ゲノム情報の臨床への応用展開におけるセキュリティ技術の紹介

Our Security Technology For The Clinical Application

野原 祥夫*　松田　規**

Sachio Nohara, Nori Matsuda

医療分野で開発が進められている個別化医療では、個人のヒトゲノムを解析したゲノム変異データを安全に保管しつつ診断に必要な情報を的確に医師に提供するセキュリティ技術が求められている。当社では、三菱電機株式会社が開発した秘匿検索技術を応用した暗号化技術により、ゲノム変異データの安全な利用を実現する技術開発に取り組んでいる。本稿では、秘匿検索に用いられる暗号化技術の特徴を説明し、個別化医療への対応について紹介する。

At the clinical sequencing era, high security technology for data management is necessary for personalized medicine on personal genome variants. We are developing the next-generation technology for safety access to personal genome information with using encrypted search technology provided by Mitsubishi Electric Corporation. In this article, we will show the feature of encrypted search technology and how that technology was adapted.

1. まえがき

ヒトゲノム解読のコストが個人でも利用できるレベルにまで低下してきたことから、従来から提唱されていた“Personalized Medicine（個別化医療）”が、ようやく実現されつつある。とくに米国は、“Precision Medicine Initiative”(1)により、がん・希少疾患とゲノム変異データについて100万人以上を対象とした調査を行い、適切な治療法や発症予防法の開発に集中投資する方針である。世界的にもゲノム変異データを活用した個別化医療拡大の流れは決定的なものになっている。

個別化医療においては、個人のヒトゲノムを表すゲノム変異データが治療法等のパラメータとして利用されるため、究極の個人情報と呼ばれるゲノム情報におけるプライバシー保護とデータセキュリティの強化が必須の課題である。米国“Precision Medicine Initiative”においても医療システム間で安全にデータ交換を行う情報管理技術の開発が主な投資項目の一つに挙げられている。

当社においては、ゲノム変異データの利用におけるプライバシー保護とデータセキュリティを向上させるために、医療システムへのアクセスに用いられる認証システムと暗号技術の改良を進めている。これらの技術は、金融機関等のクリティカルな運用を求められるコンピュータシステムやソフトウェアサービスにおいて既に用いられている技術と同等以上の安全性を目指している。

個別化医療においては、現在全国各地で運用が進められている医療ネットワーク上でゲノム変異データのやり取りが行われると考えられている。医療ネットワークでは、既に電子カルテ等の診断情報の共有化が推進されており、現行のシステムアクセスの認証方式では認証データのコピー等を完全に防ぐことができないため、より安全性の高い認証方法が望まれている。たとえば、トルコ共和国においては医療システムへのアクセス方法として本人認証の決定版といえる指静脈認証技術が応用されており(2)、当社でもゲノム変異データへのアクセスに指静脈認証技術を応用した遠隔アクセス方式を開発した。

データセキュリティについては、これまでも外部からの不正アクセス、関係者のミスや不正による情報漏えいへの対策としてデータベースの暗号化技術が用いられてきた。ただし、従来の暗号化技術にはデータにアクセスする際にメモリ上に復号された平文が展開されるという大きな欠点があり、メモリ上のデータを盗聴するマルウエアの存在が確認された(3)。その対策として暗号化した状態で様々な処理を行う「暗号化状態処理技術」の実用化が強く望まれている(4)。

また、個別化医療において、健康維持や疾患予防の段階では、自身の身に将来起こりえる健康上の問題を相談する遺伝カウンセラとゲノム変異データを共有し、発症

＊関西事業部　バイオメディカルインフォマティクス開発室　＊＊三菱電機株式会社

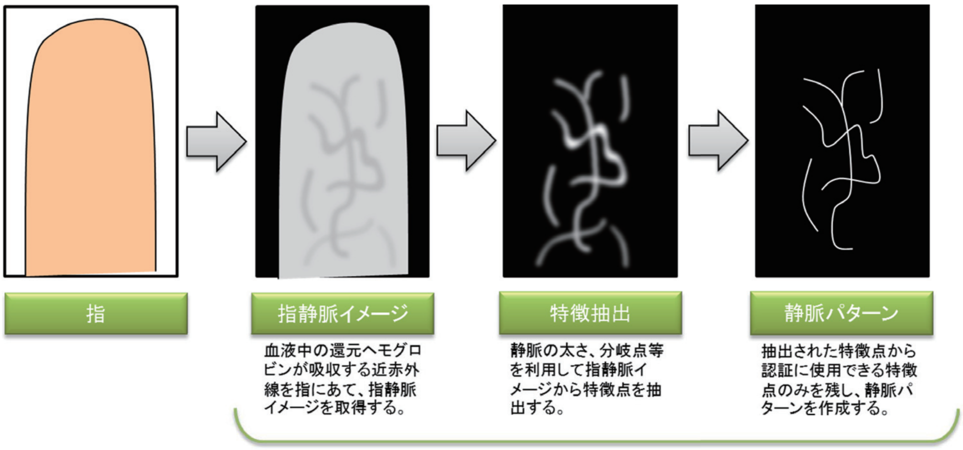
後の診療の段階では担当医師にゲノム変異データを提供する必要がある。さらに、ゲノム変異データと疾患の関係は非常に複雑で未知の部分が多いことから研究者による二次利用も強く望まれており、これらのニーズに対して安全にアクセス制御を実現できる仕組みが不可欠である。当社では、暗号技術の改良により「暗号化状態処理技術」と安全なアクセス制御技術を実現する次世代の暗号技術として、秘匿検索技術を応用したゲノム変異データベース暗号化技術の開発を進めている。

2. 指静脈認証

2.1 指静脈認証方式の原理

現在、個人認証の安全性を強化する場合は、主にワンタイムパスワードや指紋認証が用いられている。しかし、ワンタイムパスワードでは発行機器の紛失・盗難の恐れがあり、指紋認証では指紋テープ等の偽造による不正アクセスを完全に防ぐことは非常に困難である。これらの問題を解決する方式として、静脈パターンを利用した認証システムが開発されてきた。静脈認証の原理は、図１に示すように静脈の血管パターン画像から抽出した血管の分岐等の特徴点をマッチングするもので、この技術は、経年変化に強く、細部の形状が複雑で偽造することが困難であり、かつ認証精度が高いという特徴を有している（表１）。

当社では、小型の認証デバイスで静脈認証を実現可能な株式会社モフィリア社[*1]製の指静脈認証装置を採用した。この認証装置は、CMOSセンサと反射散乱方式の組み合わせで機器が小型化されており、ワンタイムパスワードや指紋認証の置き換えとしては最適であると考えている。



血液中の還元ヘモグロビンが吸収する近赤外線を指にあて、指静脈イメージを取得する。

静脈の太さ、分岐点等を利用して指静脈イメージから特徴点を抽出する。

抽出された特徴点から認証に使用できる特徴点のみを残し、静脈パターンを作成する。

指の角度などにより、静脈パターンが異なる可能性があるため、認証に使用する静脈パターンは一般的に複数回取得する。

図１　静脈認証の原理

表１　生体認証方式の比較

	静脈	指紋	顔	虹彩	音声	サイン
認証精度	◎	◎	△	◎	△	△
偽造容易性	◎	△	△	○	△	△
経年変化耐性	◎	△	△	○	△	○
心理的抵抗性	○	△	○	△	◎	◎
サイズ	○	◎	◎	△	◎	○
導入コスト	○	◎	○	△	○	○
登録可能ユーザ数	◎	○	◎	◎	◎	◎

静脈は構造が複雑であるため、偽造が難しく、年齢を重ねても構造が変化しないという特性を持っている。また、体内にあるため、傷などの外的要因にも強い。

【評価項目説明】
認証精度…生体情報を利用した認証の精度の高さ
偽造容易性…生体情報を偽造するし易さ
経年変化耐性…年齢を重ねたときの生体情報の変化のしやすさ
心理抵抗性…認証装置を使うときの心理的な抵抗感
サイズ…認証装置のサイズ
導入コスト…認証装置を導入するときのコスト
登録可能ユーザ数…認証に使える生体情報を登録できるユーザ数

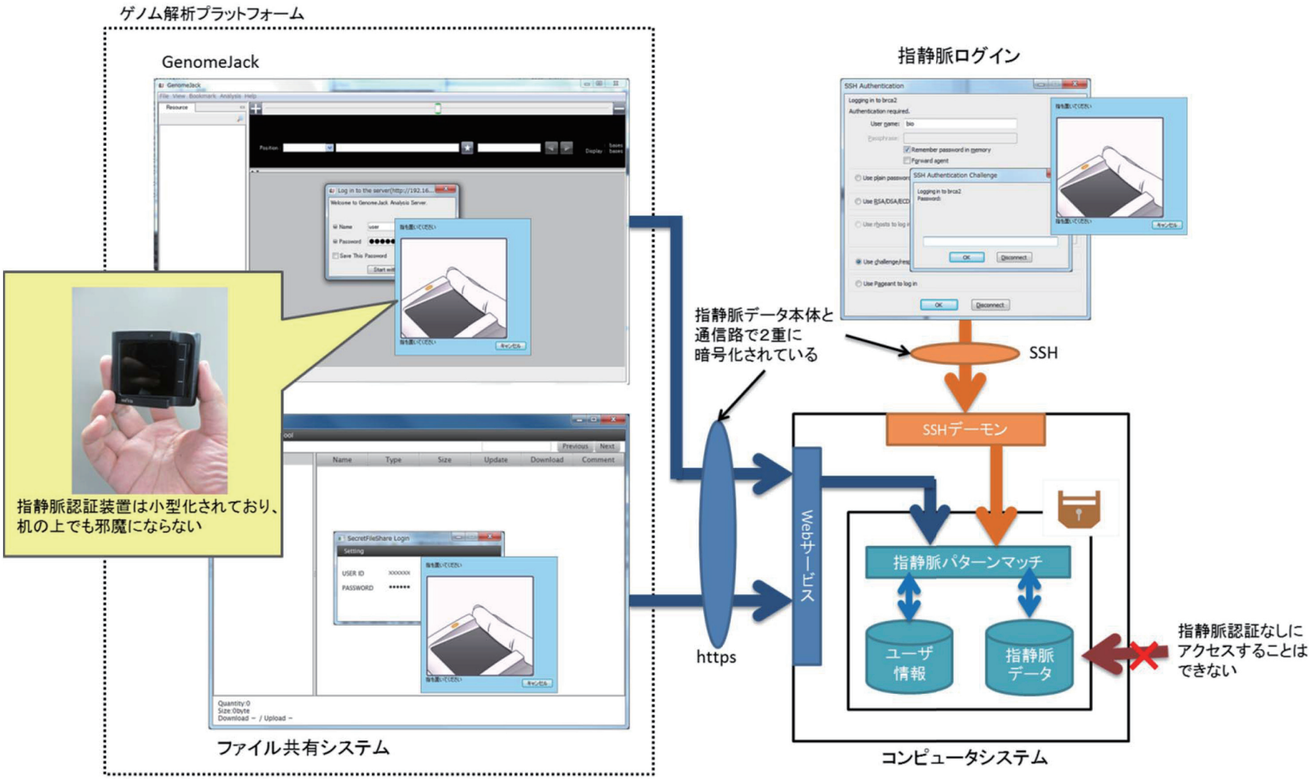


図２　指静脈認証を導入した当社製品のシステム構成

2.2　当社製品への指静脈認証の導入

当社では、個別化医療に用いられる次世代ゲノムシーケンサのデータ解析ソフトウェアシステムとして、GenomeJack(5)を製品化している。GenomeJackは、ゲノム変異データをコンピュータシステムに登録するファイル共有システム、データ解析を行う解析ソフトウェア群からなるサーバソフトウェアとクライアント側で保管されたデータや解析結果を閲覧するゲノムブラウザから構成されている。今回GenomeJackを改良して、ゲノムブラウザのアクセス認証に通常のユーザ名/パスワード認証に加えて指静脈認証も採用した。また、GenomeJackのサーバが稼働するコンピュータシステムのメンテナンス時にシステム管理者がコマンドライン端末にアクセスする場合の安全性向上にも配慮し、遠隔からのコマンドライン端末へのログイン認証に指静脈認証デバイスを連携させる仕組みを開発した（図２）。

3.　秘匿検索技術のゲノム変異データへの応用

3.1　秘匿検索技術の特徴

暗号化状態でデータ検索が可能な暗号化状態処理技術は様々な方式が開発されてきている。その中でも、当社が使用する秘匿検索技術*2は、数学的にも世界トップレベルの安全性を有すると考えられている内積述語暗号(6)(7)を応用した技術である。

秘匿検索技術では、データベース、通信経路、メモリ上において処理状態にかかわらずデータは常に暗号化された状態であるため、データが盗聴された場合でも平文の暴露を防止することができる。図３は、既存の暗号化検索技術として一般に用いられているTDE（Transparent Data Encryption）との秘匿検索技術の暗号化状態の一貫性の差を示している。TDEは、サーバ側で検索処理が実行される直前に復号されるため、メモリダンプによってマルウエアやサーバ管理者の盗聴のリスクが避けられない。一方、秘匿検索技術では、平文の検索キーワードは即座に暗号化され、そのままの状態でサーバ側に送信されて処理が実行される。検索結果の閲覧時においてもヒットしたレコードの復号化はクライアント側での閲覧直前に実行されるため、サーバや通信路で平文の暴露が一切なく、安全性が非常に高いという特徴を持つ。

常時暗号化された状態で検索を実現する技術としては、同じ元データは必ず同じ暗号化データになるという確定的暗号方式がよく用いられる。確定的暗号方式では検索などの処理時においても暗号化状態を維持できる完全な暗号化状態処理を実現できるが、平文データの文字パターン出現頻度がそのまま暗号化データの文字パターン出現頻度になるため、文字出現頻度に偏りがあるデー

＊1　http://www.mofiria.com/

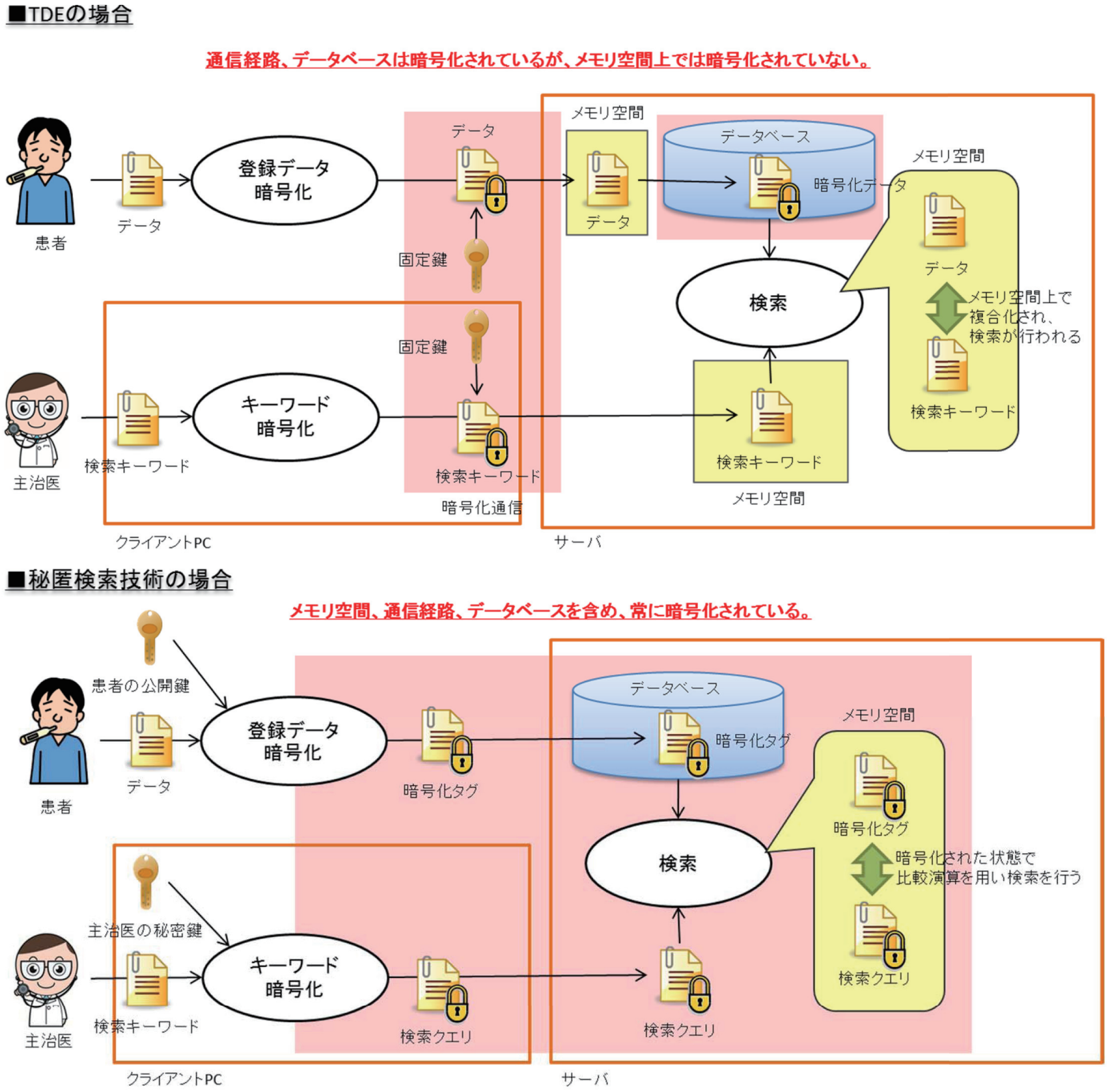


図３　秘匿検索技術とTDEの暗号化状態の比較

表２　既存暗号化検索技術との比較

	本方式	TDE	確定的暗号方式
サーバ管理者に対する機密保護	可能	閲覧可能	可能
検索者のアクセス制御	データ毎に可能	列ごとに可能	列ごとに可能
検索処理の性能	低速	高速	高速
検索の柔軟性	完全一致のみ	各種検索が可能	完全一致のみ

タの場合は、暗号化データから平文を類推されてしまうリスクが避けられない。秘匿検索技術は、内積述語暗号の特徴を生かして、暗号化データのランダム化が可能である。よって、サーバ管理者の不正などにより、大量にデータが流出した場合にも、平文の類推を防止することが可能である。

このように、秘匿検索技術は非常に安全性の高い暗号技術であるが、検索には複雑な暗号演算が必要で処理時間はNを暗号化データ数とするとO（N）のオーダーとなり、長くなる。また、原理的には検索キーワードの完全一致しか対応できない（表２）。そこで当社では、ゲノム変異データの特性を利用した暗号演算の並列化や変

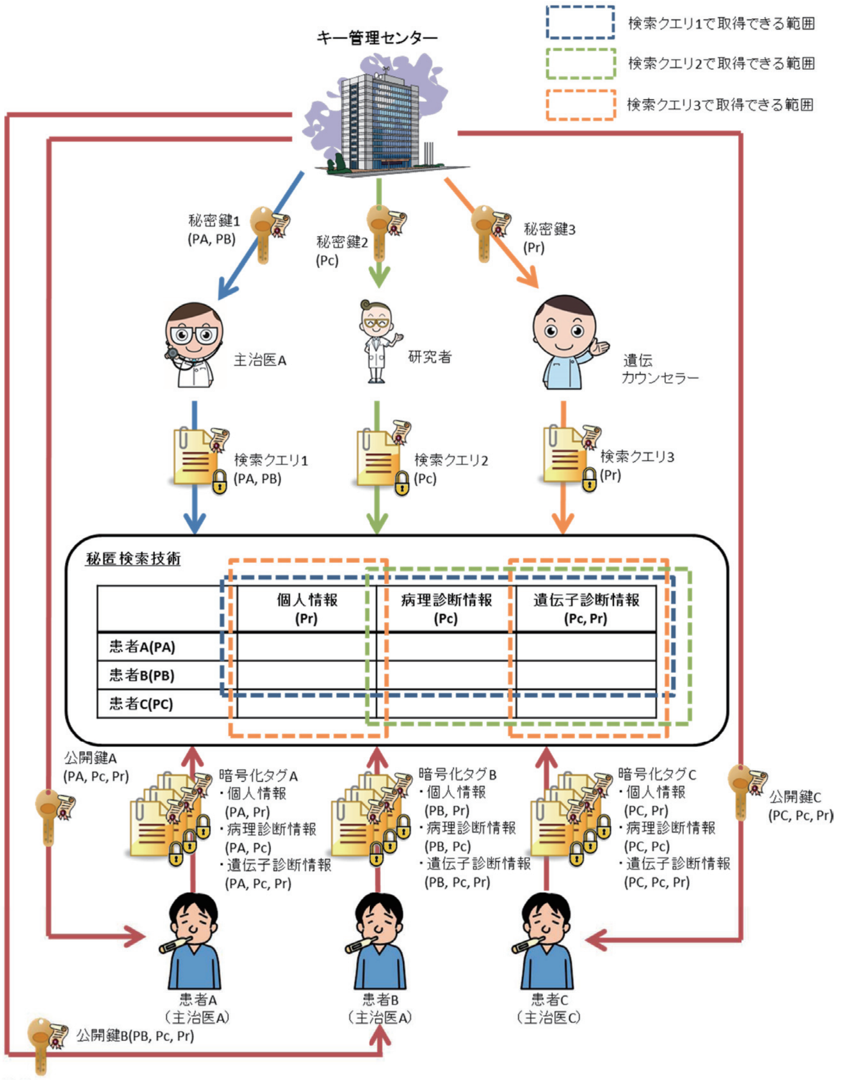
異発生パターンに適したデータ構造の設計により処理の高速化と疑似的な範囲検索を実現した*3。

3.2　秘匿検索技術によるアクセス制御の実現

一般的な暗号化状態処理技術である確定的暗号方式は１種類の秘密鍵による単純な仕組みであり、暗号化された情報をコンピュータネットワーク上で共有利用する形態には適していない。そこで、データへのアクセス制御を利用者のグループ単位で実現する鍵共有管理方式が出現してきた(8)(9)。ただし、これらの方式は、グループ単位で鍵が必要になることから管理が複雑であり、またデータ単位でアクセス制御ができないという欠点があった。一方、秘匿検索技術は、鍵や暗号化タグ、暗号化データおよび検索クエリに属性を埋め込むことができるという特徴を有しており、これらの属性を利用することで検索や復号化の権限制御を実現可能である(10)(11)。

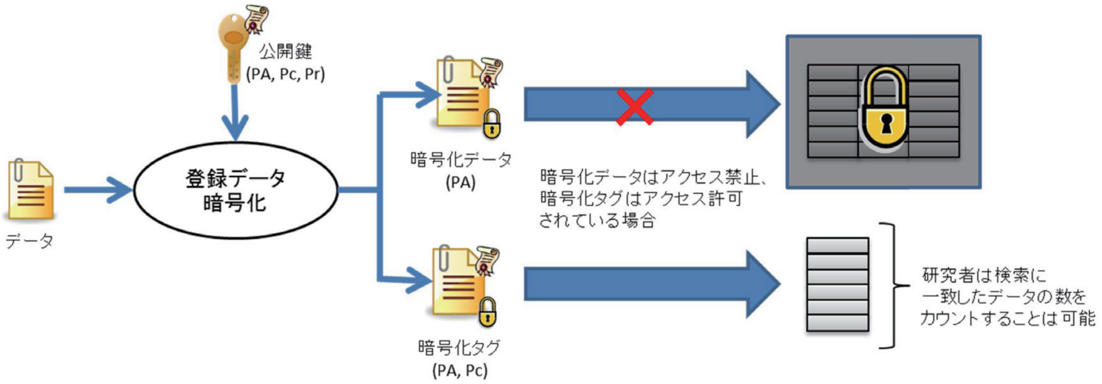
秘匿検索技術では、データを暗号化する時に公開鍵、データ検索時と復号時に秘密鍵を使用する。これらの鍵には複数の属性を埋め込むことができ、埋め込まれた属性は検索時に評価することができる。これらの特徴を生かしてデータ暗号化用の公開鍵にアクセス許可を与えるフラグ、データ検索用の秘密鍵にアクセス権限を示すフラグを埋め込んで利用者に配布し、検索時の演算により許可されたデータにだけヒットする仕組みを構築できる。これにより、個々の利用者が管理すべき鍵は公開鍵と秘密鍵の合計２つに限られ、鍵の管理が簡略化できる。

図４および５に、ゲノム変異データの利用を想定した秘匿検索技術の適用例を示す。ここでは、検索用インデ



特徴
1.　秘密鍵にアクセス権限情報、公開鍵にアクセス許可情報を組み込むことができる
2.　検索クエリは秘密鍵のアクセス権限情報を組み込むことができる
3.　暗号化タグは公開鍵のアクセス許可情報を組み込むことができる
これらの特徴によって、下記のことが実現できる
1.　検索クエリのアクセス権限情報と暗号化タグのアクセス許可情報を比較することでアクセス可否の判断ができる
2.　暗号化タグ作成時にアクセス制御を個人宛て(PA, PB, PC)、職種宛て(Pc, Pr)と分けることで取得できるデータの範囲を変更できる

図４　秘匿検索技術によるアクセス制御例



暗号化データ	暗号化タグ	使用用途
公開許可	公開許可	データを検索し、ヒットしたデータを復元するときに使用される。
公開許可	公開禁止	別のデータで検索し、関連するデータを復元するときに使用される。
公開禁止	公開許可	データを復元しないが、該当データの個数を取得するために使用される。

研究目的に統計情報が二次利用可能になる

図５　二次利用を想定した検索とデータ復号化のアクセス制御の独立性

ックスに相当するデータを暗号化タグ、保管するデータ項目本体を暗号化データと称している。暗号化タグと暗号化データのアクセス許可を個別に設定することで、検索はできるがデータの中身は復号化できないという制御が可能になる。図４では、それぞれの患者ごとに氏名などの個人情報、手術などで得られた組織の病理診断情報およびゲノムから得られた遺伝子診断情報が存在すると想定した場合に、遺伝カウンセラ、主治医およびデータを二次利用する研究者に対してどのようにアクセス制御が行われるかを示している。患者Ａ、Ｂ、Ｃのデータに対して、遺伝カウンセラにはすべての患者の遺伝子診断情報と個人情報、データを二次利用する研究者には全患者の遺伝子診断情報の統計情報のみ利用できるようにし、患者ＡおよびＢの主治医にはすべての情報にアクセスできるがＣの情報にはアクセスできない、というアクセス制御を実現したい場合、鍵管理センターでそれぞれの利用者向けにアクセス許可とアクセス権限が埋め込まれた公開鍵と秘密鍵を生成して配布する。

この場合、鍵に埋め込まれるアクセス属性の要素は、以下のように定義される。

ＰＡ：患者Ａの全データにアクセス
ＰＢ：患者Ｂの全データにアクセス
ＰＣ：患者Ｃの全データにアクセス
Ｐｃ：全患者の個人情報と遺伝子診断情報にアクセス
Ｐｒ：全患者の病理診断情報と遺伝子診断情報の暗号化タグにのみアクセス

患者Ａに配布される公開鍵には、ＰＡ、ＰｃおよびＰｒが、同様に患者Ｂの公開鍵にはＰＢ、ＰｃおよびＰｒ、患者Ｃの公開鍵にはＰＣ、ＰｃおよびＰｒが埋め込まれている。この例では患者ＡとＢの主治医は同一人物であり、検索に用いる秘密鍵にはＰＡとＰＢが埋め込まれている。遺伝カウンセラに配布される秘密鍵にはＰｃのみ、研究者に配布される秘密鍵にはＰｒのみが埋め込まれている。これにより、主治医には担当患者のデータ、遺伝カウンセラにはカウンセリングに必要なデータへのアクセスが実現でき、研究者には図５に示すとおり研究に必要とされる病理診断情報と遺伝子診断情報のヒット数だけが与えられる。

4. むすび

個別化医療において、指静脈認証方式と秘匿検索技術の組み合わせにより医師だけではなく遺伝カウンセラや研究者などに対しても適切なアクセス権限を設定することができ、安全にゲノム変異データを利用する環境を構築できることを示した。

秘匿検索技術においては、ゲノム変異データが発生した時点で暗号化することで、あらゆる状況においても元の平文が保護されることから、医療ネットワークやクラ

＊２　三菱電機株式会社 情報技術総合研究所で開発された暗号化方式
＊３　数値データなどは範囲検索を行う場合、数値を領域に区切って量子化することで、範囲検索をシミュレートする方法がある

ウドでの利用にとどまらず、端末にダウンロードした状態やDVDなどのメディアに複写した場合でも安全性を保つことができる。情報漏洩を防止するために従来は非常に煩雑であったメディアの破棄手続きなども非常に単純化できる。

また、この技術を利用すると、ゲノム変異データを研究に利用する場合、IC（Informed Consent）[*4]に基づくアクセス権限と権限者を暗号化したデータに内蔵できるため、データ二次利用の承認制御が格段に簡単になり、かつ個人情報とゲノム情報を一体化して管理しても安全なため、匿名化の仕組みが不要になることが期待される。今後は、秘匿検索技術を個別化医療のシステムにさらに利用しやすくするため、処理の高速化や鍵属性の管理システムの開発などの改良を進めていく予定である。

最後にこれまで開発を支えてくださった方々にここでお礼申し上げたい。

[*4] インフォームドコンセントの略。ゲノム情報の提供に関して、研究の目的、意義、方法、予測される結果等の説明を受け、自由意思に基づく文書によるゲノム情報提供者の同意のことを言う。インフォームドコンセントを得られていないデータは研究目的での２次利用はできない。

（１） President Obama’s Precision Medicine Initiative

（２） http://www.mofiria.com/-/news/news44

（３） Verison Business, “2009 Data Breach Investigations Supplement Report,” Verison Business, 2009.

（４） 清藤武暢. 四方順司, 高機能暗号を活用した情報漏えい対策「暗号化状態処理技術」の最新動向, 銀行金融研究所機関紙「金融研究」,第33巻第4号, 2014/10

（５） 谷嶋成樹他, 次世代ゲノムブラウザGenomeJackの開発, MSS技報 vol22, 2012.03.30, 13-24

（６） T.Okamoto, K.Takashima, “Hierarchical Predicate Encryption for Inner-Products”, ASIACRYPT 2009, LNCS Vol. 5912, 2009

（７） A. Lewko, T. Okamoto, A. Sahai, K. Takashima, and B. Waters, Fully Secure Functional Encryption: Attribute-Based Encryption and (Hierarchical) Inner Product Encryption, EUROCRYPT 2010, LNCS Vol. 6110, 2010

（８） H.A.Park, J.W.Byun, D.H.Lee, “Secure Index Search for Groups”, TrustBus 2005, LNCS Vol.3592, 2005

（９） P.Wang, H.Wang, J.Pieprzyk, “Keyword Field-Free Conjunctive Keyword Searches on Encrypted Data and Extension for Dynamic Groups”, CANS 2008, LNCS Vol.5339, 2008

（10） 松田, et al. 階層的内積述語暗号を用いたデータセンタシステムの検討, 4F2-3, SCIS2010, 2010

（11） 松田, et al. 検索可能暗号の高速化とWebアプリケーションへの適用方式に関する提案, DS-2, DICOMO2013, 2013.