

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

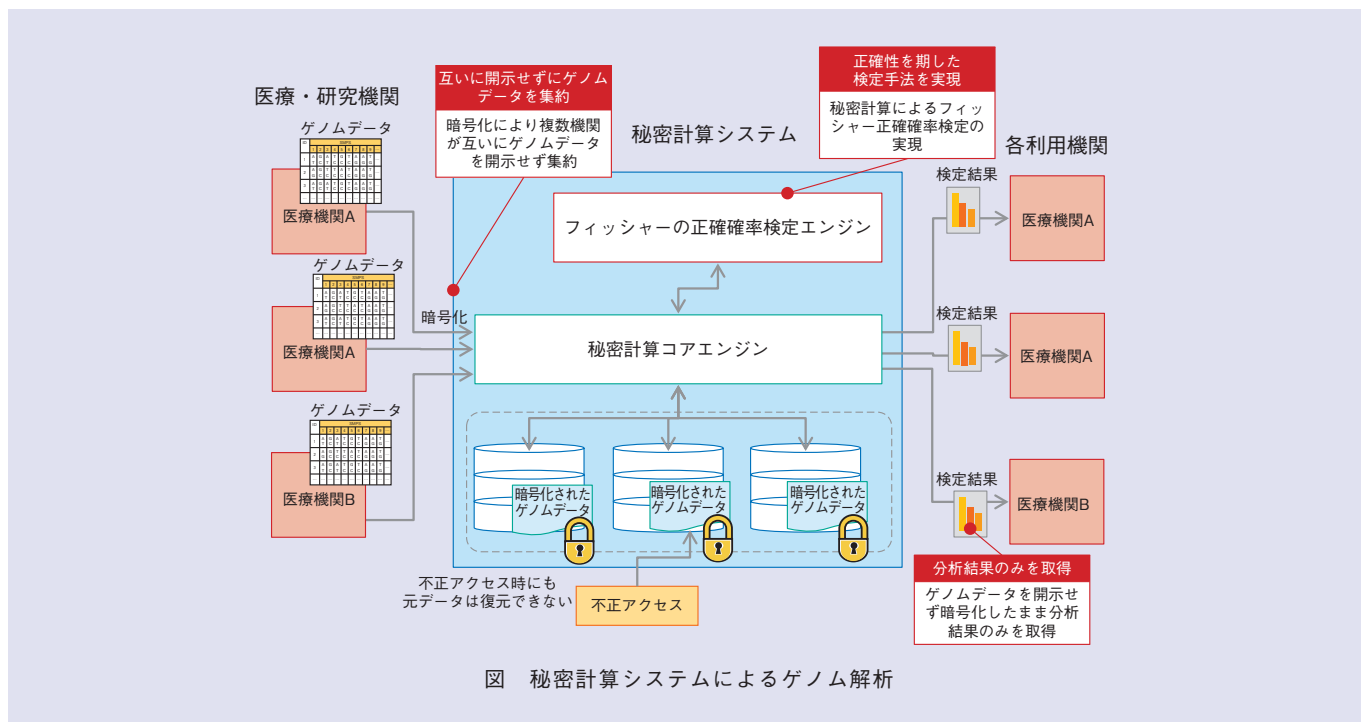
NTTと国立大学法人東北大学 東北メディカル・メガバンク機構 (ToMMo) の三澤計治助教 (バイオメディカル情報解析分野), 荻島創一准教授 (バイオクリニカル情報学分野), 長崎正朗教授 (バイオメディカル情報解析分野) の研究グループは, ゲノム情報を暗号化したまま, 複数の研究機関が持つゲノム情報を相互に開示することなく正確に分析する, プライバシー保護データマイニング技術によるゲノム解析手法を共同で開発しました。近似であるカイ二乗検定による解析ではなく, 正確性を期したフィッシャー正確確率検定*¹を世界で初めて実現し, 疾病情報と遺伝子との相関を調査するのに要する時間が, 従来¹年の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, 医療のさらなる発展へとつながることが期待されます。

■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 (GWAS: Genome Wide Association Study)*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

- *1 フィッシャー正確確率検定: 2つ以上のカテゴリ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: ゲノム全体にわた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。



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を研究開発しています。

このたび、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。

■フィッシャー正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

個人のゲノム情報を守りつつ社会全体で利活用できるツールをめざして

研究者 紹介

長崎 正朗

東北大学
東北メディカル・メガバンク機構
教授

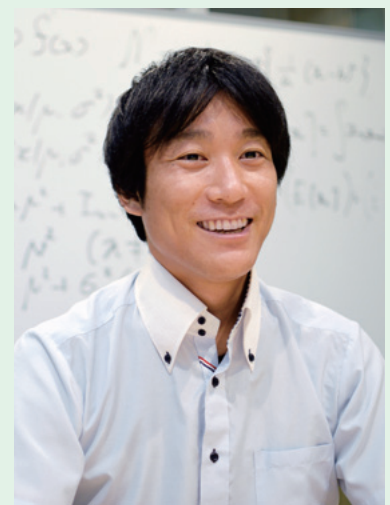
ゲノム情報をより深く知ることによっていろいろなことが分かってきています。今後、大規模なゲノム情報を統合解析することで、病気やさまざまな形質（例えば髪の毛の色や目の色）と関係する個人ごとのゲノム情報の違いがますます分かってくると考えられています。お酒に強い弱いゲノム情報で決まっているように、さまざまな薬の効き具合の個人差がゲノム情報の違いによって説明できるようになると考えられます。このことで、自分に合う薬を適切な量服用できるようになるなど、さまざまな恩恵を将来的に受けられると考えられています。

このように重要なゲノム情報ですが、パスワードと違い一生変えることができない個人個人が持つ大切な情報になります。

しかし、まだまだ分かっていないことも多く、前述の個人ごとの薬の効き具合などの情報は、対象とする薬が効果があったかどうか過去情報と各個人のゲノム情報を大規模に収集して関連解析をすることで新たな関係を見つけていく必要があります。そのために一般的に用いられる手法の1つが今回秘密計算技術の開発対象としたフィッシャー正確率検定になります。

今回開発した方法を用いることで、複数の医療機関の個人個人のゲノム情報は秘匿したまま、1000人規模の個人ごとのゲノム情報と対象とする属性（例えば薬が効くかどうか）との関係を短時間で解析できるようになります。今後の課題として、より多くの人数でも実時間で処理できるように改良をし、より汎用性の高い手法となることを目標としています。

ゲノム情報の利用は、関係する法律や指針の改訂などもあり、さらなる慎重さが求められており、ますます秘密計算技術は1つの不可欠なツールとなると考えています。



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。

■今回開発した技術の特長

NTTおよびToMMol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、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）。

◆問い合わせ先

NTTサービスイノベーション総合研究所

企画部広報担当

TEL 046-859-2032

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

URL <http://www.ntt.co.jp/news2016/1607/160712a.html>

異分野技術交流による苦勞とシナジー効果を実感

研究者 紹介

千田 浩司

NTTセキュアプラットフォーム研究所
主任研究員

データを暗号化したまま計算可能とし情報漏洩の問題解決をめざす「秘密計算」技術の研究をはじめ10年以上経ちますが、その間、秘密計算の研究開発は大きく進展し、実用化への期待が年々高まっているのを感じます。10年前は少し複雑な計算でも数時間かかるため、当時のハイスペックサーバ20台を使って平均値等の単純な計算をデモンストレーションしていました（世界トップ性能でしたが、「真空管式コンピュータよりも遅いね」とコメントいただいたのが印象的です）。

現在は度重なる情報漏洩事件や機微な情報を利活用しようとする動きを背景に、秘密計算の実用化に向けた研究開発プロジェクトが国内外で多くみられます。ゲノム情報は非常に機微であり、かつ創薬など利活用の社会的価値が高い情報でもあるため、ゲノム解析の研究者と秘密計算の研究者が手を組み「ゲノム解析秘密計算」の研究開発が近年急速に立ち上がってきています。

いろいろなお縁があり、東北大学 東北メディカル・メガバンク機構のゲノム解析の研究者とゲノム解析秘密計算の共同研究をする機会を得ました。はじめは互いの最低限の知識を理解するだけで四苦八苦でしたが、2年目半ばになる今ではシナジー効果を実感しています。特に今回の成果は、我々がこれまで培った秘密計算の技術やノウハウを結集させてゲノム解析の研究者のニーズにこたえたものであり、世界をリードするゲノム解析秘密計算技術です。今後は大規模データ対応等の課題もありますが、まずはできることから早期に実用化させ、安心・安全なゲノム情報利活用によるゲノム解析の発展に寄与できれば幸甚です。

