからの最初のリリースとして、1,070人分の基本属性情報、健康調査情報、全ゲノム配列解析情報を、制限付公開<sup>\*3</sup>しました。本データベースは、定められた登録・審査の手続きを経て、全国の研究者に利用され、データシェアリングの基盤となります。

大規模なバイオバンク<sup>\*4</sup>は、先進国各国で次世代医療をつくる必須の基盤として構築が進められています。これまで構築されているバイオバンクでは、生体試料とそれに付随する性別、年齢等の基本属性情報や診療情報などを管理しているのが通例です。多くのバイオバン

クでは、自らゲノム解析等の大規模な解析は行わず、他の機関である解析センターに生体試料を提供して実施されてきました\*5。したがって、バイオバンクでは基本属性情報や診療情報、検体情報等を管理し、ゲノム情報は他の機関のデータベースに格納される、という別々の管理になることがほとんどです。

ToMMoでは、複合バイオバンクとして、バイオバンクと解析センター機能を併せ持ち、コホート調査の参加者の健康調査及び生体試料の解析情報の全て、即ち基本属性情報、調査票情報、生理学検査情報、検体検査情報、診療情報、MRI画像情報等のみならず、ゲノム・オミックス情報を網羅する東北メディカル・メガバンク 統合データベース「dbTMM」を開発しました。健康調査及び生体試料の解析情報の、大規模で多種多様な情報を統合することで、診療情報だけでなく、体質（ゲノム・オミックス情報）、体調（検体検査情報、生理学検査情報）、環境要因（調査票情報）などを統合した精確な層別化とフェノタイピングを提供することができ、ゲノム医療実現化の推進のプラットフォームとなります。この統合データベース「dbTMM」により、遺伝子と環境が複雑に作用して発症する多因子疾患の病因の解明などを、情報を安全に保ちながら、より多くの研究者が参画して研究が進められるように促進することができますと考えています。

また、今回の最初のリリースでは、「dbTMM」に登録される予定のデータのうち、2013年度のコホート調査参加者1,070人由来の、血液・尿の検体検査情報及び調査票（生活習慣）の回答情報、全ゲノム配列情報を、試料・情報分譲申請審査に基づく制限付公開をしました。

### 【背景と経緯】

東北大学と岩手医科大学は、2012年から東日本大震災の被害からの復興事業として、東北メディカル・メガバンク計画に取り組み、東北大学は東北メディカル・メガバンク機構（ToMMo）を、岩手医科大学はいわて東北メディカル・メガバンク機構（IMM）をそれぞれ設立して、事業を進めています。両機構は、宮城・岩手両県の住民15万人を対象にコホート調査を進めており、2016年4月現在で、13万人を超える協力を得ています。

コホート調査参加者から提供頂いた血液等の試料や調査票は、バイオバンクに格納されて、全国の多くの研究者の利用に供されます（試料・情報分譲）が、東北メディカル・メガバンク計画では多くの試料についてゲノム配列などの解析を両機構により行って、高精度で網羅的な解析データを分譲することを計画してきました。多様な解析により、質や種類の大きく異なる膨大なデータが生み出されますが、それらを効率よく有用な研究の対象とするためには、統合され整理されたデータベースの構築が不可欠です。

### 【統合データベースの概要】

統合データベースは、東北メディカル・メガバンク計画のバイオバンク事業において、前向きに収集した大規模な基本属性情報、調査票情報、生理学検査情報、検体検査情報、診療情報、MRI画像情報、ゲノム・オミックス情報等を統合したデータベースです。収集したすべての情報がひとつのデータベースに統合されていることで、あらゆるゲノム情報や健康調査情報や診療情報等に基づく表現型情報の属性でコホートを精確に層別化し、精度の高いゲノム医療実現のための解析研究を実施することができます。研究者は、試料・情報分譲の申請\*6にあたって、研究計画を立案するために、本バイオバンクにどのような試料・情報があるかを検索し、研究者からの申請と審査に従い、本バイオバンクの統合データベースから情報を抽出・分譲することができます。

試料・情報分譲審査委員会\*7における、平成27年度の分譲対象についての決定\*8にしたがって、2013年度に収集した試料・情報のなかで、データフリーズ\*9を終えたものを分譲対象とし、2013年度宮城県の地域住民コホート特定健康診査\*10の、1,070人分の下記の種類の基

本情報、健康調査情報、全ゲノム情報を格納した最初のリリースを制限付アクセスで公開しました。

統合データベースは大規模な全ゲノム解析に基づく2,100万の一塩基多様体の情報を格納する、文字通りビッグデータのデータベースであり、その能力を有効に活かせるよう高速な検索機能を開発しました。また、基本属性情報、調査票情報、全ゲノム情報など多岐にわたった項目に対して、すべての個別の項目で任意の検索を可能とする検索機能を開発しました。統合データベースでは、ゲノムの一塩基多様体情報などの網羅的な解析情報と、調査票に基づく生活習慣、採血に基づく一般検査や医用画像などに由来する従来の健康情報の、両方にわたる検索条件を指定できます。例えば、特定の染色体の特定の位置で特定の配列をもつ人において、ある生理学検査項目が特定の値以上の人が何人いるか、というような検索も可能になります。

健康調査情報、全ゲノム情報、診療情報、MRI情報、検体情報等の大規模データを格納するため、セキュリティを担保したデータベースの設計としました。また、検索後に得られる層別化した集団の属性情報について、母集団との有意差解析を実施して、統計学的に母集団と有意差のみられる属性を検定することで、検索後層別化集団の統計学的自動特徴付けを実現しました。大規模データを収載したデータベースの検索では、数千万にもものぼるすべての属性情報を表示して確認しながら検索することは不可能です。しかし、この検索後層別化集団の統計学的自動特徴付けの機能により、層別化された集団がどのような特徴のある集団であるかを確認しながら検索することが可能となります。この機能はゲノム・オミックス情報や健康・医療情報の大規模データについて、そのデータの特徴をよく理解する大きな助けとなり、ゲノム医療の研究を進めるうえで、データから得られるヒントとなることが期待されます。

なお、統合データベースの利用は、試料・情報分譲申請の利用者登録<sup>\*11</sup>を行った後、統合データベースの利用申請が必要です。

### 【今回制限付公開した情報の詳細】

1. 基本属性情報 年齢、性別
2. 健康調査由来の情報
  - (1) 血液・尿の検体検査情報
    - ・血液学的検査（血球計算、白血球分画等）
    - ・免疫学的検査（アレルギー検査等）
    - ・生化学的検査
    - ・尿検査
  - (2) 調査票（生活）情報
    - ・基本情報
    - ・運動について
    - ・飲酒について
    - ・喫煙について
    - ・ストレスについて
    - ・家族構成および健康状態について
      - 家族構成
      - 現在の健康状態
      - 輸血経験、生活習慣病の治療
      - 罹患歴
    - ・体質について
    - ・仕事の状況について
    - ・睡眠について
    - ・人とのつながりについて
    - ・東日本大震災の記憶について
    - ・東日本大震災での被災と体験について
    - ・女性の健康について
  - (3) 特定健康診査情報
3. 全ゲノム配列情報
  - ・個人毎変異情報（一塩基多様体）

## 【今後の計画】

今後、本事業の進展にあわせて、格納する対象者と種類の拡大に取り組んでいきます。2016年度以降、段階的に格納する情報の対象者規模と種類を拡大していき、最終的には、15万人分についてのデータベースを完成させて制限付公開する予定です。

また、格納されるこれらの情報はあまりに大規模であるため、有効に利活用するために大規模データ解析を実施し、その結果を提供していく予定です。そして、このコホートを層別化して利活用するにあたり、体質（ゲノム・オミックス情報）、体調（検体検査情報、生理学検査情報）、環境要因（調査票情報）などを統合したフェノタイピングを実施し、フェノタイプの情報を提供する予定で、これにより、より精度の高いゲノム医療実現の研究を促進することができると考えています。

## 【用語解説】

- \*1. コホート調査：ある特定の人々の集団を一定期間にわたって追跡し、生活習慣などの環境要因・遺伝的要因などと疾病の関係を解明するための調査。
- \*2. フェノタイピング：フェノタイプ（表現型）は、一般にある生物のもつ遺伝子型が形質として表現されたものです。ゲノム医療実現に向けた研究におけるフェノタイピングとは、診断病名、検査値、投薬などの診療情報、ゲノム・オミックス情報などを統合して、コホート参加者が罹患した疾患・疾患サブタイプをフェノタイプとして登録すること。
- \*3. 制限付公開（制限付アクセス）：利用申請の上で許可された場合の公開（アクセス）。
- \*4. バイオバンク：生体試料を収集・保管し、研究利用のために提供を行う。東北メディカル・メガバンク計画のバイオバンクは、コホート調査の参加者から血液・尿などの生体試料を集める。
- \*5. 例えば、大規模なバイオバンクとして英国のUKバイオバンクが有名だが、UKバイオバンクによって収集された試料のゲノム解析は主に別機関であるサンガー研究所によって担われている。
- \*6. 試料・情報分譲の申請：本年度より、データフリーズの終わった検体由来の試料・情報から段階的に分譲を開始している。現在、申請できる資格は、日本の学術研究機関（非営利機関）とそれ以外の日本に本社のある企業等の研究所に所属する研究者。
- \*7. 試料・情報分譲審査委員会：2013年、ToMMoは、試料・情報の限定的な利用や外部分譲に向けた仕組み作り、ルール制定・論点整理を行い、本格的試料・情報の分譲開始までのロードマップを作成することを目的とした試料・情報分譲室を設けた。2013年12月に、国内有識者による第1回試料・情報分譲審査委員会を開催し、現在までに8回の委員会が開催されている。
- \*8. 2015年度の方譲対象についての決定：2015年度は、以下のような情報が分譲可能。1) 採血・採尿によって得られた検査結果情報（血球値、白血球分画、生化学検査、総IgE、特異IgE値など）、2) 調査票による情報（身長、体重、生活習慣、運動量、飲酒、喫煙、ストレス、家族構成、現在の健康状態、罹患歴、仕事の状況、睡眠、人とのつながり、女性の健康、精神面（K6、CES-D））、3) 特定健康診査の結果、4) 遺伝子多型情報（一塩基多様体頻度情報、一塩基多様体・変異の個人毎変異情報）
- \*9. データフリーズ：対象とするデータについて、ある時点でデータクリーニングをし、固定すること。
- \*10. 地域住民コホート特定健康診査：地域住民コホート調査は、宮城県と岩手県の20歳以上の方に健康調査への参加をお願いしている。この調査の一部は、自治体を実施する特定健康診査の会場に、東北大学や岩手医科大学のスタッフが伺い、住民の方々に、本

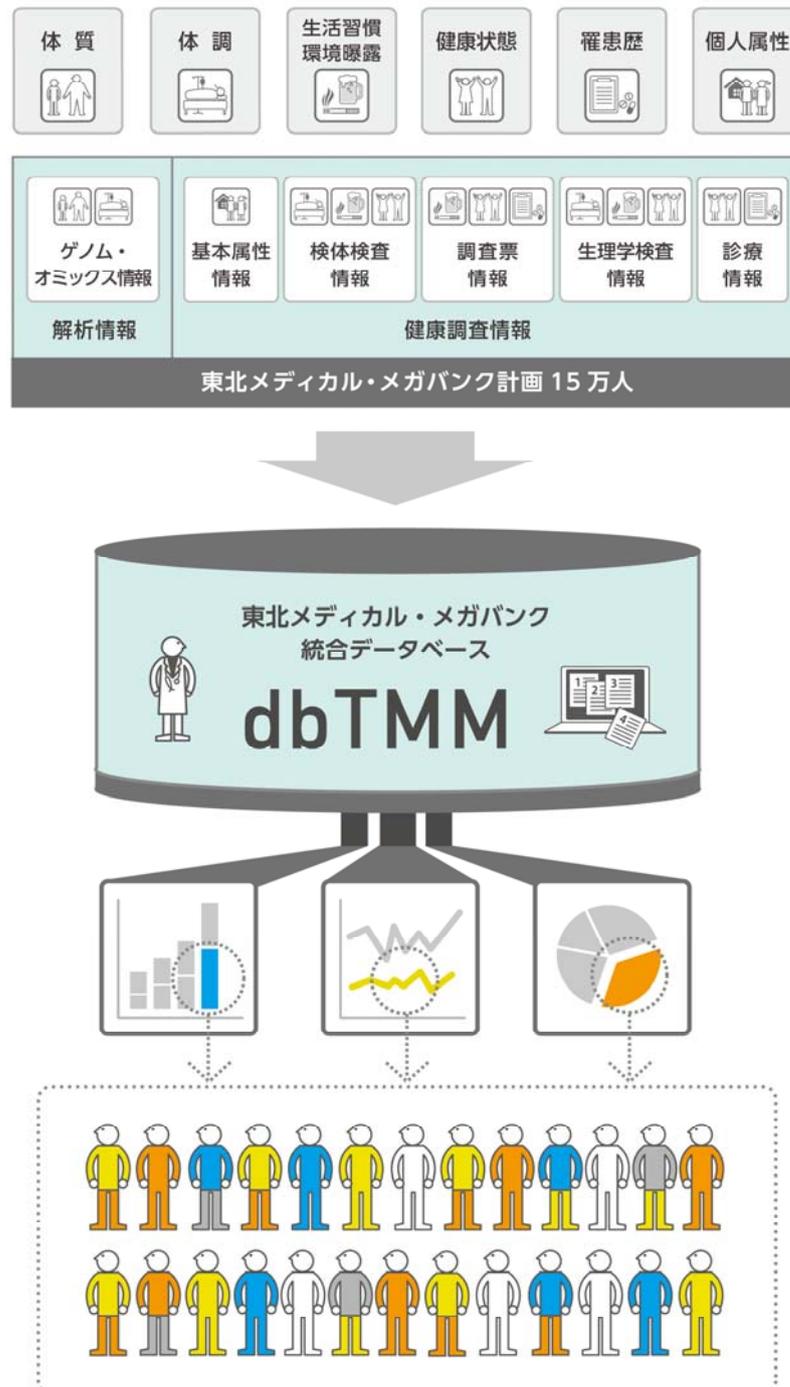
事業・調査について説明の上、同意を頂いた方々から、追加の採血および調査票への記入協力依頼等を行う。

\*11. 試料・情報分譲申請の利用者登録：利用者登録の手順は、バイオバンク試料・情報関連ウェブサイトの「試料・情報分譲申請の流れ」を参照。

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/>

【参考図】

1 東北メディカル・メガバンク 統合データベース dbTMM 概念図



## 2 東北メディカル・メガバンク 統合データベース dbTMM 画面構成例

### 詳細検索

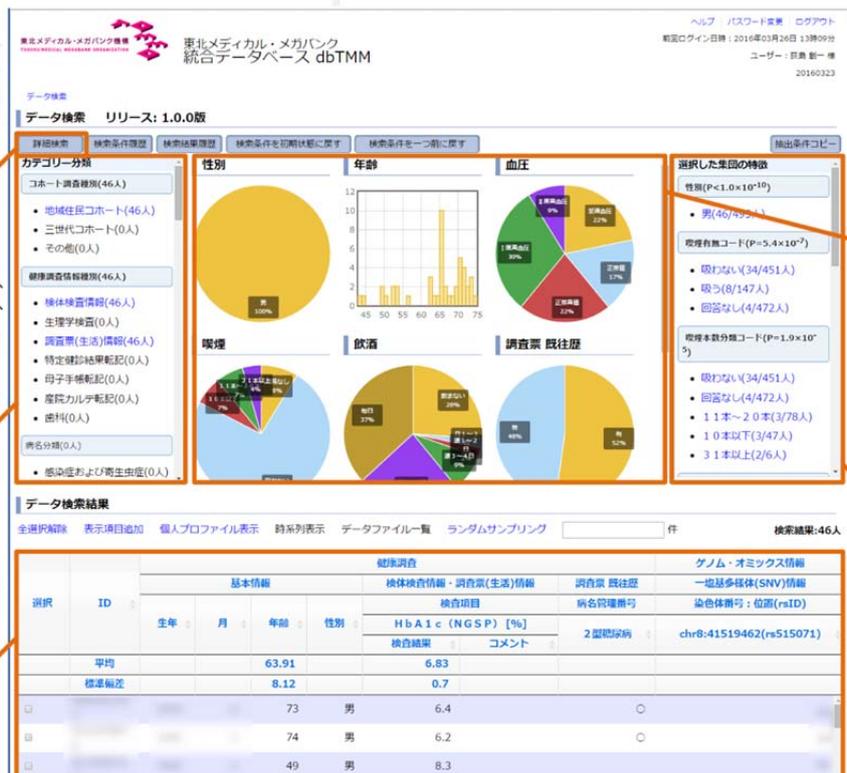
健康調査情報、全ゲノム配列情報の任意の項目の任意の条件での検索やその組合せ(AND/OR)による検索

### カテゴリ

コホート調査種別、健康調査情報種別、病名分類(ICD-10)、ゲノム・オミックス情報種別等のファセットによる該当件数・検索

### データ表示

検索により層別化した集団のデータの任意の項目の表示



### 層別化した集団のグラフ

検索により層別化した集団の代表的な項目のグラフの表示

### 層別化した集団の特徴

検索により層別化した集団の特徴付けとして、層別化した集団と母集団とで比較して、有意差のある項目の表示

## 【参考】

### <東北メディカル・メガバンク計画について>

本計画は、東日本大震災を受け、被災地住民の健康不安の解消に貢献するとともに、個別化予防等の東北発の次世代医療を実現するため、ゲノム情報やオミックス情報を含むコホート研究等を実施し、被災地域の復興を推進する、国の復興事業として行われているものです。平成27年度より、国立研究開発法人 日本医療研究開発機構が本計画の研究支援担当機関の役割を果たしています。

被災地に医療関係人材を派遣して地域医療の復興に貢献するとともに、15万人規模の地域住民コホートと三世代コホートを形成し、そこで得られる生体試料、健康情報、診療情報等を収集してバイオバンクを構築します。さらに、ゲノム情報、オミックス情報、診療情報等を解析することで、個別化医療等の次世代医療に結びつく成果を創出することを目指しています。また、得られた生体試料や解析成果を同意の内容等に十分留意し、個人情報保護のための匿名化等の適切な措置を施した上で、外部に提供することや、コホート調査や解析研究を行うための多様な人材の育成も行っています。

本計画の事業の実施は、東北大学東北メディカル・メガバンク機構と岩手医科大学いわて東北メディカル・メガバンク機構とが連携して行っています。

**【お問い合わせ先】**

(統合データベースに関すること)

東北大学東北メディカル・メガバンク機構

統合データベース室長

准教授 荻島 創一 (おぎしま そういち)

電話番号：022-274-6032

Eメール：[ogishima@megabank.tohoku.ac.jp](mailto:ogishima@megabank.tohoku.ac.jp)

(報道に関すること)

東北大学東北メディカル・メガバンク機構

広報戦略室長

長神 風二 (ながみ ふうじ)

電話番号：022-717-7908

ファックス：022-717-7923

Eメール：[f-nagami@med.tohoku.ac.jp](mailto:f-nagami@med.tohoku.ac.jp)

(AMED 事業に関すること)

日本医療研究開発機構 (AMED)

バイオバンク事業部 基盤研究課

電話番号：03-6870-2228

Eメール：[kiban-kenkyu@amed.go.jp](mailto:kiban-kenkyu@amed.go.jp)