



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



東北大学

2016年7月12日

報道機関各位

日本電信電話株式会社
国立大学法人東北大学 東北メディカル・メガバンク機構

複数の研究機関が持つゲノムデータを相互に開示せず 分析する解析手法を開発

～プライバシー保護データマイニング技術によるフィッシャー正確確率検定を世界で初めて実現～

日本電信電話株式会社（以下 NTT、本社：東京都千代田区、代表取締役社長：鶴浦博夫）と国立大学法人東北大学 東北メディカル・メガバンク機構（以下、ToMMo）の三澤計治助教（バイオメディカル情報解析分野）、荻島創一准教授（バイオクリニカル情報学分野）、長崎正朗教授（バイオメディカル情報解析分野）の研究グループは、ゲノム情報を暗号化したまま、複数の研究機関が持つゲノム情報を相互に開示することなく正確に分析する、プライバシー保護データマイニング技術によるゲノム解析手法を共同で開発しました。近似であるカイ二乗検定による解析ではなく、正確性を期したフィッシャー正確確率検定^{*1}を世界で初めて実現し、疾病情報と遺伝子との相関を調査するのに要する時間が、従来の1年以上から約20分に短縮可能な手法（1000人分のデータによる試算）を開発することで、プライバシー保護データマイニング技術によるゲノム解析の実現に必要な基盤技術を確立しました。本成果によって、複数の研究機関が安全にゲノムデータを持ち寄り分析することで、医療の更なる発展へとつながることが期待されます。

なお本成果は、NTTとToMMoによる共同研究「ゲノム情報を分散・秘匿化したままセキュアな環境で情報解析するための基盤技術開発のフェジビリティスタディ」によるもので、2016年7月14日より開催されるコンピュータセキュリティ研究会の研究発表会にて発表します。

1. 研究の背景・経緯

近年のDNA配列決定技術の進歩により、大量のゲノムデータが解析されるようになりました。これに伴い、ゲノムワイド関連解析（Genome Wide Association Study: GWAS）^{*2}などの手法を用いて、疾病情報と遺伝子の関連性を発見する研究が国際的にも進められています。また、複数の研究機関が保有するゲノムデータを横断的に分析することで、より大規模なデータを使って分析を進めたいというニーズも高まっています。

しかしながらゲノムデータは、個人の遺伝情報という機微性が高い性質のため慎重な取り扱いが求められ、複数の研究機関の間で相互に制限なく開示し合うことは容易ではありませんでした。このため、複数の組織が保有する情報を安全に活用する技術として、プライバシー保護データマイニングと呼ばれる技術が注目されており、その手法の1つとしてNTTではデータを暗号

化したまま計算する秘密計算技術^{※3}を研究開発しています。

この度、NTT が秘密計算技術を、ToMMo がゲノム解析技術を担当し、複数の研究機関が互いにゲノムデータを開示することなく、暗号化したままで比較し、かつゲノムワイド関連解析において正確性を期すために必要となるフィッシャー正確確率検定を現実的な時間で処理できる手法を共同で開発しました。

2. フィッシャー正確確率検定の概要と課題

ゲノムワイド関連解析では疾病情報と遺伝子の関連性を発見するために、ゲノムデータを患者群と対照群に分け、数百万の変異個所について遺伝子変異の有無を集計した分割表を作成します。それぞれの分割表に対し、検定という統計手法を用いて患者群と対照群に遺伝子変異の有無との関連性があるかどうかを調べます。有意的な関連性がある場合は、その遺伝子個所と疾病情報に何らかの関連があることが分かります。秘密計算技術を用いて複数の研究機関がゲノムデータを暗号化したままゲノムワイド関連解析を行うことで、それぞれがゲノムデータを開示することなく、より高い信頼性の下で疾病情報と遺伝子の関連性を分析することができます（図1）。

これまで、カイ二乗検定を用いてゲノムデータを暗号化したままでゲノムワイド関連解析を実現した先行研究がありました。しかし、カイ二乗検定は近似であり、分割表に現れる数字が小さい時には、統計検定の結果が正しくなくなる恐れがあるため、このような場合にはフィッシャーの正確確率検定を行う必要があります。しかし、フィッシャー正確確率検定では、大きな数の階乗計算が必要となり、かつ数百万もの遺伝的多様性を1つずつ解析する必要があることから、これまでの方式では秘密計算技術の適用が困難という課題がありました。

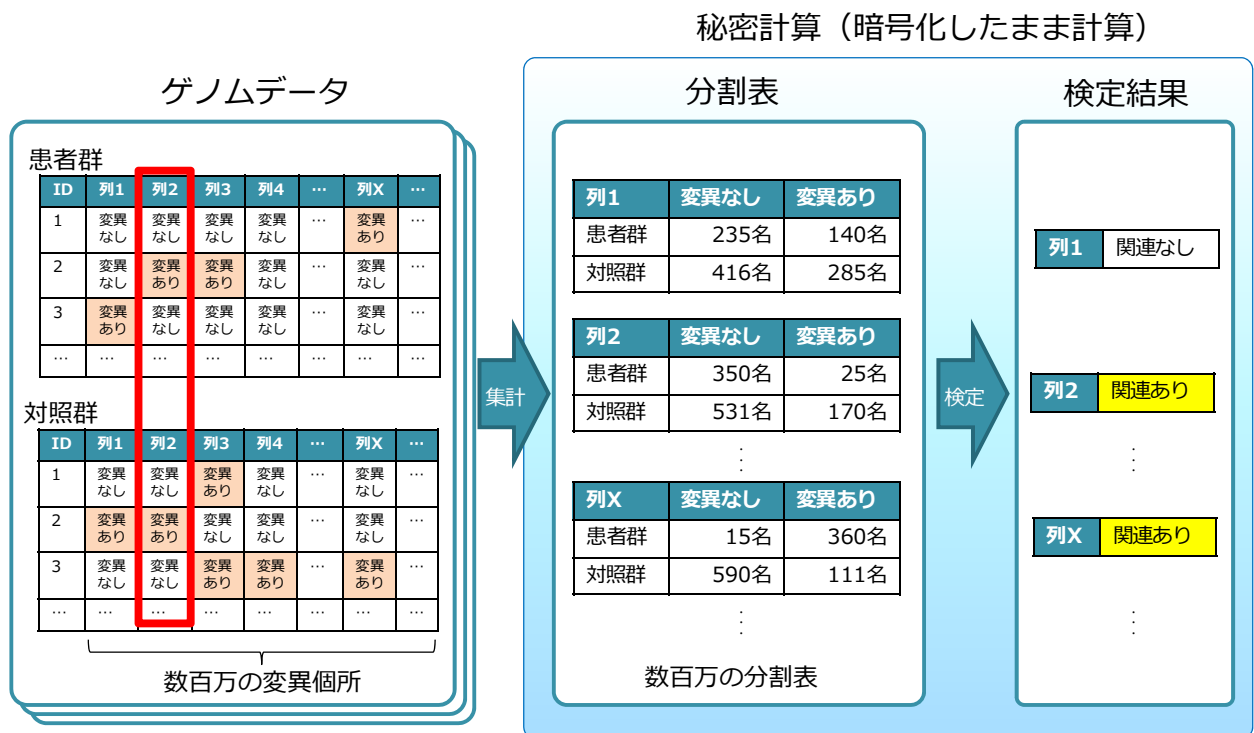


図1. ゲノムワイド関連解析による疾病情報と遺伝子の関連性分析

3. 今回開発した技術の特長

NTT および ToMMo は、ゲノムデータを暗号化したまま行う数百万回ものフィッシャー正確確率検定を、以下の3つの方法で計算時間を削減し、従来の1年以上から約20分程度に短縮可能な手法（1000人分のデータによる試算）を開発しました（図2）。

(i) フィッシャー正確確率検定を秘密計算で効率良く実現する方法

- ・フィッシャー正確確率検定で関連性があるとみなされる、全てのありうる分割表の数字パターンを書き出しておき、その中に検定を行う変異箇所の分割表のパターンが含まれるかどうかを判定する方法。
- ・階乗計算結果の対数をテーブルに記憶し、必要な対数を秘密計算で効率良く取得する方法。

(ii) 検定実行回数を削減する方法

- ・フィッシャー正確確率検定法よりも簡単な演算で遺伝子変異箇所候補を絞り込む方法。

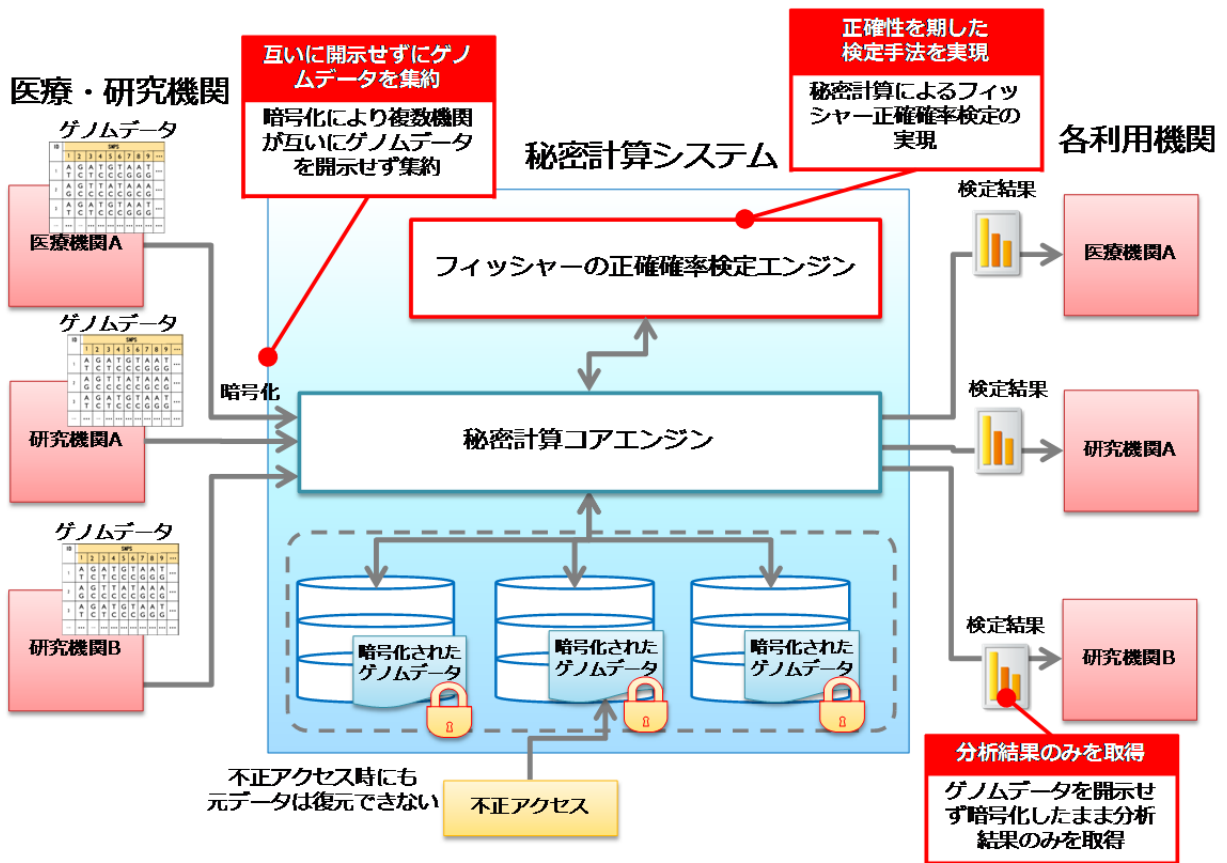


図2. 秘密計算システムによるゲノム解析

4. 今後の展開

今回開発した技術により、1000人規模のゲノムデータに対して安全かつ効率的な解析が可能となりました。今後は、それ以上の規模のゲノムデータに対しても、より効率的な解析が可能な手法の検討を行います。また、フィッシャー正確確率検定やゲノムワイド関連解析以外の重要なゲノム解析手法についても秘密計算技術によるゲノム解析を実現する手法の開発に取り組むことで、医療の更なる発展へとつながることが期待されます。

本成果は、2016年7月14日～15日に開催されるコンピュータセキュリティ研究会の研究発表会にて発表します。

(会場：中市コミュニティーホール Nac 〒753-0086 山口県山口市中市町 3-13)

発表演題名： 秘密計算フィッシャー正確検定(1) ～標本数が少ない場合 主著者： NTTセキュアプラットフォーム研究所主任研究員 千田浩司
発表演題名： 秘密計算フィッシャー正確検定(2) ～標本数が多い場合 主著者： NTTセキュアプラットフォーム研究所研究員 濱田浩気
発表演題名： プライバシ保護ゲノム解析のための秘密計算フィッシャー正確検定 主著者： NTTセキュアプラットフォーム研究所 長谷川聡

【用語解説】

- ※1 フィッシャー正確確率検定：
2つ以上のカテゴリーの独立性について判定を行う方法。
- ※2 ゲノムワイド関連解析 (Genome-Wide Association Study: GWAS)：
ゲノム全体に渡り患者群とコントロール群との遺伝子頻度を統計的に解析することによって、疾患関連遺伝子を発見する手法。
- ※3 秘密計算技術：
データを暗号化したまま処理し、互いにデータを一切秘密にしたまま、処理結果のみを得ることができる技術。

【本件に関するお問合せ先】

日本電信電話株式会社
サービスイノベーション総合研究所
企画部広報担当

E-mail : randd@lab.ntt.co.jp

Tel : 046-859-2032

国立大学法人東北大学 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)

E-mail : f-nagami@med.tohoku.ac.jp

Tel : 022-717-7908