

令和 5 年 6 月 20 日現在

機関番号：15101

研究種目：基盤研究(C)（一般）

研究期間：2020～2022

課題番号：20K07313

研究課題名（和文）臨床自然言語処理と網羅的遺伝子解析による新たな遺伝病診断システムの開発

研究課題名（英文）Development of a new genetic testing by clinical natural language processing and comprehensive gene analysis

研究代表者

難波 栄二（NANBA, Eiji）

鳥取大学・研究推進機構・教授

研究者番号：40237631

交付決定額（研究期間全体）：（直接経費） 3,300,000円

研究成果の概要（和文）：本研究は臨床自然言語処理（CNLP）と次世代シーケンサー（NGS）を用いた自動診断アルゴリズム（システム）の医療実装を目的とした。神経疾患である歯状核赤核淡蒼球ルイ体萎縮症の模擬カルテを作成し、Human Phenotype Ontologyを用いたCNLP抽出の検討を試みた。2021年に保険収載されたマイクロアレイ染色体検査の診断体制も検討した。さらに、CNLPの参考として発達障害の代表的な疾患である脆弱X症候群（FXS）とダウン症候群（DS）に関するツイートに関する情報の分析を行った。残念ながら、新型コロナウイルス感染症パンデミックのために、十分な研究成果を得ることはできなかった。

研究成果の学術的意義や社会的意義

2023年6月9日「良質かつ適切なゲノム医療を国民が安心して受けられるようにするための施策の総合的かつ計画的な推進に関する法律案」が参議院本会議で可決され成立した。この法律では、ゲノム医療の研究や医療応用を推進するとともに、生命倫理への適切な配慮、ゲノム情報による不当な差別が行われることのないようにすることなどが定められている。本研究では十分な成果は得られなかったが、生命倫理などに配慮したゲノム情報自動診断装置は重要となり、さらに研究を推進することが必要である。

研究成果の概要（英文）：This study aimed to implement a medical algorithm that can automatically diagnose with high accuracy using Clinical Natural Language Processing (CNLP) and Next Generation Sequencing (NGS). We created a simulated patient record for Dentatorubral-Pallido-Luysian Atrophy, a neurological disease, and attempted to examine CNLP extraction using the Human Phenotype Ontology (HPO). We also examined the diagnostic structure of the microarray chromosome examination, which was covered by national medical insurance in 2021. Furthermore, as a reference for CNLP, we conducted an analysis of information related to tweets over the course of 2022 concerning Fragile X Syndrome (FXS) and Down Syndrome (DS), both representative developmental disorders. Unfortunately, due to the COVID-19 pandemic, we were unable to obtain sufficient research results.

研究分野：人類遺伝学

キーワード：臨床自然言語 次世代シーケンサー マイクロアレイ染色体検査 遺伝学的検査 ゲノム医療 人工知能(AI) Social Network Service Twitter

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

稀少難病の分野において次世代シーケンサー (NGS) を用いたゲノム医療が推進されており、申請者らも NGS を用いた研究に取り組み、この分野での研究実績をもつ。NGS を用いた稀少難病の診断ではエクソーム解析が主流であったが、技術開発が進み、全ゲノム解析が利用されるようになってきている。今後は全ゲノム解析を用いることにより、診断率が向上することが期待されている。

しかし、NGS を用いた診断において時間がかかり効率化が難しいのは、遺伝子バリエーション情報と臨床情報とを照合し、膨大な NGS データから疾患の原因となる遺伝子バリエーションを特定して診断を確定させる工程である。現在、日本の国家プロジェクトとして未診断疾患イニシアチブ (IRUD) があり、NGS を用いた解析によって多くの遺伝性疾患の診断実績をあげている。鳥取大学医学部附属病院は IRUD 診断拠点病院となっており、申請者らは診断委員会の中心的な役割を担っている。IRUD における課題は、検体を診断拠点病院から解析拠点へ送付した後、その診断結果が戻ってくるまでに半年以上の時間がかかっていることである。これは、候補の遺伝子バリエーションをコンピュータプログラムで自動的に絞り込んだ後、経験豊富な専門家が臨床情報と照合し、さらに論文などの情報をもとに総合的に判断する工程で多くの時間を要することが背景にある。この課題は、NGS を用いた遺伝性疾患の診断の医療実装を進めるにあたって大きな障壁となり、その対応が必要となる。現在、欧米諸国では臨床情報から自動的に必要な情報を抽出し、NGS データと連携し、迅速かつ効率的な診断を実施する研究が進められている。実際に、新生児集中治療室で 1 日以内に患児の遺伝性疾患を特定する成果も発表されている (Clark M et al. Sci Transl Med 2019)。この報告では NGS 解析の効率化とともに、AI による臨床自然言語処理 (CNLP) により、専門家の検討を経ることなく自動診断する試みが行われている。しかし、日本ではこのような研究は行われておらず、日本語のカルテに適したシステムを作ってゆく必要がある。

申請者らは、NGS 解析において臨床遺伝専門医、認定遺伝カウンセラー、ジェネティックエキスパート、臨床検査技師、バイオインフォマティシャンなどによる検討会 (エキスパートパネル) を毎週開催し、データの検討を行っている。しかし、遺伝子バリエーションの絞り込みは、これまでの経験に頼ることも多く、時間がかかっている。より効率的で標準的な方法が必要であり、その検討を行っている。

近年、「疾患において表出する臨床症状」を「標準化した用語」で表すデータベース Human Phenotype Ontology (HPO) (<https://hpo.jax.org/app/>) が公開され、このデータベースを元に臨床情報から候補疾患を自動的にリストアップする Phenomizer などのアルゴリズムが Web 上で利用できるようになったが、未だ十分なものではない。また、HPO と NGS データを組み合わせた補助診断ツールの開発が最近開始されており、申請者らは 5 つのツール (PhenIX、hiPHIVE、eXtasy-order statistics、eXtasy-combined max、Phen-Gen) を検討し、PhenIX、hiPHIVE が補助診断ツールとしての機能が比較的高いことを明らかにした (Ebiki et al. Yonago Acta Medica 2019)。しかし、これらのアルゴリズムではインプットする HPO が限られており、その診断結果は満足できるものではない。多様な臨床症状に対応し診断を効率化するには、電子カルテから自動的に HPO を抽出し、PhenIX、hiPHIVE などに自動的に導入して診断できるシステムの構築が必要となる。

このシステムの開発によって遺伝性疾患の診断効率を向上させることにより、日本のゲノム医療をさらに推進することが期待できる。

2．研究の目的

本研究の当初の目的は、研究開始当初の背景から臨床自然言語処理（CNLP）と次世代シーケンサー（NGS）を用いた自動診断アルゴリズム（システム）の医療実装であった。しかし、この研究期間内に、染色体レベルの網羅的解析技術を用いたマイクロアレイ染色体検査が保険収載された。マイクロアレイ技術も NGS と同様に網羅的解析であり、本技術を用いた診断は、酷似している。そのために、マイクロアレイ染色体検査の効率的な診断についても検討を行った。さらに、Social Network Service(SNS)から情報を分析することが、臨床自然言語処理の情報に役立つと考えられ、Twitter を用いて検討を行った。

3．研究の方法

1) 自然言語処置（CNLP）の検討

神経疾患の一つである歯状核赤核淡蒼球ルイ体萎縮症（DRPLA）の模擬患者カルテの作成を行い、CNLP を抽出のための Human Phenotype Ontology を用いたアルゴリズム、さらに、CNLP のアルゴリズムの特性分析を試みた。

2) マイクロアレイ染色体検査の検討

UCSC Genome Browser (<http://genome.ucsc.edu/index.html>)、Decipher (<https://www.deciphergenomics.org/>) などの Web 情報サイト、さらに慶應義塾大学医学部臨床遺伝学センターから提供されている「マイクロアレイ染色体検査の結果解釈の補助ソフトウェアツール」(補助ソフトウェアツール) などを利用し、マイクロアレイ染色体検査での効率的な診断についても検討した。

3) SNS からの情報の検討

発達障害の代表的な疾患である脆弱 X 症候群（FXS）とダウン症候群（DS）に例とし、これらの疾患に関する 2022 年 1 年間のツイートの情報を Python 言語による Tweepy を用い収集し、類似点と相違点を客観的に分析した。

4．研究成果

1) CNLP の検討

DRPLA の模擬患者カルテを元に、CNLP として、tidytext、txt2vec、quanteda などの特性の分析を開始したが、明確な差を見出すことはできなかった。さらに、症例報告論文から CNLP アルゴリズムを用いた Term さらに HPO の抽出を行ない、そのデータを元に実際の電子カルテでの検討を構想したが、新型コロナウイルス感染の拡大により、具体的に検討を進めることができなかった。

2) マイクロアレイ染色体検査の検討

補助ソフトウェアツールは、使い方が簡便で、アプトプットされるデータは最終的に診断に結びつく病的バリエーションのみとなっており専門知識がなくても利用できる点がメリットである。しかし、対象疾患については英国 DECIPHER プロジェクトが定義している 6 6

疾患と限定的であるため、専門知識をもつ者が、UCSC Genome Browser

(<http://genome.ucsc.edu/index.html>)や Decipher

(<https://www.deciphergenomics.org/>)などの Web 情報サイトを利用した検討を加える必要があった。今後は、臨床自然言語処理 (CNLP) を組み合わせ、UCSC Genome Browser や Decipher と連携した自動解析アルゴリズムを開発し、効率的な診断体制を構築することが課題と考えられた。

3) SNS からの情報の検討

FXS に関するツイートは 3,732 件、DS に関するツイートは 2,528 件あり、FXS では 72,612 語、DS では 36,894 語が検出された。頻度に基づく単語リストでは、FXS と DS の共通の上位 10 語に「意識」「子供」が含まれ、FXS では「遺伝子」や「研究」が特徴的であった。共起ネットワーク分析では、FXS と DS の単語間の関係が可視化され、キーワードに基づいてサブグループに分類された。FXS では「キャリア」に関するグループが見られたが、DS にはなかった。この結果は、FXS と DS に関する情報の違いを示している。今回の検討では、Twitter の情報の発信源が家族なのか、研究者なのか、医療の専門家なのかは分類することはできなかったが、多くの人々が利用する SNS 上の情報を利用し CNLP システムの精度の向上を図り、より実用的な自動診断システムの開発に貢献すると考えられた。

4) 今後の研究の方向について

新型コロナウイルス感染パンデミックのために、残念ながら、本研究では予定どおりの研究成果は上がらなかった。2023 年 6 月 9 日、「良質かつ適切なゲノム医療を国民が安心して受けられるようにするための施策の総合的かつ計画的な推進に関する法律案」が参議院本会議で可決され成立した。この法律では、ゲノム医療の研究や医療応用を推進するとともに、生命倫理への適切な配慮、ゲノム情報による不当な差別が行われることのないようにすることなどが定められている。

このような背景から、AI による臨床自然言語処理 (CNLP) などを用い、ゲノム情報の取り扱いやセキュリティに配慮した、医療用のゲノム解析情報自動診断装置を開発してゆくことが必要になる。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計0件

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

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 | 岡崎 哲也 (OKAZAKI Tetsuya) (30465299) | 鳥取大学・医学部附属病院・助教 (15101) | |
| 研究分担者 | 足立 香織 (ADACHI Kaori) (50609237) | 鳥取大学・研究推進機構・准教授 (15101) | |

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 |
|---------|---------|
| | |