

令和 5 年 6 月 12 日現在

機関番号：32657

研究種目：基盤研究(B)（一般）

研究期間：2019～2021

課題番号：19H04210

研究課題名（和文）ナノポアシーケンサーを用いたRNA二次構造決定法の開発

研究課題名（英文）RNA secondary structure prediction using nanopore sequencers

研究代表者

佐藤 健吾（Sato, Kengo）

東京電機大学・システムデザイン工学部・教授

研究者番号：20365472

交付決定額（研究期間全体）：（直接経費） 10,200,000円

研究成果の概要（和文）：RNA二次構造特異的に化学修飾を引き起こす化合物でRNA配列を処理し、ナノポアシーケンサーでその化学修飾を直接読み取ることによって二次構造プロファイルを計測する方法を確立した。深層学習によるRNA二次構造予測法MXfold2を開発し、世界最高精度を達成した。さらに、RNA二次構造特異的な化学修飾の反応度を考慮し、これになるべく適合するRNA二次構造を予測する手法を実装した。

研究成果の学術的意義や社会的意義

RNAの構造と機能の網羅的な相関解析に期待が集まっている。ここでの基盤技術であるRNA二次構造決定法の多くは、RNAの構造に大きな影響を与えるRNA修飾の存在を無視しており、特に配列長が長く塩基修飾が含まれているRNA配列に関して未だに十分な予測精度とは言えない。エピトランスクリプトームを意識したRNAの構造と機能の網羅的な相関解析へ向けて、RNA修飾を考慮した二次構造予測を実現し、予測精度を改善することが重要な課題となっている。本研究の成果はこの課題を克服するための基盤となる技術である。

研究成果の概要（英文）：We established a method to measure secondary structure profiles by treating RNA sequences with compounds that induce chemical modifications specific to RNA secondary structures and directly reading the chemical modifications with a nanopore sequencer. We developed MXfold2, a deep learning method for RNA secondary structure prediction, and achieved the world's highest accuracy. Furthermore, we implemented a method to predict the RNA secondary structure that fits the reactivity of the chemical modification specific to the RNA secondary structure as much as possible.

研究分野：バイオインフォマティクス

キーワード：バイオインフォマティクス RNA二次構造 ナノポアシーケンサー RNA修飾 深層学習

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

非コード RNA の配列情報解析においては、配列のみでなく構造を考慮する必要がある。タンパク質コード領域は、コドン使用頻度やフレームシフトを起こさないといった選択圧があるために、配列自体が高度に保存される傾向にある。それに対し、非コード RNA の多くは同じ機能を持つもの同士であっても配列相同性はあまり高くない一方で、立体構造を形成することによって機能を発揮することから、非コード RNA の機能と構造の間には強い相関があると考えられている。

実験的に RNA 立体構造を決定する手法の多くは高コストかつ低スループットであるため、RNA 分子内に形成される塩基対の集合として規定される RNA 二次構造を計算機で予測する手法が広く用いられてきた。近年の RNA 二次構造予測技術の進歩により、短い配列に関しては高精度の予測が可能となったものの、長い RNA 配列からの二次構造予測には精度の限界があり、配列長が 500 塩基を超えたあたりから予測精度が著しく悪化する傾向にある。

次世代シーケンサー (NGS) の普及により、大量の DNA 配列を高速かつ低コストで解読することが可能となり、RNA 二次構造に関しても NGS を利用した解析手法がいくつか開発されている。RNA から DNA への逆転写伸長反応を阻害する化学修飾をループ部位の RNA 塩基のみに付加後に逆転写し、NGS で cDNA 配列を読むことによって、塩基対を形成しにくい部位のシグナル (二次構造プロファイル) を得ることができる。得られた二次構造プロファイルを手がかりに計算機で RNA 二次構造を推定することによって、予測精度の向上を実現している。

近年、第 3 世代シーケンサーの一つであるナノポアシーケンサーが注目を集めている。ナノポアシーケンサーは DNA の伸長反応に依らないため、長い配列を一気に解読してロングリードを得ることが可能という特長がある。また、塩基相補性に依らずナノポアを通過する分子を直接観測するため、RNA を逆転写した cDNA ではなく RNA を直接読む direct RNA-seq が可能であるという特長があり、メチル化など DNA/RNA に付加された塩基修飾を検出できる可能性が報告されている。

2. 研究の目的

本研究では、RNA の構造と機能の網羅的な相関解析へ向けて、その基盤となる RNA 二次構造決定のための新しい技術を開発する。具体的には、RNA 二次構造特異的な化学修飾を引き起こす化合物で RNA 配列を処理し、ナノポアシーケンサーでその化学修飾を直接読み取ることによって二次構造プロファイルを計測する方法を確立する。また、得られた二次構造プロファイルを RNA 二次構造予測に利用することによって予測精度の劇的な向上を目指す。

3. 研究の方法

ナノポアシーケンサーを用いた RNA 二次構造決定法の開発を実現するために、以下の各項目を実施する。

課題(1) ナノポアシーケンサーで読み取り可能な二次構造特異的修飾の条件検討

詳細な二次構造情報が解明されている rRNA をナノポアシーケンサーを用いてシーケンシングするために、HeLa 細胞から抽出した Total RNA にポリ A 鎖を付加し、ナノポアシーケンサーを用いて Direct RNA-Seq を行う。次に非塩基対に特異的に化学修飾を施す 1-methyl-7-nitroisatoic anhydride (1M7) を使用して、熱変性させたポリ A 鎖付き Total RNA を化学修飾し、Direct RNA-Seq を行うことで化学修飾によるリードの品質の差異を調べる。また、HeLa 細胞に直接 1M7 を 3 条件の濃度で添加し、Total RNA にポリ A 鎖を付加した後 Direct RNA-Seq を行う。Tombo (Stoiber et al. 2016) を用いて、得られた化学修飾データとネガティブコントロールのデータを比較することで化学修飾部位の検出を行う。

課題(2) 様々な修飾塩基に対応した direct RNA-seq 用ベースコールアルゴリズムの開発

修飾塩基に対応した direct RNA-seq 用ベースコールアルゴリズムを開発し、ベースコール精度の向上を目指す。時系列モデルには再帰的ニューラルネットワークを採用する。ONT 社が中心となって公開しているヒト培養細胞のトランスクリプトームの direct RNA-seq データを利用し、既知の修飾塩基を教師データとしてモデルを学習する。

課題(3) 二次構造特異的な化学修飾の検出アルゴリズムの開発

同一のサンプルを 3 種類の条件、すなわち SHAPE 試薬を入れた条件 (SHAPE+)、SHAPE 試薬を入れない条件 (SHAPE-)、熱変性によって二次構造を壊してから SHAPE 試薬を入れた条件 (DC) でシーケンシングする。SHAPE+条件では、塩基対を形成しない塩基に SHAPE 修飾が付加される。SHAPE-条件では、SHAPE 修飾は一ヶ所も付加されない。DC 条件では、配列中に塩基対が形成されないため均等に SHAPE 修飾が付加される。これらの電流波形を参照ゲノム配列にマッピングし、統計的検定によって差異を検出することによって、SHAPE 修飾

の位置を特定する．そして，検出結果を課題 (2) における教師データとして使用することによって，参照ゲノム配列にマッピングすることなく対照条件 (SHAPE-, DC) なしで SHAPE 修飾を検出できるようにベースコールアルゴリズムを改良する．

課題(4) 二次構造プロファイルを利用した RNA 二次構造予測アルゴリズムの開発

従来型のシーケンサーにより得られる二次構造プロファイルは，リファレンスゲノムの各塩基に対して，リードのマッピング結果に基づいて推定される RT-stop や RT-mutate の割合から，二次構造特異的な SHAPE 修飾が付加される傾向(リアクティビティ)を計算する．リアクティビティの値から疑似自由エネルギーを算出し，二次構造予測に反映する(Deigan et al., PNAS, 2009)．一方，本研究において開発するナノポアシーケンサーを用いた二次構造プロファイルの計測では，各リードのベースコール時に SHAPE 修飾を検出する．これにより塩基対を形成しない塩基を直接同定することができるため，二次構造予測や RNA と他分子(タンパク質や低分子化合物)の相互作用予測などに対して，二次構造プロファイルをより簡潔に反映させることが可能となる．

4. 研究成果

各実験条件のシーケンス結果を表 1 に示した mRNA のシーケンスデータのマッピング率が 97.29%であったのに対し，Total RNA のシーケンスデータのマッピング率は 59.09%と低い結果となった．これは mRNA に比べ，Total RNA の平均リード長が短く，マッピングがより困難であったと考えられる．また，熱変性させた RNA に化学修飾を施したサンプルはネガティブコントロールに比べて平均リード長が短く，平均 Q スコアが低い結果となった．これは熱変性させることで RNA 二次構造を壊し，過剰に化学修飾が付加されることでシーケンスが中断されたと考えられる．また，In-cell(+)に関しても 1M7 修飾濃度が高くなる程平均 Q スコアが下がることから，同様の考察が得られる．

表 1 リード品質の比較

| サンプル | 総リード | 平均リード長(b) | 平均Qスコア | マッピング率 |
|------------------|-----------|-----------|--------|--------|
| mRNA | 481,370 | 841.4 | 11.5 | 97.29% |
| In-cell(+) 10 mM | 481,774 | 633.6 | 9.0 | 66.22% |
| In-cell(+) 25 mM | 896,603 | 342.4 | 7.3 | 40.55% |
| In-cell(+) 80 mM | 422,592 | 292.3 | 7.1 | 37.73% |
| In-cell(-) | 1,095,920 | 452.1 | 8.7 | 59.09% |
| DC(-) | 93,639 | 617.9 | 8.0 | 44.30% |
| DC(+) | 61,193 | 274.3 | 7.3 | 28.88% |

各細胞内 RNA 修飾条件での 1M7 修飾検出結果を表 2 に示した．5S rRNA 配列中の非塩基対の割合に比べ，修飾検出位置と非塩基対位置が一致した割合(位置 0 の割合)が低いことが分かった．しかし，修飾検出位置が非塩基対位置から 1~2 離れた結果の割合(位置 0~2 の割合)を正解に含めると修飾条件濃度が高いほど割合が高くなることが分かった．これはシーケンスの際に，ナノポア内に 5 塩基含まれた電流値を計測しているため，実際の修飾位置の前後 2 塩基に影響を与えたと考えられる．

表 2 1M7 修飾の検出

| | 修飾条件 | 位置0 | 位置1 | 位置2 | 負 | 位置0の割合 | 位置0~2の割合 | 非塩基対率 |
|---------|-------|-----|-----|-----|---|--------|----------|-------|
| 5S rRNA | 10 mM | 3 | 1 | 0 | 5 | 33.3% | 44.4% | 19.0% |
| | 25 mM | 1 | 0 | 3 | 4 | 12.5% | 50.0% | |
| | 80 mM | 1 | 1 | 4 | 1 | 14.3% | 85.7% | |

深層学習を用いた RNA 二次構造予測法 MXfold2 を開発した．熱力学モデルと深層学習を組み合わせることによって RNA 二次構造予測の精度を向上させることに成功し，既存手法との比較実験によって世界最高精度であることを確認した．機械学習・深層学習に基づく既存手法のいくつかが訓練データへの過学習による致命的な欠陥があることを示した一方で，本手法は熱力学モデルを組み合わせることで過学習を克服し，高精度だけでなく頑健な RNA 二次構造予測を実現した(図 1)．本手法によって計算される RNA 二次構造の評価値が熱力学安定性と強い相関を持つことが示唆された．また，MXfold2 に対し，1M7 など RNA 二次構造特異的な化学修飾の反応度を考慮し，これになるべく適合する RNA 二次構造を予測する手法を実装した．

MXfold2 とは異なる方法での深層学習による RNA 二次構造予測手法を開発した．この手法は深層学習による RNA 二次構造予測法の既存の方法と類似した手法である．この手法を用いて，訓練データに対する過剰適合に関する大規模な実験を行った．

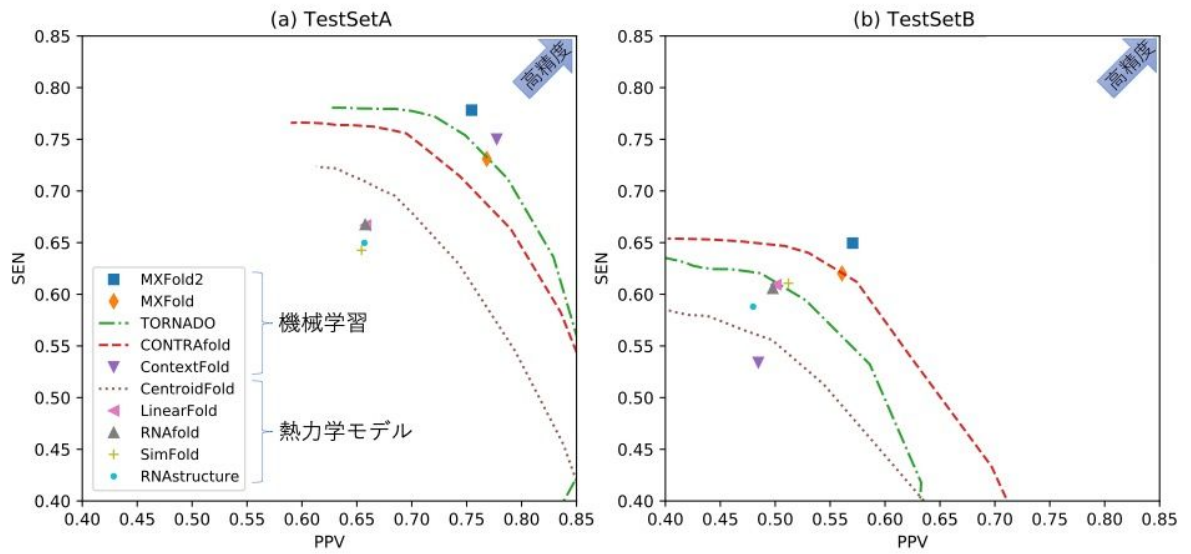


図 1 既存手法との精度比較。PPV は予測した塩基対が正解構造に含まれる割合（正解率）を表し、SEN は正解構造に含まれる塩基対を予測できた割合（網羅率）を表す。

5. 主な発表論文等

〔雑誌論文〕 計8件（うち査読付論文 7件/うち国際共著 0件/うちオープンアクセス 7件）

| | |
|--|---------------------------|
| 1. 著者名 Sato Kengo, Akiyama Manato, Sakakibara Yasubumi | 4. 巻 12 |
| 2. 論文標題 RNA secondary structure prediction using deep learning with thermodynamic integration | 5. 発行年 2021年 |
| 3. 雑誌名 Nature Communications | 6. 最初と最後の頁 941 |
| 掲載論文のDOI（デジタルオブジェクト識別子） 10.1038/s41467-021-21194-4 | 査読の有無 有 |
| オープンアクセス オープンアクセスとしている（また、その予定である） | 国際共著 - |
| 1. 著者名 Taneda Akito, Sato Kengo | 4. 巻 22 |
| 2. 論文標題 A Web Server for Designing Molecular Switches Composed of Two Interacting RNAs | 5. 発行年 2021年 |
| 3. 雑誌名 International Journal of Molecular Sciences | 6. 最初と最後の頁 2720 ~ 2720 |
| 掲載論文のDOI（デジタルオブジェクト識別子） 10.3390/ijms22052720 | 査読の有無 有 |
| オープンアクセス オープンアクセスとしている（また、その予定である） | 国際共著 - |
| 1. 著者名 Jayakumar Vasanthan, Ishii Hiromi, Seki Misato, Kumita Wakako, Inoue Takashi, Hase Sumitaka, Sato Kengo, Okano Hideyuki, Sasaki Erika, Sakakibara Yasubumi | 4. 巻 21 |
| 2. 論文標題 An improved de novo genome assembly of the common marmoset genome yields improved contiguity and increased mapping rates of sequence data | 5. 発行年 2020年 |
| 3. 雑誌名 BMC Genomics | 6. 最初と最後の頁 243 |
| 掲載論文のDOI（デジタルオブジェクト識別子） 10.1186/s12864-020-6657-2 | 査読の有無 有 |
| オープンアクセス オープンアクセスとしている（また、その予定である） | 国際共著 - |
| 1. 著者名 Sato Kengo, Hamada Michiaki | 4. 巻 25 |
| 2. 論文標題 Recent trends in RNA informatics: a review of machine learning and deep learning for RNA secondary structure prediction and RNA drug discovery | 5. 発行年 2023年 |
| 3. 雑誌名 Briefings in Bioinformatics | 6. 最初と最後の頁 bbad186 |
| 掲載論文のDOI（デジタルオブジェクト識別子） 10.1093/bib/bbad186 | 査読の有無 有 |
| オープンアクセス オープンアクセスとしている（また、その予定である） | 国際共著 - |

| | |
|---|-----------------------|
| 1. 著者名 Sato Kengo, Kato Yuki | 4. 巻 23 |
| 2. 論文標題 Prediction of RNA secondary structure including pseudoknots for long sequences | 5. 発行年 2021年 |
| 3. 雑誌名 Briefings in Bioinformatics | 6. 最初と最後の頁 bbab395 |
| 掲載論文のDOI (デジタルオブジェクト識別子) 10.1093/bib/bbab395 | 査読の有無 有 |
| オープンアクセス オープンアクセスとしている (また、その予定である) | 国際共著 - |

| | |
|--|---------------------------|
| 1. 著者名 Kashiwagi Shunya, Sato Kengo, Sakakibara Yasubumi | 4. 巻 11 |
| 2. 論文標題 A Max-Margin Model for Predicting Residue-Base Contacts in Protein-RNA Interactions | 5. 発行年 2021年 |
| 3. 雑誌名 Life | 6. 最初と最後の頁 1135 ~ 1135 |
| 掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/life11111135 | 査読の有無 有 |
| オープンアクセス オープンアクセスとしている (また、その予定である) | 国際共著 - |

| | |
|---|---------------------------|
| 1. 著者名 Akiyama Manato, Sakakibara Yasubumi, Sato Kengo | 4. 巻 13 |
| 2. 論文標題 Direct Inference of Base-Pairing Probabilities with Neural Networks Improves Prediction of RNA Secondary Structures with Pseudoknots | 5. 発行年 2022年 |
| 3. 雑誌名 Genes | 6. 最初と最後の頁 2155 ~ 2155 |
| 掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/genes13112155 | 査読の有無 有 |
| オープンアクセス オープンアクセスとしている (また、その予定である) | 国際共著 - |

| | |
|--|------------------------|
| 1. 著者名 Akiyama Manato, Sato Kengo | 4. 巻 1 |
| 2. 論文標題 RNA Secondary Structure Prediction Based on Energy Models | 5. 発行年 2023年 |
| 3. 雑誌名 RNA Structure Prediction (Methods in Molecular Biology series) | 6. 最初と最後の頁 89 ~ 105 |
| 掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-1-0716-2768-6_6 | 査読の有無 無 |
| オープンアクセス オープンアクセスではない、又はオープンアクセスが困難 | 国際共著 - |

〔学会発表〕 計18件（うち招待講演 2件 / うち国際学会 8件）

| |
|---|
| 1. 発表者名 Yuki Hotta, Yasubumi Sakakibara and Kengo Sato |
| 2. 発表標題 深層強化学習を用いた二次構造に基づくRNA配列の設計 |
| 3. 学会等名 第9回生命医薬情報学連合大会 (IIBMP2020) |
| 4. 発表年 2020年 |

| |
|--|
| 1. 発表者名 Akiyama, M., Sato, K., Sakakibara, Y. |
| 2. 発表標題 A max-margin training of RNA secondary structure prediction integrated with the thermodynamic model |
| 3. 学会等名 Noncoding RNAs: Mechanism, Function and Therapies, Keystone Symposia (国際学会) |
| 4. 発表年 2020年 |

| |
|--|
| 1. 発表者名 Akiyama, M., Sato, K., Sakakibara, Y. |
| 2. 発表標題 A max-margin training of RNA secondary structure prediction integrated with the thermodynamic model |
| 3. 学会等名 RNA Informatics (国際学会) |
| 4. 発表年 2019年 |

| |
|---|
| 1. 発表者名 佐藤健吾 |
| 2. 発表標題 熱力学モデルを統合した深層学習によるRNA二次構造予測 |
| 3. 学会等名 第21回情報科学技術フォーラム (FIT2022) (招待講演) |
| 4. 発表年 2022年 |

| |
|---|
| 1 . 発表者名 Sato, K., Akiyama, M., Sakakibara, Y. |
| 2 . 発表標題 RNA secondary structure prediction using deep learning with thermodynamic integration |
| 3 . 学会等名 RNA meeting 2021 (国際学会) |
| 4 . 発表年 2021年 |

| |
|---|
| 1 . 発表者名 Sato, K., Akiyama, M., Sakakibara, Y. |
| 2 . 発表標題 RNA secondary structure prediction using deep learning with thermodynamic integration |
| 3 . 学会等名 Noncoding RNAs: Biology and Applications (国際学会) |
| 4 . 発表年 2021年 |

| |
|--|
| 1 . 発表者名 Kato, Y., Sato, K., Havgaard, JH., Kawahara, Y. |
| 2 . 発表標題 Deep learning-based prediction of potential RNA G-quadruplexes with D-Quartet |
| 3 . 学会等名 The 29th Intelligent Systems for Molecular Biology and the 20th European Conference on Computational Biology (ISMB/ECCB 2021) (国際学会) |
| 4 . 発表年 2021年 |

| |
|--|
| 1 . 発表者名 Sato, K., Akiyama, M., Sakakibara, Y. |
| 2 . 発表標題 RNA secondary structure prediction using deep learning with thermodynamic integration |
| 3 . 学会等名 The 29th Intelligent Systems for Molecular Biology and the 20th European Conference on Computational Biology (ISMB/ECCB 2021) (国際学会) |
| 4 . 発表年 2021年 |

| |
|--|
| 1. 発表者名 Sato, K., Akiyama, M., Sakakibara, Y. |
| 2. 発表標題 RNA secondary structure prediction using deep learning with thermodynamic integration |
| 3. 学会等名 Noncoding RNA World: From: Mechanism to Therapy (国際学会) |
| 4. 発表年 2021年 |

| |
|---|
| 1. 発表者名 Sato, K., Akiyama, M., Sakakibara, Y. |
| 2. 発表標題 RNA secondary structure prediction using deep learning with thermodynamic integration, |
| 3. 学会等名 第10回生命医薬情報学連合大会, 日本バイオインフォマティクス学会2021年年会 |
| 4. 発表年 2021年 |

| |
|---|
| 1. 発表者名 Kawaguchi, K., Sakakibara, Y., Sato, K. |
| 2. 発表標題 プライバシー保護技術を用いた遺伝子発現差異解析 |
| 3. 学会等名 第10回生命医薬情報学連合大会, 日本バイオインフォマティクス学会2021年年会 |
| 4. 発表年 2021年 |

| |
|---|
| 1. 発表者名 Kengo Sato, Yuki Kato |
| 2. 発表標題 Prediction of RNA secondary structure including pseudoknots for long sequences |
| 3. 学会等名 情報処理学会第68回バイオ情報学研究会 |
| 4. 発表年 2021年 |

| |
|---------------------------------------|
| 1. 発表者名 佐藤健吾, 秋山真那斗, 榎原康文 |
| 2. 発表標題 MXfold2: 深層学習を用いたRNA二次構造予測 |
| 3. 学会等名 第44回日本分子生物学会年会 (招待講演) |
| 4. 発表年 2021年 |

| |
|---|
| 1. 発表者名 Naoki Mikamo, Yasubumi Sakakibara and Kengo Sato |
| 2. 発表標題 Extending a deep learning-based RNA secondary structure prediction algorithm for RNA modifications |
| 3. 学会等名 第23回日本RNA学会年会 |
| 4. 発表年 2022年 |

| |
|---|
| 1. 発表者名 Naoki Mikamo, Yasubumi Sakakibara and Kengo Sato |
| 2. 発表標題 Extending a deep learning-based RNA secondary structure prediction algorithm for RNA modifications |
| 3. 学会等名 第11回生命医薬情報学連合大会 (IIBMP 2022) |
| 4. 発表年 2022年 |

| |
|---|
| 1. 発表者名 Naoki Mikamo, Yasubumi Sakakibara and Kengo Sato |
| 2. 発表標題 Extending a deep learning-based RNA secondary structure prediction algorithm for RNA modifications |
| 3. 学会等名 情報処理学会第72回バイオ情報学研究会 |
| 4. 発表年 2022年 |

| |
|--|
| 1. 発表者名 Sato, K., Akiyama, M., Sakakibara, Y. |
| 2. 発表標題 RNA secondary structure prediction using deep learning with thermodynamic integration |
| 3. 学会等名 RNA Nanotechnology (Gordon Research Conference) (国際学会) |
| 4. 発表年 2023年 |

| |
|------------------------------------|
| 1. 発表者名 川口開登, 榊原康文, 佐藤 健吾 |
| 2. 発表標題 プライバシー保護技術を用いた遺伝子発現差異解析 |
| 3. 学会等名 情報処理学会第73回バイオ情報学研究会 |
| 4. 発表年 2023年 |

〔図書〕 計0件

〔産業財産権〕

〔その他〕

| |
|---|
| MXfold2 Server http://www.dna.bio.keio.ac.jp/mxfold2/ RNA二次構造予測で世界最高精度を達成 https://www.keio.ac.jp/ja/press-releases/2021/2/12/28-78076/ |
|---|

6. 研究組織

| | 氏名 (ローマ字氏名) (研究者番号) | 所属研究機関・部局・職 (機関番号) | 備考 |
|-------|--|---------------------------------------|----|
| 研究分担者 | 加藤 有己 (Kato Yuki) (10511280) | 大阪大学・医学系研究科・助教 (14401) | |

6. 研究組織（つづき）

| | 氏名 (ローマ字氏名) (研究者番号) | 所属研究機関・部局・職 (機関番号) | 備考 |
|-------------------|---|---|----|
| 研究 分 担 者 | 河原 行郎 (Kawahara Yukio) (80542563) | 大阪大学・医学系研究科・教授 (14401) | |

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

| 共同研究相手国 | 相手方研究機関 |
|---------|---------|
| | |