

# 人工知能と CRISPR による 口腔細菌叢のデイスバイオーシス検出方法の検討

渡辺 孝康<sup>1,2</sup>, 芝 多佳彦<sup>3</sup>, 中野 善夫<sup>1,2</sup>

## Detection of dysbiosis within oral microbiota by using artificial intelligence and CRISPR

Takayasu Watanabe<sup>1,2</sup>, Takahiko Shiba<sup>3</sup>, Yoshio Nakano<sup>1,2</sup>

### Abstract

Recently, dysbiosis, which is an imbalance in the state of microbiota in terms of bacterial composition, has been focused on as a cause of infectious diseases. However, it is difficult to identify dysbiotic patterns of bacterial composition because the composition originally differs among individuals. On the other hand, artificial intelligence is a growing technology that enables the recognition, identification, and/or elucidation of problems using a large amount of data. In this study, we detected dysbiosis within oral microbiota by using deep learning, which is an artificial intelligence method that uses a pattern recognition algorithm with a neural network. In six patients with periodontitis and four with peri-implantitis, clustered regularly interspaced short palindromic repeat (CRISPR) regions were extracted from the nucleotide sequence data of microbiota and used for deep learning. The repertoire of CRISPR spacers was diverse among the patients and seemed to be almost specific to each patient. Neither hierarchical clustering nor principal component analysis allowed the patients to be clearly distinguished according to their diseases. On the other hand, deep learning was able to distinguish the patients according to their diseases with a high degree of accuracy, suggesting its potential for use as a basic technology in dental sciences for detecting oral dysbiosis.

**Key words** : microbiota, dysbiosis, CRISPR, deep learning

### はじめに

う蝕や歯周炎といった細菌感染症では、コッホの原則として知られる通り、発症の原因をただ一種類の細菌に追究する考え方が主流であっ

た。その一方で、細菌がある環境にて生息する際に、多種類が共生して細菌叢を形成していることが明らかとなって以来、細菌叢における細菌の種類のバランスに生じる乱れ、すなわちデイスバイオーシスに病因を追究する考え方が

<sup>1</sup> 日本大学歯学部 基礎自然科学分野 (化学)

<sup>2</sup> 日本大学歯学部総合歯学研究所 機能形態部門  
〒101-8310 東京都千代田区神田駿河台1-8-13

<sup>3</sup> 東京医科歯科大学大学院歯学総合研究科 歯周病学分野  
〒113-8510 東京都文京区湯島1-5-45  
(受理: 2021年7月8日)

<sup>1</sup> Department of Chemistry, Nihon University School of Dentistry

<sup>2</sup> Division of Functional Morphology, Dental Research Center, Nihon University School of Dentistry  
1-8-13 Kanda-Surugadai, Chiyoda-ku, Tokyo 101-8310, Japan

<sup>3</sup> Department of Periodontology, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University  
1-5-45 Yushima, Bunkyo-ku, Tokyo 113-8510, Japan

近年広まってきている<sup>1)</sup>。これを助ける研究手法として、次世代高速シーケンサーの技術発展が大いに貢献しており、遺伝子情報として塩基配列を決定するのに手間と労力を要したサンガー法に代わって、一度に大量の塩基配列情報を取得できるツールとして、いまやマイクロアレイでさえもとって代われようとしている<sup>2,3)</sup>。しかし、次世代高速シーケンサーで取得した塩基配列データは、その量的優位性ゆえに、細菌叢がディスバイオーシスであるか否かをデータセットから簡易に視認し判別することが困難である。

そこで我々は、ビッグデータを用いて事象の認識・特定・判別を可能とする人工知能技術に着目した。なかでも深層学習は、脳神経が形成するニューラルネットワークを応用してパターン認識のためのアルゴリズムを利用する技術で、画像認識など幅広い分野にて利用されている<sup>4,5)</sup>。本研究では、歯の周囲に生じる歯周炎、ならびに歯科インプラントの周囲に生じるインプラント周囲炎の細菌叢データから、深層学習を用いてディスバイオーシスを検出する方法を検討した。その際、細菌叢データの中でも特に、細菌がウイルスなどの感染を受けた履歴、すなわち獲得免疫を担う遺伝領域としてゲノム内に保持している clustered regularly interspaced short palindromic repeat (CRISPR) を用いた<sup>6-8)</sup>。CRISPR は、等間隔に並んだ同一のリピート配列、およびその間に挟まる形で位置する様々なスペーサー配列から構成される。スペーサー配列は、過去に感染したウイルスなどの遺伝子の一部を保存しておく役割があり、免疫記憶領域であるといわれる。スペーサー配列のバリエーションは、同じ種類の細菌であっても株ごとに多様であるため<sup>9)</sup>、これに対して深層学習を適用すれば、ディスバイオーシスの検出に役立つものと考えた。

## 材料と方法

歯周病に罹患した6名、およびインプラント周囲炎に罹患した4名の患者の炎症病変の細菌叢検体について次世代高速シーケンサーにて解析した細菌DNAの塩基配列情報を利用した(<https://www.ddbj.nig.ac.jp/>; DRA010104)<sup>10)</sup>。解析ソフトウェア CRISPR Recognition Tool (version 1.2)<sup>11)</sup>にて、リピート配列・スペーサー配列のいずれも1個あたり30~40塩基対、かつ3個以上のリピート配列となるような条件を指定して、データからCRISPR領域のみを抽出した。次いで、得られたCRISPRの中からスペーサー配列のみを選別し、解析ソフトウェア CD-HIT (version 4.8.1)<sup>12)</sup>を用いて、互いに類似したスペーサーを整理した。この情報をもとに、患者ごとに保有するスペーサーの種類を調べ、ヒートマップとして図示した。また、階層クラスタリングおよび主成分分析により、保有するスペーサーの種類に基づく患者間の近縁性を調べた。さらに、歯周炎とインプラント周囲炎の識別程度を、解析ソフトウェア R パッケージ h2o (version 3.32.0.1)を用いた深層学習にて調べた。以上の解析には、東京大学医科学研究所ヒトゲノム解析センターのスーパーコンピュータを利用した。

## 結果

歯周炎およびインプラント周囲炎の患者計10名から、合計で2,525個のスペーサー配列が抽出された。スペーサー数は最小の患者で41個、最大の患者で615個であった。2,525個のスペーサーにおいて、互いに類似したスペーサーを調べたところ、2,052種類に整理された。その保有状況をヒートマップにて図示すると、1,934種類のスペーサーは1名の患者のみに認められ、残りのうち115種類は2人、3種類は3人

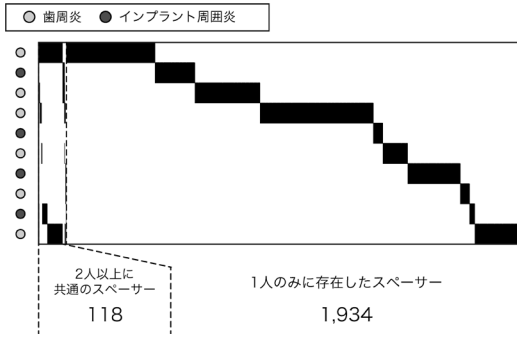


図 1. 患者毎のスペーサー分布

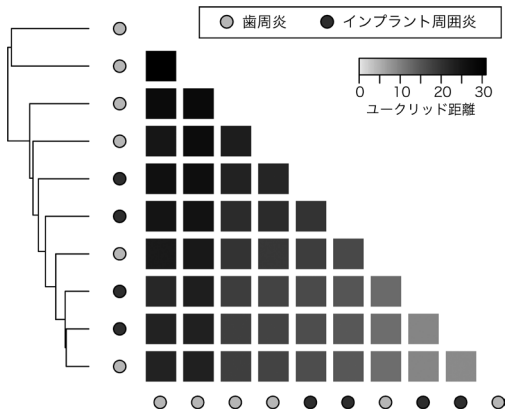


図 2. 階層クラスタリングによる患者間の類似性

の患者に共通して認められた (図 1)。

患者間の類似性に関する解析では、階層クラスタリングにおいて、歯周炎の患者とインプラント周囲炎の患者は明確なクラスターとして分割されず、両者は樹形図において混在していた (図 2)。主成分分析でも同様に、個々の患者を表すプロットは歯周炎とインプラント周囲炎とで明確なクラスターを形成せず、両者が座標平面上で混在する分布となった (図 3)。

ここで、深層学習をスペーサーデータに対して適用し、歯周炎とインプラント周囲炎の識別程度を調べた。その際、患者 1 名のデータだけを除外して学習させ、除外データの疾患を予測させるアルゴリズムを、除外患者を変更しながら繰り返した。その結果、Tanh, Rectifier, Maxout

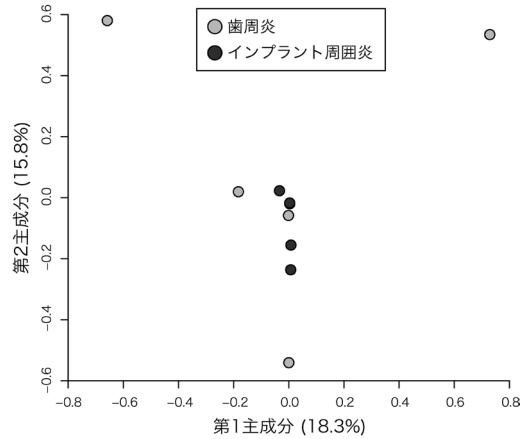


図 3. 主成分分析による患者間の類似性

表 1. 深層学習における活性化関数毎の正解率

活性化関数	Dropoutなし	Dropoutあり
Tanh	80.0	60.0
Rectifier	80.0	70.0
Maxout	90.0	70.0

をベースとした 6 種類の活性化関数のうち、Tanh に dropout を付加した活性化関数において最も低い正解率 60.0% となった。一方、Maxout に dropout を付加しなかった活性化関数において最大の正解率 90.0% となった (表 1)。

## 考察

ゲノム解読されている細菌のうち、CRISPR は 5 割の真正細菌、9 割の古細菌が保有することで知られ<sup>7,13)</sup>、例えば歯周病原細菌 *Porphyromonas gingivalis* では、ゲノム内に最大で 4ヶ所の CRISPR を有し、1ヶ所の CRISPR あたりのスペーサー数は多い場合で 100 個を超えることもある<sup>9)</sup>。本研究ではまず、スペーサーの種類を整理した段階で、2,052 種類のうち 94.2% にも上る 1,934 種類のスペーサーが 1 人の患者のみから抽出された。このことから、10 名の患者におけるスペーサーのバリエー

ションはやはり患者ごとに特有であることが示唆された。すなわち、本研究は解析対象を歯周炎・インプラント周囲炎という2種類の疾患に限定したにもかかわらず、疾患の種類に関係なく、患者ごとに特有なスパーサー保有パターンを示したものといえる。

次いで、患者間におけるスパーサーの種類類似性を調べる際、歯周炎には歯周炎特有の、インプラント周囲炎にはインプラント周囲炎特有のスパーサー保有パターンが、何らかの形で現れると期待された。しかし予想に反して、階層クラスタリングでは、歯周炎・インプラント周囲炎の患者が混在する樹形図となった。階層クラスタリングは近縁の患者同士を階層的に整理する手法で、特定のクラスターが認められる場合、樹形図でクラスター間の枝長が長くなる<sup>9)</sup>が、本研究ではそのような結果は得られなかった。このような、疾患にかかわらず患者が混在するという結果は、階層クラスタリングとは異なるアルゴリズムを用いる主成分分析でも同様であった。主成分分析は、1,934種類のスパーサーの保有状況という多次元データを、視覚的に理解が容易な2次元に縮約するための手法で、本研究では第1主成分には本来の分散の18.3%が、第2主成分には15.8%が縮約された。主成分分析の座標平面上では、プロットが近いほど患者間のスパーサー保有パターンの類似性が高いことを示すため、例えば歯周炎の患者のプロットとインプラント周囲炎患者のプロットが離れた位置にそれぞれ集積していれば、疾患ごとに明確なクラスターが形成されたことを示す。しかしながら、本研究ではそのような明確なクラスターは観察されなかった。

これまでの研究では、細菌叢由来のDNAならびにRNAデータを解析すると、歯周炎とインプラント周囲炎とで異なる特徴が見出されることが報告されているが<sup>14-16)</sup>、以上の結果をまとめると、本研究では解析対象をCRISPRに

絞り込んだことによって、却って疾患ごとの特徴を見出しにくいデータセットになったものと考えられる。また、深層学習を適用せずに、従来多用されてきた階層クラスタリングや主成分分析のような解析手法を用いるだけでは、患者を疾患ごとに識別することが困難であるといえる。

一方、深層学習を適用すると、主要なパラメーターの一つである活性化関数によって正解率に60.0%から90.0%の開きがあるものの、総じて高い正解率のもとで患者を歯周炎・インプラント周囲炎に識別することがわかった。深層学習では、視覚的な識別が困難なデータパターンを認識・識別することができるため、本研究のように患者ごとの特異性が高く疾患ごとの特徴を見出しにくいスパーサー情報というデータであっても、精度よく疾患の識別ができることが示された。CRISPRのみに絞り込んだデータセットは、前述の通り従来の解析手法では疾患ごとの特徴を見出しにくかったにもかかわらず、CRISPRデータと深層学習を用いることで疾患の高精度な識別を成し得たといえる。

## まとめ

細菌叢に由来するDNAデータの中でも、CRISPRにおけるスパーサー情報は患者ごとの特異性が高いとみられ、歯周炎・インプラント周囲炎の特徴を見出しにくいデータセットであった。このことは特に、従来用いられてきた階層クラスタリングや主成分分析といった解析手法の適用下では顕著であった。その一方で、深層学習を用いることによって、スパーサー情報に基づいて高精度で疾患を識別し得ることが示され、細菌叢におけるディスバイオシス検出の一手法として今後、対象患者数を充実させつつ適用方法や精度向上などを探究する必要があると考えられる。また、CRISPRにはスパー

サーだけでなくリポート配列にも細菌の種類ごとの多様性があることが知られており<sup>17)</sup>, これを対象とした深層学習の適用にも検討の余地がある。本研究は, 深層学習という人工知能の一手法を, 細菌叢に由来するビッグデータに適用した研究であり, 人工知能とビッグデータを歯科医療に応用する上での基礎的知見として将来的な応用が期待されるであろう。

### 謝辞

本論文に記載した研究は, 令和元年度および令和2年度日本大学歯学部佐藤奨学金研究費, 公益財団法人富徳会研究者助成金, および公益財団法人ロッセ財団奨励研究助成の援助を受けた。

### 【参考文献】

- 1) Rosier BT, Marsh PD, Mira A (2018) Resilience of the oral microbiota in health: mechanisms that prevent dysbiosis. *J Dent Res* 97, 371-380.
- 2) Loman NJ, Constantinidou C, Chan JZ, Halachev M, Sergeant M, Penn CW, Robinson ER, Pallen MJ (2012) High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nat Rev Microbiol* 10, 599-606.
- 3) 渡辺孝康 (2018) ヒトの口腔内菌叢解析の特徴と応用事例. *ペット栄養学会誌*, 21, 154-157.
- 4) Bolón-Canedo V, Remeseiro B (2020) Feature selection in image analysis: a survey. *Artif Intell Rev* 53, 2905-2931.
- 5) Tsuei HJ, Tsai WH, Pan FT, Tzeng GH (2020) Improving search engine optimization (SEO) by using hybrid modified MCDM models. *Artif Intell Rev* 53, 1-16.
- 6) Marraffini LA, Sontheimer EJ (2010) CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat Rev Genet* 11, 181-190.
- 7) 渡辺孝康, 中川一路, 丸山史人 (2013) 原核生物の新規な獲得免疫機構 CRISPR/Cas システム. *化学と生物*, 51, 440-443.
- 8) 渡辺孝康, 村瀬一典, 中川一路, 丸山史人 (2015) 原核生物の獲得免疫に見られる新機構から紐解かれてきたゲノム進化. *化学療法の領域*, 31, 118 (2240)-128(2250).
- 9) Watanabe T, Nozawa T, Aikawa C, Amano A, Maruyama F, Nakagawa I (2013) CRISPR regulation of intraspecies diversification by limiting IS transposition and intercellular recombination. *Genome Biol Evol* 5, 1099-1114.
- 10) Komatsu K, Shiba T, Takeuchi Y, Watanabe T, Koyanagi T, Nemoto T, Shimogishi M, Shibasaki M, Katagiri S, Kasugai S, Iwata T (2020) Discriminating microbial community structure between peri-implantitis and periodontitis with integrated metagenomic, metatranscriptomic, and network analysis. *Front Cell Infect Microbiol* 10, 596490.
- 11) Bland C, Ramsey TL, Sabree F, Lowe M, Brown K, Kyrpides NC, Hugenholtz P (2007) CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* 8, 209.
- 12) Fu L, Niu B, Zhu Z, Wu S, Li W (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150-3152.

- 13) Bhaya D, Davison M, Barrangou R (2011) CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu Rev Genet* 45, 273-297.
- 14) Maruyama N, Maruyama F, Takeuchi Y, Aikawa C, Izumi Y, Nakagawa I (2014) Intraindividual variation in core microbiota in peri-implantitis and periodontitis. *Sci Rep* 4, 6602.
- 15) Shiba T, Watanabe T, Kachi H, Koyanagi T, Maruyama N, Murase K, Takeuchi Y, Maruyama F, Izumi Y, Nakagawa I (2016) Distinct interacting core taxa in co-occurrence networks enable discrimination of polymicrobial oral diseases with similar symptoms. *Sci Rep* 6, 30997.
- 16) 芝多佳彦, 竹内康雄, 渡辺孝康, 小柳達郎, 丸山緑子, 加地博一, 村瀬一典, 丸山史人, 中川一路, 和泉雄一 (2017) インプラント周囲炎と歯周炎の細菌学的な違いを科学する. *歯界展望*, 129, 823-826.
- 17) Sorek R, Lawrence CM, Wiedenheft B (2013) CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annu Rev Biochem* 82, 237-266.