

令和 4 年 6 月 6 日現在

機関番号：32612

研究種目：新学術領域研究（研究領域提案型）

研究期間：2017～2021

課題番号：17H06410

研究課題名（和文）人工知能を用いた化学コミュニケーション空間の多様性と共通性の解明

研究課題名（英文）Deep analysis of chemical communication space using artificial intelligence technology

研究代表者

榊原 康文（Sakakibara, Yasubumi）

慶應義塾大学・理工学部（矢上）・教授

研究者番号：10287427

交付決定額（研究期間全体）：（直接経費） 62,100,000円

研究成果の概要（和文）：本研究の目的は、多種多様な化学コミュニケーションを統一的に表現するモデルの開発である。タンパク質化合物結合予測を網羅的かつ高精度に行うバーチャルスクリーニングシステムである次世代COPICATを開発し、最新のどの既存手法よりも高い精度を達成した。天然化合物を扱うための自己符号化器（NP-VAE）を開発し、巨大分子構造を射影した潜在空間を獲得することに成功した。領域の班員から送られた1,900種類の化合物データを用いて、世界で初めての天然物・巨大分子構造の潜在空間を構築した。機械学習と専門家のもつドメイン知識のフィードバック戦略により、新規PKCリガンド候補を多数発見することに成功した。

研究成果の学術的意義や社会的意義

これらの研究成果は、化学コミュニケーションの理解と制御に学術的に貢献するとともに、医薬品や農薬などの開発にも寄与することが期待される。とくに、世界で初めて構築した天然物・巨大分子構造の潜在空間は本領域でしか成しえない成果である。それに基づいてAIプラットフォームを完成することにより、化合物を介したあらゆる化学コミュニケーションを統一的に理解し、医薬農薬の創薬や共生などの生命現象の解明に資することになる。

研究成果の概要（英文）：The purpose of this research is to develop a model that represents a wide variety of chemical communication in a unified manner. We have developed the next-generation COPICAT, which is a virtual screening system that comprehensively and highly accurately predicts protein-compound interactions, and achieved higher accuracy than the state-of-the-art existing methods. We have developed a variational auto-encoder (NP-VAE) for handling natural compounds and succeeded in acquiring a chemical latent space that encodes natural macromolecular structures. A latent space of natural compounds and macromolecular structures was constructed using 1,900 types of compound data provided from the members of this research project. We succeeded in discovering a large number of new PKC ligand candidates through machine learning and expert domain knowledge feedback strategies.

研究分野：バイオインフォマティクス

キーワード：ケミカルスペース 深層学習 マルチオミックス バーチャルスクリーニング 人工知能

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。

1. 研究開始当初の背景

(1) ビッグデータの到来は生物・生命科学の分野においても顕著である。解読されたゲノムの数 (Genome Online DB) は、1997年時点では48ゲノム、進行中のプロジェクトが302ゲノムであったのに対して、2017年時点では41,488ゲノムが解読され、35,489ゲノムが解読中となっていた。また、創薬のリード化合物の候補となる低分子化合物データベース (PubChem) には2億1895万化合物、細胞内の網羅的遺伝子発現プロファイルを集めたデータベース (NCBI GEO) には175万細胞サンプル、がん細胞に対する抗がん剤の感受性を網羅的に計測したがん細胞株百科事典 (CCLE) には56.1ギガバイトにおよぶ情報がすでに蓄積されている。さらに、次世代シーケンサーなどのハイスループット計測技術の開発により、指数関数的にデータが増加し、更新されている。生命科学におけるこれらのデータの特徴は、基本的にすべてオープンソースである。

(2) ビッグデータを網羅的に解析する手法を開発するための手段の一つが統計的学習手法の利用である。膨大なデータの中から注目する特性を識別・説明する特徴量とルールを発見し、それらを未知データに適用することで意義のある予測を行う。統計的学習手法の中で、最もよく用いられ、多分野で優れた性能を示してきた手法である SVM を用いて、研究代表者の榊原は、タンパク質・化合物相互作用の網羅的予測システム COPICAT を開発し、数千万という膨大な化合物データベース PubChem からリード化合物を探索することを可能にした。コンピュータ予測とウェット検証実験を統合するフィードバック戦略手法を確立し、前立腺がんの原因遺伝子であるアンドロゲン受容体を標的タンパク質として、新規アンタゴニスト T5853872 の発見に成功した。本研究は世界的に注目され、Ultra High-Throughput Virtual Screening という分野のさきがけともなった。

2. 研究の目的

(1) 本研究は、人工知能分野における深層学習を適用して、同種間・微生物叢内・微生物宿主間などの多種多様な化学コミュニケーションを統一的に表現するモデルを開発する。その目的のため、タンパク質化合物結合予測を網羅的かつ高精度に行うバーチャルスクリーニングシステムである次世代 COPICAT を開発、および多様な分子構造をもつ化合物群で構成されるケミカルスペースから射影される潜在空間 (latent space) を深層学習を用いて構築する。

(2) 獲得された潜在空間から化学コミュニケーションにおいて頻出する生物活性リガンドの特異的部分構造の抽出と機能解析を行う。とくに天然化合物の巨大分子構造から学習された潜在空間を、本領域の AI プラットフォームを構築する基盤として活用する。

深層学習手法により化合物ライブラリーの分子構造群を潜在空間に射影する第一の目的は、ケミカルスペースの多様性を数理的な空間で表現をして、それを効果的に可視化することにある。これにより、化合物の構造や機能の分布を情報学的に解析することが可能となり、また視覚的に確認をすることができる。第二の目的は、獲得した潜在空間をもとに、目的の機能に最適化された新規の化合物構造を設計することにある。例えば、潜在空間における抗がん剤分子標的薬の分布が判明すれば、その周辺空間から新たに化合物を生成することで分子標的薬としてより最適化された化合物構造を設計することができる。このような目的のために、深層学習の分野において、化合物の自己符号化器 (JT-VAE) が提案されてきたが、どの手法も低分子化合物を対象としていたため、天然化合物のような巨大分子構造を扱うことができなかった。

(3) 本領域の入江班 (A02 班) との共同研究により、人工知能をもちいた新規 PKC リガンドの探索を行う。また、バイオ医薬品のデザインのための人工知能技術を開発する。

(4) 任意のシュドノット構造を含む RNA 二次構造予測の厳密解法は、その計算量が極めて複雑なクラスである。これまでの多くの RNA 二次構造予測法はシュドノット構造を無視した予測を行う。しかしながら、シュドノット構造は翻訳やスプライシングの制御、リボソームのフレームシフトなどに関与することが知られており、シュドノット構造を考慮した RNA 二次構造予測手法が求められている。mRNA やウイルス RNA のような長い配列に対しても高速かつ高精度でシュドノット構造を含む RNA 二次構造を予測できる手法の開発を目指すことにした。また、タンパク質-RNA 相互作用 (PRI) は多くの生物学的プロセスに必須であり、PRI に関与する配列や構造の側面を理解することは、そのプロセスを解明するために重要である。しかし、タンパク質-RNA 相互作用の複雑な構造を実験的に決定することは高価で時間もかかるため、計算機による PRI の予測方法が開発されてきた。これらの手法の多くは、タンパク質の RNA 結合領域のみ、あるいは RNA のタンパク質結合モチーフのみを予測することに主眼を置いており、PRI の残基-塩基間のコンタクト全体を予測する方法は、まだ十分な精度が得られていない。これらの手法の中には、全てのタンパク質や RNA の配列について利用可能ではない立体構造や相同配列を特定する必要があるものもある。そこで本研究では、配列情報と配列から予測される構造情報のみを用いてタンパク質と RNA の残基-塩基コンタクトを予測する手法の開発を目指すことにした。

3. 研究の方法

(1) タンパク質の特徴、化合物の特徴、複数タイプのインタラクトームデータを統合して、タンパク質-化合物相互作用を予測する畳み込みニューラルネットワークを用いた深層学習ベースの手法“次世代 COPICAT”を開発した(図1)。インタラクトームデータは、タンパク質-タンパク質相互作用と化合物-化合物相互作用から構成されている。タンパク質データはアミノ酸配列をベクトル表現に変換した後に畳み込みニューラルネットワーク(CNN)にかけることで、分子構造データに基づくタンパク質特徴ベクトルを獲得した。

化合物データについては、フィンガープリント法として標準的に用いられる ECFP を分子構造データに基づく化合物特徴ベクトルとした。マルチインタラクトームデータはそれぞれ、グラフの分散表現学習フレームワークである node2vec を適用することで各ノードの特徴ベクトルを取り出した。最終層では、4つの異なるモダリティの特徴ベクトルを同一の潜在空間に射影し、類似度を計算することにより、相互作用の予測値を出力した。類似度の計算として、cos 類似度を拡張した element-wise product と全結合層を組み合わせた手法を開発した。

(2) 天然化合物を扱うための自己符号化器(NP-VAE)を今回新たに開発し、巨大分子構造を射影した潜在空間を獲得することにはじめて成功した。NP-VAEは、深層学習手法の Tree-LSTM や VAE を駆使し、部分構造抽出と学習アルゴリズムを改良することにより開発を達成した。

NP-VAEは前処理、Encoder、Decoderの三段階から構成される(図2)。前処理では化合物を一定の規則に基づいて分解し単純化した木構造に変換する。代表的な官能基や結合に注目して分解する規則を考案し適用したことによって、より詳細な構造上の特徴抽出が可能となった。Encoder部分で Tree LSTM を用いることで潜在変数 Z を獲得した。Decoder部分では、 Z を入力として化合物を部分構造ごとに再構成するアルゴリズムを用いた。さらに、立体構造を加味した化学潜在空間を構築するために、Decoder部分に三次元 ECFP を予測する経路を追加した。

(3) 本領域の入江班が独自に所有する PKC リガンドの結合評価データ(正例: 339個, 負例: 936個)を教師データとして、グラフ畳み込みニューラルネットワークにもとづく結合予測モデルを学習した。この予測モデルを用いて、PubChem に登録された全ての化合物(約9千万化合物)に対して網羅的な探索を行った。予測結果に対して入江班のドメイン知識による検討を行い、明らかに PKC に結合しないと考えられる化合物を負例として予測モデルにフィードバックし、再学習した予測モデルを用いて2ラウンド目の探索を行った(図3)。

バイオ医薬品のデザインのために、人工知能とウェット実験を組み合わせたタンパク質の機能改良アプローチを提案した。このアプローチでは、まず、ウェット実験によって改変対象タンパク質のアミノ酸配列をランダムに変異させた変異体ライブラリーを作成して機能評価を行う。次に、機能評価データを教師データとして、人工知能による機能予測モデルを学習する。この予測モデルをもちいて、アミノ酸配列空間から高機能変異体の網羅的予測を行う。さらに、予測された変異体のウェット実験による機能評価を行い、教師データとしてフィードバックする。このようなサイクルを繰り返すことで、タンパク質の機能を改良していく。

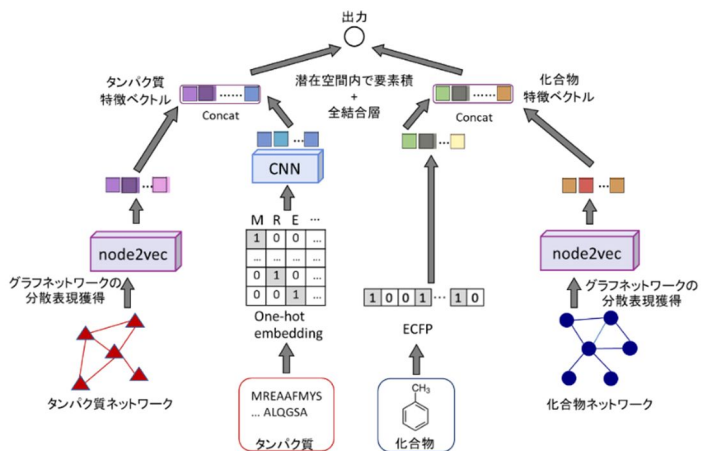


図1: 次世代 COPICAT のアーキテクチャー: タンパク質化合物相互作用予測にマルチオミクスデータを統合した深層学習手法

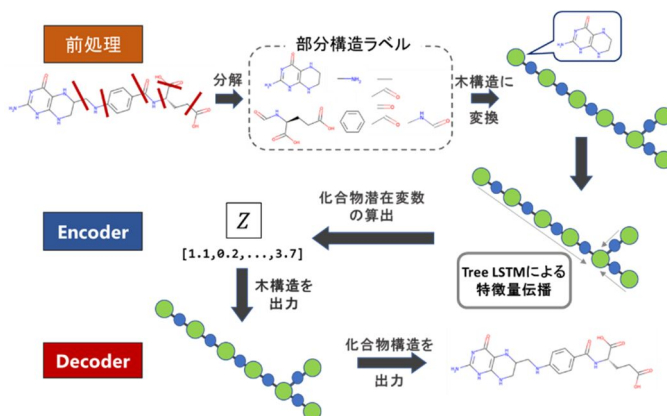


図2: 自己符号化器(NP-VAE)のアーキテクチャー

Encoder部分で Tree LSTM を用いることで潜在変数 Z を獲得した。Decoder部分では、 Z を入力として化合物を部分構造ごとに再構成するアルゴリズムを用いた。さらに、立体構造を加味した化学潜在空間を構築するために、Decoder部分に三次元 ECFP を予測する経路を追加した。

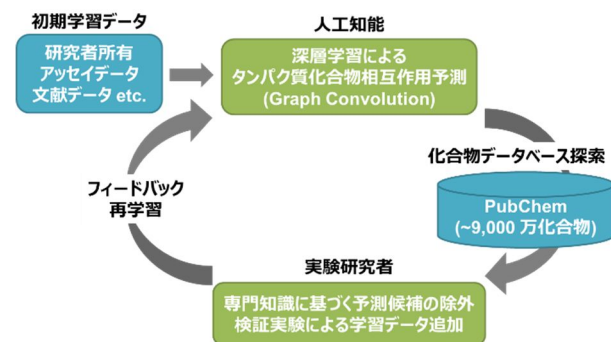


図3: 人工知能と実験研究者によるフィードバック戦略

(4) IPknot の高速化のために、ベースとなる計算モデルに解析手法高速化の点で良い近似を実現する LinearPartition モデルを採用した。LinearPartition モデル自体はシュードノットを考慮した計算は行わないが、これに IPknot による近似解法を組み合わせることによって、配列長に対して線形の計算量でシュードノットを考慮した RNA 二次構造予測を実現した。さらに、あらかじめ予測精度を見積もる指標 pseudo-expected accuracy をシュードノット構造に適用し、これに基づいて配列ごとに最適なパラメータを自動的に選択する方法を開発し、これによって高精度化を実現した。

タンパク質と RNA の残基-塩基コンタクトの予測を、整数計画問題として定式化した。配列の組み合わせや予測される二次構造などの配列ベースの特徴に基づき、スコアリング関数を最大化する残基-塩基間コンタクトマップを予測する。スコアリング関数は、既知の PRI と 3 次元構造から最大マージンのフレームワークを用いて学習される。

4. 研究成果

(1) タンパク質化合物相互作用予測にマルチオミクスデータを統合した深層学習手法を開発し、最新のどの既存手法よりも高い精度を達成することに成功した。本手法をもって次世代 COPICAT の開発が達成されたことを示した。

タンパク質の特徴、化合物の特徴、複数タイプのインタラクトームデータを統合して、タンパク質-化合物相互作用を予測する畳み込みニューラルネットワークを用いた深層学習ベースの手法を開発した (J Cheminform, 2021)。インタラクトームデータは、タンパク質-タンパク質相互作用と化合物-化合物相互作用から構成されている。そして、難易度の異なる 3 つのベンチマークデータセットを用意し、それぞれについて性能評価を行った。データベース STRING および STITCH からマルチインタラクトームであるタンパク質-タンパク質相互作用ネットワークと化合物-化合物相互作用ネットワークを取得し、22,831 のタンパク質-化合物相互作用データの予測を行なった。分子構造データとマルチインタラクトームデータを統合した本手法は、タンパク質-化合物相互作用予測のための最先端の既存手法を上回る精度を示した (表 1)。

表 1 : 次世代 COPICAT と既存手法との性能比較

	Baseline			Unseen compound-test			Hard		
	AUROC	AUPRC	F-value	AUROC	AUPRC	F-value	AUROC	AUPRC	F-value
統合モデル	0.972	0.954	0.900	0.890	0.842	0.727	0.882	0.834	0.714
分子構造モデル	0.956	0.927	0.868	0.869	0.786	0.657	0.851	0.770	0.662
ネットワークモデル	0.947	0.920	0.853	0.831	0.759	0.661	0.780	0.706	0.601
Graph CNN	0.917	0.850	0.794	0.804	0.679	0.637	0.707	0.563	0.427
NeoDTI	0.956	0.905	0.872	0.823	0.773	0.621	0.790	0.715	0.297
SVM	0.805	0.651	0.743	0.765	0.603	0.689	0.652	0.500	0.481
Random Forest	0.873	0.767	0.837	0.770	0.635	0.697	0.605	0.452	0.364

(2) 天然化合物を扱うための自己符号化器 (NP-VAE) を新たに開発し、巨大分子構造を射影した潜在空間を獲得することに成功した (Pacifichem 2021)。NP-VAE の精度は、従来の深層学習手法よりも高かった。先行研究 JTVAE との精度比較のため、最大原子数が 38 以下の低分子だけを含む化合物データベース ZINC の 5,000 化合物を学習させた後、学習データセットの精度を JTVAE と比較した。NP-VAE は巨大分子構造を高い再構成精度で出力することに成功した (表 2)。次に、領域の班員から送られた 1,900 種類の化合物データを用いて、世界で初めての天然物・巨大分子構造の潜在空間を構築した (図 4)。同時に、本領域の化合物群に特異的な分子構造として、1,000 以上の部分分子構造が同定された。抗がん剤の化学療法薬と分子標的薬という観点から化学潜在空間を観察すると、化学療法薬は左側

表 2 : 自己符号化器 (NP-VAE) と既存手法との性能比較

既存モデルと本手法	再構成成功率	最大原子数
GVAE (Kusner <i>et al.</i> , 2017)	7.2%	38
SD-VAE (Dai <i>et al.</i> , 2018)	43.5%	38
JTVAE (Jin <i>et al.</i> , 2018)	76.3%	38
NP-VAE (Ochiai <i>et al.</i> , 2021)	94.4%	551

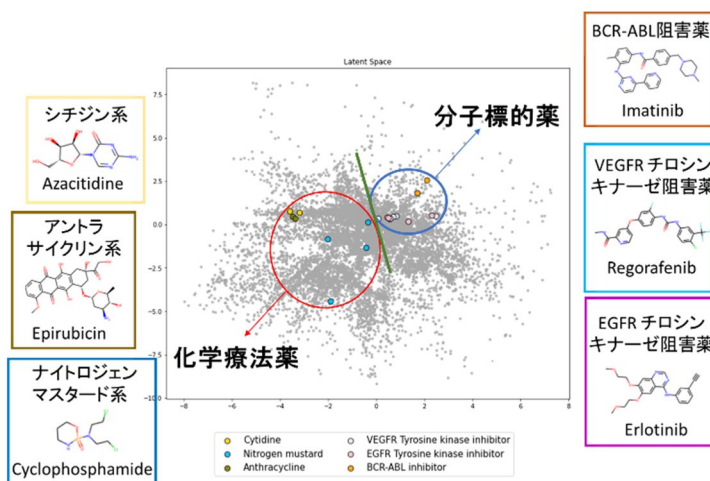


図 4 : 天然化合物の化学潜在空間と特徴的化合物の例

に、分子標的薬は右側に分布しているという興味深い結果が得られた。特に分子標的薬は局所的に分布していることがわかる。

(3) 新規 PKC リガンド探索については、人工知能と専門家のもつドメイン知識のフィードバック戦略により、新規 PKC リガンド候補を多数発見することに成功した(図5)。興味深いことに、これらの候補化合物には PKC リガンドとして新規な骨格構造をもつものが 15 個含まれており、このうち 2 個についてはウェット実験によって PKC への結合が確認された。以上の結果について、Pacifichem 2021 において口頭発表され、国際学術誌 Chem Commun に採択された。

バイオ医薬品デザインのためのタンパク質の機能改良アプローチについては、まず、蛍光タンパク質の蛍光色の改変を対象として研究を行い、本アプローチでタンパク質の機能改変が行えることを実証した(ACS Synth Biol, 2018)。また、機能予測モデルの学習において教師データの組成が配列空間の探索に与える影響を調べた(ACS Calalysis, 2021)。自然言語処理分野における最新の深層学習モデルであるトランスフォーマーにタンパク質の進化的情報を取り入れる手法を開発し、機能予測モデルの性能を向上することに成功した(Brief Bioinform, 2021)。主要なバイオ医薬品である抗体医薬品につ

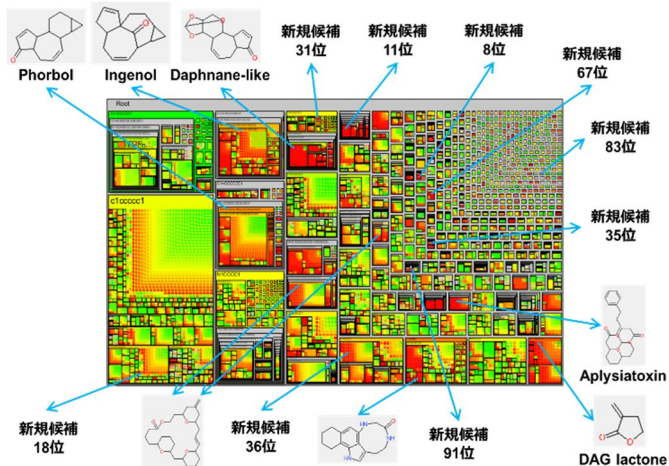


図5：新規 PKC リガンド候補の骨格構造による分類

いて、標的との結合親和性を予測する手法を開発した(Sci Rep, 2020)。これらはバイオ医薬品を含む機能性タンパク質のデザインに人工知能を導入した先駆的な成果であり、日本バイオインフォマティクス学会ポスター賞、生命情報科学若手の会 Best Activator Award、日本蛋白質科学会ポスター賞、国際会議 ISMB/ECCB Best Talk Award を受賞した。

(4) 本研究で開発した IPknot++ は、網羅的なベンチマークにおいて幅広い条件でシュードノットを含まない RNA 二次構造予測と同等の計算速度でありながら(図6)、シュードノットを含む配列に対しても良好な予測精度であることを示した(表3)。

本研究で開発した RNA タンパク質相互作用予測法は、残基-塩基間コンタクト予測において F 値 0.66 を達成した。また、タンパク質の RNA 結合部位予測の既存手法との比較実験によって、結合部位の予測精度が既存手法と同等であることが示唆された。

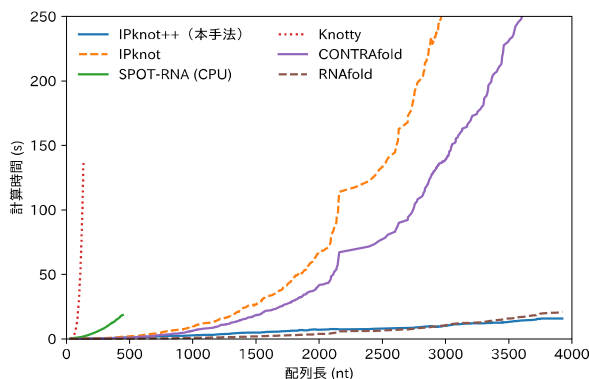


図6：他手法との計算時間の比較

表3：配列長ごとの他手法との F 値の比較

配列長	12-150 塩基		151-500 塩基		501-4381 塩基	
	PK なし	PK あり	PK なし	PK あり	PK なし	PK あり
IPknot++(本手法)	0.681	0.552	0.492	0.482	0.433	0.428
IPknot	0.669	0.500	0.480	0.461	0.212	0.317
Knotty	0.641	0.550	-	-	-	-
SPOT-RNA	0.658	0.621	0.462	0.479	-	-
CONTRAfold	0.682	0.519	0.500	0.479	0.425	0.415
RNAfold	0.668	0.472	0.474	0.442	0.361	0.347

これらの研究成果は、化学コミュニケーションの理解と制御に学術的に貢献するとともに、医薬品や農薬などの開発にも寄与することが期待される。とくに、世界で初めて構築した天然物・巨大分子構造の潜在空間は本領域でしか成しえない成果である。それに基づいて AI プラットフォームを完成することにより、化合物を介したあらゆる化学コミュニケーションを統一的に理解し、医薬農薬の創薬や共生などの生命現象の解明に資することになる。

5. 主な発表論文等

〔雑誌論文〕 計26件（うち査読付論文 25件 / うち国際共著 0件 / うちオープンアクセス 18件）

1. 著者名 Akiyama Manato, Sakakibara Yasubumi	4. 巻 4
2. 論文標題 Informative RNA base embedding for RNA structural alignment and clustering by deep representation learning	5. 発行年 2022年
3. 雑誌名 NAR Genomics and Bioinformatics	6. 最初と最後の頁 lqac012
掲載論文のDOI（デジタルオブジェクト識別子） 10.1093/nargab/lqac012	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 Junpei Maki, Asami Oshimura, Chihiro Tsukano, Ryo Yanagita, Yutaka Saito, Yasubumi Sakakibara, Kazuhiro Irie	4. 巻 -
2. 論文標題 AI and computational chemistry-accelerated development of an alotaketal analogue with conventional PKC selectivity	5. 発行年 2022年
3. 雑誌名 Chemical Communications	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.1039/D2CC01759H	査読の有無 有
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 -
1. 著者名 Yoshimura Yuka, Hamada Akifumi, Augey Yohann, Akiyama Manato, Sakakibara Yasubumi	4. 巻 1
2. 論文標題 Genomic style: yet another deep-learning approach to characterize bacterial genome sequences	5. 発行年 2021年
3. 雑誌名 Bioinformatics Advances	6. 最初と最後の頁 vbab039
掲載論文のDOI（デジタルオブジェクト識別子） 10.1093/bioadv/vbab039	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 Watanabe Narumi, Ohnuki Yuuto, Sakakibara Yasubumi	4. 巻 13
2. 論文標題 Deep learning integration of molecular and interactome data for protein-compound interaction prediction	5. 発行年 2021年
3. 雑誌名 Journal of Cheminformatics	6. 最初と最後の頁 s13321021005133
掲載論文のDOI（デジタルオブジェクト識別子） 10.1186/s13321-021-00513-3	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 Liang Kuo-ching, Sakakibara Yasubumi	4. 巻 22
2. 論文標題 MetaVelvet-DL: a MetaVelvet deep learning extension for de novo metagenome assembly	5. 発行年 2021年
3. 雑誌名 BMC Bioinformatics	6. 最初と最後の頁 s12859020037376
掲載論文のDOI (デジタルオブジェクト識別子) 10.1186/s12859-020-03737-6	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, Nishimura Osamu, Kadota Mitsutaka, Hirose Naoki, Sano Hiromi, Murakawa Yasuhiro, Yamamoto Yumiko, Nakaya Masataka, Tsukiyama Tomoyuki, Seita Yasunari, Nakamura Shinichiro, Kawai Jun, Sasaki Erika, Ema Masatsugu, Kuraku Shigehiro, Kawaji Hideya, Sakakibara Yasubumi	4. 巻 8
2. 論文標題 Chromosomal-scale de novo genome assemblies of Cynomolgus Macaque and Common Marmoset	5. 発行年 2021年
3. 雑誌名 Scientific Data	6. 最初と最後の頁 s41597021009356
掲載論文のDOI (デジタルオブジェクト識別子) 10.1038/s41597-021-00935-6	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 Saito Yutaka, Oikawa Misaki, Sato Takumi, Nakazawa Hikaru, Ito Tomoyuki, Kameda Tomoshi, Tsuda Koji, Umetsu Mitsuo	4. 巻 11
2. 論文標題 Machine-Learning-Guided Library Design Cycle for Directed Evolution of Enzymes: The Effects of Training Data Composition on Sequence Space Exploration	5. 発行年 2021年
3. 雑誌名 ACS Catalysis	6. 最初と最後の頁 14615 ~ 14624
掲載論文のDOI (デジタルオブジェクト識別子) 10.1021/acscatal.1c03753	査読の有無 有
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 -

1. 著者名 Suzuki Godai, Saito Yutaka, Seki Motoaki, Evans-Yamamoto Daniel, Negishi Mikiko, Kakoi Kentaro, Kawai Hiroki, Landry Christian R., Yachie Nozomu, Mitsuyama Toutai	4. 巻 7
2. 論文標題 Machine learning approach for discrimination of genotypes based on bright-field cellular images	5. 発行年 2021年
3. 雑誌名 npj Systems Biology and Applications	6. 最初と最後の頁 s4154002100190w
掲載論文のDOI (デジタルオブジェクト識別子) 10.1038/s41540-021-00190-w	査読の有無 有
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 -

1. 著者名 Yamaguchi Hideki, Saito Yutaka	4. 巻 22
2. 論文標題 Evolving protocols for Transformer-based variant effect prediction on multi-domain proteins	5. 発行年 2021年
3. 雑誌名 Briefings in Bioinformatics	6. 最初と最後の頁 bbab234
掲載論文のDOI (デジタルオブジェクト識別子) 10.1093/bib/bbab234	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 Tajima Naoyuki, Kumagai Toshitaka, Saito Yutaka, Kameda Tomoshi	4. 巻 113
2. 論文標題 Comparative analysis of the relationship between translation efficiency and sequence features of endogenous proteins in multiple organisms	5. 発行年 2021年
3. 雑誌名 Genomics	6. 最初と最後の頁 2675 ~ 2682
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.ygeno.2021.05.037	査読の有無 有
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 -

1. 著者名 Irumagawa Shin, Kobayashi Kaito, Saito Yutaka, Miyata Takeshi, Umetsu Mitsuo, Kameda Tomoshi, Arai Ryoichi	4. 巻 11
2. 論文標題 Rational thermostabilisation of four-helix bundle dimeric de novo proteins	5. 発行年 2021年
3. 雑誌名 Scientific Reports	6. 最初と最後の頁 s41598021869522
掲載論文のDOI (デジタルオブジェクト識別子) 10.1038/s41598-021-86952-2	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 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 s1286402066572
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 Kurumida Yoichi, Saito Yutaka, Kameda Tomoshi	4. 巻 10
2. 論文標題 Predicting antibody affinity changes upon mutations by combining multiple predictors	5. 発行年 2020年
3. 雑誌名 Scientific Reports	6. 最初と最後の頁 s41598020763698
掲載論文のDOI (デジタルオブジェクト識別子) 10.1038/s41598-020-76369-8	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 Lee Jong-Hun, Saito Yutaka, Park Sung-Joon, Nakai Kenta	4. 巻 27
2. 論文標題 Existence and possible roles of independent non-CpG methylation in the mammalian brain	5. 発行年 2020年
3. 雑誌名 DNA Research	6. 最初と最後の頁 dsaa020
掲載論文のDOI (デジタルオブジェクト識別子) 10.1093/dnares/dsaa020	査読の有無 有
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 -

1. 著者名 Saito Yutaka, Kitagawa Wataru, Kumagai Toshitaka, Tajima Naoyuki, Nishimiya Yoshiyuki, Tamano Koichi, Yasutake Yoshiaki, Tamura Tomohiro, Kameda Tomoshi	4. 巻 9
2. 論文標題 Developing a codon optimization method for improved expression of recombinant proteins in actinobacteria	5. 発行年 2019年
3. 雑誌名 Scientific Reports	6. 最初と最後の頁 8338
掲載論文のDOI (デジタルオブジェクト識別子) 10.1038/s41598-019-44500-z	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 Nakato R, Yutaka Saito, et al.	4. 巻 12
2. 論文標題 Comprehensive epigenome characterization reveals diverse transcriptional regulation across human vascular endothelial cells	5. 発行年 2019年
3. 雑誌名 Epigenetics & Chromatin	6. 最初と最後の頁 77
掲載論文のDOI (デジタルオブジェクト識別子) 10.1186/s13072-019-0319-0	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 Hirohara Maya, Saito Yutaka, Koda Yuki, Sato Kengo, Sakakibara Yasubumi	4. 巻 19
2. 論文標題 Convolutional neural network based on SMILES representation of compounds for detecting chemical motif	5. 発行年 2018年
3. 雑誌名 BMC Bioinformatics	6. 最初と最後の頁 526
掲載論文のDOI (デジタルオブジェクト識別子) 10.1186/s12859-018-2523-5	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, Sakakibara Yasubumi	4. 巻 16
2. 論文標題 A max-margin training of RNA secondary structure prediction integrated with the thermodynamic model	5. 発行年 2018年
3. 雑誌名 Journal of Bioinformatics and Computational Biology	6. 最初と最後の頁 1840025 ~ 1840025
掲載論文のDOI (デジタルオブジェクト識別子) 10.1142/S0219720018400255	査読の有無 有
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 -

1. 著者名 Saito Yutaka, Oikawa Misaki, Nakazawa Hikaru, Niide Teppei, Kameda Tomoshi, Tsuda Koji, Umetsu Mitsuo	4. 巻 7
2. 論文標題 Machine-Learning-Guided Mutagenesis for Directed Evolution of Fluorescent Proteins	5. 発行年 2018年
3. 雑誌名 ACS Synthetic Biology	6. 最初と最後の頁 2014 ~ 2022
掲載論文のDOI (デジタルオブジェクト識別子) 10.1021/acssynbio.8b00155	査読の有無 有
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 -

1. 著者名 Aoto Yoshimasa, Okumura Kazuhiro, Hachiya Tsuyoshi, Hase Sumitaka, Wakabayashi Yuichi, Ishikawa Fuyuki, Sakakibara Yasubumi	4. 巻 8
2. 論文標題 Time-Series Analysis of Tumorigenesis in a Murine Skin Carcinogenesis Model	5. 発行年 2018年
3. 雑誌名 Scientific Reports	6. 最初と最後の頁 12994
掲載論文のDOI (デジタルオブジェクト識別子) 10.1038/s41598-018-31349-x	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 Aoki Genta, Sakakibara Yasubumi	4. 巻 34
2. 論文標題 Convolutional neural networks for classification of alignments of non-coding RNA sequences	5. 発行年 2018年
3. 雑誌名 Bioinformatics	6. 最初と最後の頁 i237 ~ i244
掲載論文のDOI (デジタルオブジェクト識別子) 10.1093/bioinformatics/bty228	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 亀田 倫史, 齋藤 裕, 及川 未早来, 梅津 光央, 津田宏治.	4. 巻 21
2. 論文標題 機械学習支援による蛋白質高機能化	5. 発行年 2019年
3. 雑誌名 分子シミュレーション研究会誌「アンサンブル」	6. 最初と最後の頁 34 ~ 38
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
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 -

1. 著者名 Aoto Yoshimasa, Hachiya Tsuyoshi, Okumura Kazuhiro, Hase Sumitaka, Sato Kengo, Wakabayashi Yuichi, Sakakibara Yasubumi	4. 巻 12
2. 論文標題 DEClust: A statistical approach for obtaining differential expression profiles of multiple conditions	5. 発行年 2017年
3. 雑誌名 PLoS One	6. 最初と最後の頁 e0188285
掲載論文のDOI (デジタルオブジェクト識別子) 10.1371/journal.pone.0188285	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 Ito Naofumi, Katoh Kaoru, Kushige Hiroko, Saito Yutaka, Umemoto Terumasa, Matsuzaki Yu, Kiyonari Hiroshi, Kobayashi Daiki, Soga Minami, Era Takumi, Araki Norie, Furuta Yasuhide, Suda Toshio, Kida Yasuyuki, Ohta Kunimasa	4. 巻 8
2. 論文標題 Ribosome Incorporation into Somatic Cells Promotes Lineage Transdifferentiation towards Multipotency	5. 発行年 2018年
3. 雑誌名 Scientific Reports	6. 最初と最後の頁 1634
掲載論文のDOI (デジタルオブジェクト識別子) 10.1038/s41598-018-20057-1	査読の有無 無
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 -

[学会発表] 計20件(うち招待講演 3件/うち国際学会 7件)

1. 発表者名 Liang KC, Sakakibara Y
2. 発表標題 MetaVelvet-DL: a MetaVelvet deep learning extension for de novo metagenomics assembly
3. 学会等名 International Conference on Bioinformatics (InCoB) 2019 (国際学会)
4. 発表年 2019年

1. 発表者名 Uehara M, Kominato M, Hase S, Inoue T, Sasaki E, Toyoda A, Sakakibara Y
2. 発表標題 Development of meta-transcriptome analysis method and its application to meta-transcriptome map of common marmoset
3. 学会等名 Intelligent Systems for Molecular Biology (ISMB) / International Society for Computational Biology (ISCB) 2019 (国際学会)
4. 発表年 2019年

1. 発表者名 榊原康文
2. 発表標題 慶應大学における医療へのA I活用の現状と動向について
3. 学会等名 公益社団法人日本技術士会神奈川県支部 第80回CPD講座（招待講演）
4. 発表年 2019年

1. 発表者名 齋藤 裕
2. 発表標題 人工知能と実験の融合による生物工学研究
3. 学会等名 公益社団法人日本技術士会神奈川県支部 第80回CPD講座（招待講演）
4. 発表年 2019年

1. 発表者名 Maya Hirohara, Yutaka Saito, Yuki Koda, Kengo Sato, Yasubumi Sakakibara
2. 発表標題 Convolutional neural network based on SMILES representation of compounds for detecting chemical motif
3. 学会等名 The 29th International Conference on Genome Informatics (GIW 2018) (国際学会)
4. 発表年 2018年

1. 発表者名 Genta Aoki and Yasubumi Sakakibara
2. 発表標題 Convolutional neural networks for classification of alignments of non-coding RNA sequences
3. 学会等名 The 26th International conference on Intelligent Systems for Molecular Biology (ISMB2018) (国際学会)
4. 発表年 2018年

1. 発表者名 Mika Uehara, Minoru Kominato, Sumitaka Hase, Takashi Inoue, Erika Sasaki, Yasubumi Sakakibara
2. 発表標題 Development of meta-transcriptome analysis method and its application to meta-transcriptome map of common marmoset
3. 学会等名 6th World Congress on Targeting Microbiota (国際学会)
4. 発表年 2018年

1. 発表者名 榊原康文
2. 発表標題 Development of meta-transcriptome analysis method and its application to meta-transcriptome map of common marmoset
3. 学会等名 第77回日本癌学会学術総会
4. 発表年 2018年

1. 発表者名 佐藤健吾
2. 発表標題 機械学習を用いたRNA二次構造予測
3. 学会等名 日本バイオインフォマティクス学会九州地域部会セミナー (招待講演)
4. 発表年 2018年

1. 発表者名 Yutaka Saito, Misaki Oikawa, Hikaru Nakazawa, Teppei Niide, Tomoshi Kameda, Koji Tsuda, Mitsuo Umetsu.
2. 発表標題 Machine-learning-guided mutagenesis for directed evolution of fluorescent proteins.
3. 学会等名 The 29th International Conference on Genome Informatics (GIW 2018). (国際学会)
4. 発表年 2018年

1. 発表者名 Misaki Oikawa, Yutaka Saito, Tomoshi Kameda, Hikaru Nakazawa, Teppei Niide, Koji Tsuda, Mitsuo Umetsu.
2. 発表標題 Machine-learning-guided mutagenesis platform for desired evolution: in the case of fluorescent protein.
3. 学会等名 The 10th Protein & Antibody Engineering Summit (PEGS Europe 2018). (国際学会)
4. 発表年 2018年

1. 発表者名 田島 直幸, 北川 航, 齋藤 裕, 西宮 佳志, 玉野 孝一, 安武 義晃, 田村 具博, 亀田 倫史.
2. 発表標題 Rhodococcus erythropolisを用いた遺伝子配列改変によるタンパク質発現調節法の開発.
3. 学会等名 日本農芸化学会 2019年度大会.
4. 発表年 2019年

1. 発表者名 田島 直幸, 熊谷 俊高, 齋藤 裕, 亀田 倫史.
2. 発表標題 リボソームプロファイリングデータから見る複数生物の内在性タンパク質の翻訳効率と配列特徴量の関係.
3. 学会等名 第13回 日本ゲノム微生物学会年会.
4. 発表年 2019年

1. 発表者名 Naoyuki Tajima, Wataru Kitagawa, Yutaka Saito, Yoshiyuki Nishimiya, Kouichi Tamano, Yoshiaki Yasutake, Tomohiro Tamura, Tomoshi Kameda.
2. 発表標題 放線菌口ドコッカス属における遺伝子配列改変による発現調節の解析.
3. 学会等名 第7回 生命医薬情報学連合大会 (IIBMP 2018).
4. 発表年 2018年

1. 発表者名 亀田 倫史, 齋藤 裕, 津田 宏治, 梅津 光央.
2. 発表標題 人工知能を用いたタンパク質高機能化.
3. 学会等名 第70回 日本生物工学会大会.
4. 発表年 2018年

1. 発表者名 及川 未早来, 齋藤 裕, 亀田 倫史, 中澤 光, 二井手 哲平, 津田 宏治, 梅津 光央.
2. 発表標題 AIはタンパク質進化を導くか? : 機械学習支援によるGFPのYFP化検証.
3. 学会等名 第70回 日本生物工学会大会.
4. 発表年 2018年

1. 発表者名 及川 未早来, 齋藤 裕, 亀田 倫史, 中澤 光, 二井手 哲平, 津田 宏治, 梅津 光央.
2. 発表標題 機械学習支援によるタンパク質進化工学検証 : GFPからYFPへ.
3. 学会等名 化学工学会 第84年会.
4. 発表年 2019年

1. 発表者名 田島 直幸, 北川 航, 齋藤 裕, 西宮 佳志, 玉野 孝一, 安武 義晃, 田村 具博, 亀田 倫史.
2. 発表標題 放線菌口ドコッカス属における遺伝子配列改変による発現調節手法の開発.
3. 学会等名 第41回 日本分子生物学会年会 (MBSJ 2018).
4. 発表年 2018年

1. 発表者名 榊原康文
2. 発表標題 人工知能を用いた化学コミュニケーション空間の多様性の解明に向けて
3. 学会等名 日本農芸化学会2018年度大会シンポジウム
4. 発表年 2018年

1. 発表者名 齋藤裕, 光山 統泰
2. 発表標題 Cosearge: HiCデータから複数のTADにまたがる遺伝子共局在を検出するアルゴリズム
3. 学会等名 2017年度 生命科学系学会合同年次大会
4. 発表年 2017年

〔図書〕 計2件

1. 著者名 亀田 倫史, 齋藤 裕, 田島 直幸, 西宮 佳志, 玉野 孝一, 北川 航, 安武 義晃, 田村 具博.	4. 発行年 2018年
2. 出版社 シーエムシー出版	5. 総ページ数 240
3. 書名 情報解析に基づく遺伝子配列変化による発現量調節. スマートセルインダストリー -微生物細胞を用いた物質生産の展望-	

1. 著者名 Yutaka Saito.	4. 発行年 2018年
2. 出版社 Elsevier	5. 総ページ数 3284 (354-362)
3. 書名 Comparative Epigenomics. Encyclopedia of Bioinformatics and Computational Biology.	

〔産業財産権〕

〔その他〕

Informative RNA base embedding for RNA structural alignment and clustering by deep representation learning: RNABERT
<https://github.com/mana438/RNABERT.git>

Genomic style: yet another deep-learning approach to characterize bacterial genome sequences
<https://github.com/friendflower94/binning-style>

メタゲノムアセンブリプログラムMetaVelvet-DL
<http://www.dna.bio.keio.ac.jp/metavelvet-dl/>

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	佐藤 健吾 (Sato Kengo) (20365472)	慶應義塾大学・理工学部(矢上)・講師 (32612)	
研究分担者	齋藤 裕 (Saito Yutaka) (60721496)	国立研究開発法人産業技術総合研究所・情報・人間工学領域・主任研究員 (82626)	

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1件

国際研究集会	開催年
The Sixteenth Asia Pacific Bioinformatics Conference (APBC2018)	2018年～2018年

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