（報告様式４）

16jm0310005h0003

平成 29 年 5 月 31 日

平 成 28 年 度　委 託 研 究 開 発 成 果 報 告 書

I. 基本情報

事　業　名：（日本語）医療分野国際科学技術共同研究開発推進事業
戦略的国際科学技術協力推進事業（SICP）日本－フィンランド 研究交流
（英　語）International Collaborative Research Program:
Strategic International Research Cooperative Program(SICP)

研究開発課題名：（日本語）個別化医療を実現するプライバシ保護ゲノム情報解析
（英　語）Privacy-preserving genomic data analysis for personalized medicine (PRIVAGEN)

研究開発担当者　（日本語）清水 佳奈
所属 役職 氏名：（英　語）Kana Shimizu

実 施 期 間：平成 28 年　4 月　1 日 ～ 平成 29 年　3 月　31 日

分担研究　（日本語）
開発課題名：（英　語）
研究開発分担者（日本語）
所属 役職 氏名：（英　語）

II. 成果の概要（総括研究報告）

次世代シークエンサーの普及により，個人のゲノム情報の解読と蓄積が急速に進んでいる．しかしながら，個人ゲノムデータは取扱いの難しさから組織間で共有することが難しいため，収集されたデータは目的も立場も異なる組織に点在したまま孤立する場合が多く，潜在的には豊富に存在するはずのデータが十分に活かされていない問題がある．このような状況の改善を目指し，本研究ではデータの中身を秘匿したまま，実用的な速度でゲノム情報を解析する技術の開発を目的とした．本事業期間の前半では，日本側は暗号技術に基づく技術，フィンランド側は差分プライバシ法に基づく技術を開発し，後半では両国が一体となって各々の開発した技術を組み合わせたプライバシ保護ゲノム解析技術の開発を実施した．また，ゲノム情報解析分野と暗号技術分野は学術的に遠い分野であるため，それぞれを専門とする研究者が一堂に集結するコミュニティは依然として稀である．本プロジェクトにおける交流を通じて，暗号技術の研究者とゲノム情報解析の研究者が相互理解を深め，本分野における研究を活性化することも目標とした．以下に本事業で達成された成果のポイントを示す．

（1）研究の成果

- 日本側では，暗号とデータ構造を効果的に組み合わせ，秘匿検索の高度化と効率化を両立する新技術の開発に成功した．従来技術は入力サイズに対して指数的な計算量が必要だったのに対して，本事業で開発した技術は線形の計算量で計算が可能であることを示した．
- 日本側では，上記の新技術を応用して，ゲノム配列検索，医療データなどの文字列検索，機械学習モデルの評価を行うことのできる秘匿計算アルゴリズムを開発した．また，開発したアルゴリズムが従来手法よりも理論的に優れていることを示したことに加え，実データにおいて理論通りの性能が得られることを実証した．
- 日本側では，上記三つのアルゴリズムを実装してインターネット上で公開することにより，技術が社会に広く普及する土台を整備した．
- フィンランド側では，差分プライバシ法を応用したベイズ法に基づく機械学習を行う新技術を開発した．また，さらなる応用技術の開発によりプライバシ保護薬剤感受性予測を行う手法を開発した．
- 日本側とフィンランド側が共同で，暗号と差分プライバシ法を組み合わせて効率的にプライバシ保護機械学習を行うことのできる技術を開発した．
- 上記の研究開発により，本事業期間中に国内学会から 5 件，研究開発者の多くが在籍する産業技術総合研究所より 1 件の受賞を達成した．

（2）人的交流の成果

- 本事業期間中に日本とフィンランドの研究者がのべ 12 回の渡航を実施し，国際ワークショップの開催を含む 5 回の交流イベントを共同開催した．
- 日本側とフィンランド側が共同開催した国際ワークショップ PRIVAGEN2015 には 10 カ国以上から約 80 名の研究者が集結した．プライバシ保護技術開発とゲノム情報解析は学術的に遠い分野であることから，双方の研究を専門とする研究者が一堂に会する機会は非常にまれであったため，多くの参加者から好評を得た．また，本分野に取り組む世界的に重要な拠点から今後の分野を担う若手の研究者が集結したため，両国の研究者にとって人的ネットワークを構築する良い機会になった．

Recent advances in DNA sequencing technology enables to obtain large-scale personal genome data, and it is expected that an efficient analysis on personal genome data leads to a biomedical innovation. However, although the high expectation, it is difficult to utilize the data including personal information because privacy issues hinder data sharing between data holders. In this project, we have studied novel principles necessary for developing privacy-preserving genomic data analysis where both a user and a server can maintain their privacy. To achieve the challenging goal, we studied two different but complementary approaches: cryptography and differential privacy. In the project, the Finnish research group, which has extensive experience in developing statistical machine learning, developed novel methods based on differential privacy that enables to improve drug sensitivity prediction. The Japanese research group, which has strength in cryptography, developed various search methods based on homomorphic encryption. During the project period, two research groups had various discussions by an in-person meeting and a meeting over the network and examined a combined approach of cryptography and differential privacy.

Below, we report more details about the research conducted by the Japanese research group. We first developed the cryptographic protocol that achieves a following model, which we named a recursive oblivious transfer: A sender has a vector v and a chooser has an index i, the chooser obtains v[v[…v[i]…] without knowing any other elements of the vector v while he/she does not leak the index i to the sender. We also analyzed the time and communication complexities of the algorithm and improved the communication complexity from O(N) to O(√N) . Next, we developed an efficient design principle of privacy-preserving search method. There are an efficient data structure/algorithm such as FM-index which run in time linear to a query length. The key idea of the design principle is to combine such a data structure and the recursive oblivious transfer to achieve linear time search without compromising privacy. Based on the design principle, we have developed following three applications. (1) Privacy-preserving search for SNP sequences that enables the user to search the longest prefix match between a query and SNP sequences. (2) Privacy-preserving substring search on biomedical text that enables the user to search substring from a document such as medical records. (3) Privacy-preserving decision tree evaluation that enables the user to obtain a predicted class label without showing his/her feature vector to the server while the server can keep the model and model parameters private against the user. Those applications are implemented and published at the web site. During the project, the Japanese research group published six referred journal articles, and received five awards from domestic academic societies and a president award from AIST.

In addition to those research results, we also engaged in activities that promote an interdisciplinary research among various related fields such as bioinformatics, genome ethics, machine learning, data-mining and cryptography. For this purpose, the Finnish group and the Japanese group jointly organized a workshop PRIVAGEN 2016 which was held as an official satellite workshop of GIW/InCoB 2016, and we had more than 80 participants from more than 10 countries. The Japanese research group also organized sessions for discussing the effect of new Japanese legislation of personal information on genome information analyses at IIBMP 2015 and IIBMP 2016.

III. 成果の外部への発表

（１）学会誌・雑誌等における論文一覧（国内誌　　5 件、国際誌　　6 件）

1. Nuttapong Attrapadung, Goichiro Hanaoka, Shota Yamada, "New Security Proof for the Boneh-Boyen IBE: Tight Reduction in Unbounded Multi-challenge Security", Proc. of 14th International Conference on Information and Communications Security (ICICS 2014). pp176-190.（査読有）
2. Fuyuki Kitagawa, Takahiro Matsuda, Goichiro Hanaoka, Keisuke Tanaka, "Efficient Key Dependent Message Security Amplification against Chosen Ciphertext Attacks", Proc. of 18th International Conference on Information Security and Cryptology (ICISC 2014). pp84-100（査読有）
3. 照屋唯紀, 縫田光司, 清水佳奈, 花岡悟一郎, "ゲノムプライバシ保護を考慮した紛失通信プロトコル", 2015 年暗号と情報セキュリティシンポジウム(SCIS2015)予稿集, 電子情報通信学会（電

子媒体）, 2015 年 1 月（査読無し）

4. Teruya T, Nuida K, Shimizu K, Hanaoka G, On Limitations and Alternatives of Privacy-Preserving Cryptographic Protocols for Genomic Data. In: The 10th International Workshop on Security (IWSEC 2015) Advances in Information and Computer Security Volume 9241 of the series Lecture Notes in Computer Science pp 242-261（査読有）
5. Shimizu K, Nuida K, Arai H, Mitsunari S, Attrapadung N, Hamada M, Tsuda K, Sakuma J, Hirokawa T, Hanaoka G, Asai K, Privacy-preserving similarity search for chemical compound databases. BMC Bioinformatics, 2015:16 (Suppl 18):S6（査読有）
6. N. Attrapadung, S. Yamada. Duality in ABE: Converting Attribute Based Encryption for Dual Predicate and Dual Policy via Computational Encodings. In CT-RSA 2015, pp. 87-105, 2015（査読有）
7. 石巻 優, 清水 佳奈, 縫田 光司, 山名 早人, 完全準同型暗号を用いた高速なゲノム秘匿検索, 2016 年暗号と情報セキュリティシンポジウム(SCIS2016)予稿集, 電子情報通信学会 （電子媒体）, 2016 年 1 月（査読無し）
8. 清水 佳奈, “生命情報科学におけるプライバシ保護検索” 日本シミュレーション学会学会誌 : シミュレーション (小特集: エネルギーシミュレーションとデータ解析), 第 35 巻 第 1 号 p26-31（依頼執筆）
9. Shimizu K, Nuida K, Rätsch G, Efficient Privacy-Preserving String Search and an Application in Genomics. Bioinformatics, (2016) 32 (11): 1652-1661（査読有）
10. 須藤弘貴, 清水佳奈 "LOUDS と準同型暗号による秘匿決定木評価", 第 9 回データ工学と情報マネジメントに関するフォーラム（DEIM2017）予稿集, 情報処理学会（電子媒体）, 2017 年 3 月（査読無し）
11. 清水佳奈, 山本奈津子, 川嶋実苗, 片山俊明, 荻島創一, "改正個人情報保護法でゲノム研究はどう変わるか？ー個人識別符号・要配慮情報としてのゲノムデータ", 2017 年 3 月号 Vol.35 No.4 p600-605（依頼執筆）

（2）学会・シンポジウム等における口頭・ポスター発表

1. Kana Shimizu, Koji Nuida, Yusuke Sakai, Nuttapong Attrapadung, Gunnar Raetsch, Goichiro Hanaoka, "Privacy Preserving Similarity for Biological Data", Biological Data Scienece, Cold Spring Harbor, 2014（海外, 口頭）
2. Kana Shimizu, "Privacy Preserving Search in Bioinformatics ", ISMB 2014, Boston, 2014（海外, 口頭）
3. Nuttapong Attrapadung, Goichiro Hanaoka, Shota Yamada, "New Security Proof for the Boneh-Boyen IBE: Tight Reduction in Unbounded Multi-challenge Security", ICICS 2014, Hong Kong, 2014. （海外, 口頭）
4. Fuyuki Kitagawa, Takahiro Matsuda, Goichiro Hanaoka, Keisuke Tanaka, "Efficient Key Dependent Message Security Amplification against Chosen Ciphertext Attacks", ICISC 2014, Seoul, 2014. （海外, 口頭）
5. 照屋唯紀, 縫田光司, 清水佳奈, 花岡悟一郎, "ゲノムプライバシ保護を考慮した紛失通信プロトコル", SCIS 2015, 福岡,2015（国内, 口頭）

6. 縫田光司，照屋唯紀，花岡梧一郎，松田隆宏，清水佳奈，"双方向の情報を秘匿可能な効率的化合物データベース検索プロトコル"，ゲノムビッグデータによるゲームチェンジ―新しい創薬・ヘルスケアへの息吹―，東京，2014（国内，口頭）
7. Shimizu K, Nuida K, Rätsch G, Efficient Privacy-Preserving String Search and an Application in Genomics. High Throughput Sequencing Algorithms & Applications (HitSeq 2015), A SIG of ISMB/ECCB 2015, Dublin Irelands, July 11, 2015（海外，口頭）
8. Kana Shimizu, Koji Nuida, Hiromi Arai, Shigeo Mitsunari, Nuttapong Attrapadung, Michiaki Hamada, Koji Tsuda, Takatsugu Hirokawa, Jun Sakuma, Goichiro Hanaoka, Kiyoshi Asai, Privacy-Preserving Search for Chemical Compound Databases, GIW/InCoB 2015 (Joint 26th Genome Informatics Workshop and 14th International Conference on Bioinformatics), Tokyo, Japan, September 9, 2015（国内，口頭）
9. N. Attrapadung, S. Yamada. Duality in ABE: Converting Attribute Based Encryption for Dual Predicate and Dual Policy via Computational Encodings. RSA Conference Cryptographers' Track 2015 (CT-RSA 2015), San Fransisco, April（海外，口頭）
10. Teruya T, Nuida K, Shimizu K, Hanaoka G, On Limitations and Alternatives of Privacy-Preserving Cryptographic Protocols for Genomic Data.,The 10th International Workshop on Security (IWSEC 2015), Nara, Japan, August 28, 2015（国内，口頭）
11. 清水佳奈, “秘匿ゲノム検索”, IPSJ-ONE (情報処理学会全国大会), 東京, 2016 年 3 月 12 日（国内，口頭）
12. 清水佳奈, 縫田光司, Gunnar Rätsch, “Efficient Privacy-Preserving Genomic Sequence Search”, 生命医薬情報学連合大会 2015, 2015 年 10 月 30 日，京都（国内，口頭）
13. 石巻 優, 清水 佳奈, 縫田 光司, 山名 早人, 完全準同型暗号を用いた高速なゲノム秘匿検索, 2016 年暗号と情報セキュリティシンポジウム(SCIS2016), 熊本市, 2016 年 1 月 20 日（国内，口頭）
14. 須藤弘貴, 神保元脩, 縫田光司, 清水佳奈, “Secure String Pattern Match Based on Wavelet Matrix”, 第 5 回生命医薬情報学連合大会, 東京, 2016 年 9 月 30 日（国内，口頭）
15. 神保元脩, 須藤弘貴, 清水佳奈, “Towards Secure Matchmaker Exchange, Secure Similarity Evaluation Using Minhash.”, 第 5 回生命医薬情報学連合大会, 東京, 2016 年 9 月 30 日（国内，ポスター）
16. 須藤弘貴, 清水佳奈, ” LOUDS と準同型暗号による秘匿決定木評価”, 第 9 回データ工学と情報マネジメントに関するフォーラム（DEIM2017），飛騨高山，2017 年 3 月 6 日（国内，口頭）
17. Masanobu Jimbo, Hiroki Sudo, Waseda, Koji Nuida, and Kana Shimizu, An alphabet-friendly privacy-preserving string search, 16th Workshop on Algorithms in Bioinformatics (WABI 2016), August 23, 2016, Aarhus（海外，ポスター）
18. Kana Shimizu, Privacy-preserving genome sequence search, 2016 International Workshop on Spatial and Temporal Modeling from Statistical, Machine Learning and Engineering perspectives (STM2016), July 23, 2016, Tokyo（国内，口頭）

（3）「国民との科学・技術対話社会」に対する取り組み

（4）特許出願