

科学研究費助成事業 研究成果報告書

平成 29 年 6 月 9 日現在

機関番号：32665

研究種目：挑戦的萌芽研究

研究期間：2014～2016

課題番号：26540171

研究課題名(和文) 会議録を活用した病理学的ファクトデータベースの構築と利用法に関する研究

研究課題名(英文) Construction and analysis of pathological fact database using minutes

研究代表者

中西 陽子 (NAKANISHI, Yoko)

日本大学・医学部・助教

研究者番号：90366592

交付決定額(研究期間全体)：(直接経費) 1,800,000円

研究成果の概要(和文)：病理診断は患者から採取された組織や細胞を形態学的に評価する。しかし近年、分子生物学や癌治療の発展によりゲノミクスなどを取り入れた新たな領域への変革も求められてきている。本研究は、学会報告を活用した病理学的情報のデータベース構築を目的として行った。日本病理学会会誌に掲載されている会議録計約22,400件(2000-2015)より、乳癌研究と肺癌研究計624件を電子化した。うち466件をMeCabによりテキスト解析し、抽出された単語の関係をワードクラウドならびに共起ネットワークで解析した結果、癌種による年次別の特徴が見出された。会議録情報の蓄積と解析により疾患毎の重要事項が継続的に更新可能となる。

研究成果の概要(英文)：Pathological diagnostics estimates the morphologies of tissues and cells that have been obtained from patients. Recently, molecular biology and cancer therapies have been developed dramatically. Therefore, new concepts to pathological diagnostics combined with various disciplines like genomics are required. In the present study, we aimed to construct the fact database of pathology using conference minutes that were preliminary but temporally information. The proceedings of Japan society of pathology were used and each abstract was changed to each text file. We analyzed words in 466 abstracts about breast cancer research and lung cancer research from total 22,400 abstracts (2000-2015) using MeCab. The correlation among each word was analyzed by Word Cloud and co-occurrence relation network. These results showed that the year's trend of each research category. Construction and analysis of pathology minute information become able to renew important topics of every disease continually.

研究分野：分子生物学 病理学 図書館情報学

キーワード：病理学 会議録 ファクトデータベース 診断 治療

1. 研究開始当初の背景

分子生物学の飛躍的な進歩により癌の治療薬の開発が目覚ましい。このため、治療効果が期待できる患者を選定するための病理学的検査分野の早急な情報整備が必要となってきた。医学分野の中で病理学は、臨床医学と基礎生命科学の接点にあり、ヒトの病気について、臓器から細胞レベルまでの形態の変化や差異に重点を置いて研究する分野である。これまでに蓄積された成果に基づいた知見により、病理診断が行われている。

国内の病理学的情報の組織化として、典型的な症例については、医療従事者向け情報提供として臓器の写真や組織の顕微鏡写真がデジタルアーカイブ化されており、病理コア画像のレファレンスデータベース (<http://pathology.or.jp/corepictures2010/index.html>) が日本病理学会ホームページ上で公開されており、教育や診断に利用されている。

また、顕微鏡観察用のスライドガラス標本全体をデジタル画像としてコンピュータに記録してモニタ上で観察するシステムが、バーチャルスライドとして開発されて以来、ガラスの破損や退色の心配がなく、遠隔診断やコンサルテーションなどにも活用されている。国外ではこのようなバーチャル画像のレファレンスデータベース構築も盛んである。

このように、病理学における形態情報の組織化は成熟しつつある。しかし近年、癌の診断においては、治療に直結する情報として個々の患者の病理組織検体から様々な分子情報を解析して提供する必要性が生じてきた。従来の病理学に加えて新たな専門分野の知識が必要となり、その内容は多岐にわたる。さらに、近年、形態だけではなく分子情報を加えた診断概念の改訂や、治療法の更新が次々に行われている状況である。治療薬の開発から承認までの期間も早くなっていることから病理学周辺の各種情報の集約は容易ではない。

そこで本研究では、診断ならびに治療の変化に呼応する病理診断支援システムの構築を目的とした。疾患ごとに異なり、加速度的に更新される診断、治療に関連する情報を速やかに集約するため、本研究では、速報性に優れたものの一過性の情報となっている学会発表に着目した。学会の開催毎に新しい情報が速報的に更新される会議録を活用を試みた。

2. 研究の目的

本研究では、最新情報に富む学会報告の活用に着目し、会議録を利用したファクトデータベースの構築と、個別化医療への支援を目指した二次利用の有用性について検討することを目的とした。

3. 研究の方法

(1) データの選定

学会報告の会議録は医学中央雑誌を代表とする様々な書誌データベースに収録されているが、本文のテキストデータは収録されていない。本研究では、日本病理学会総会議録を対象とし、会議録本文より抽出した用語を対象データとした。対象年度と対象の会議は2000年から2015年に開催された総会（春季開催）とした。データ源である日本病理学会会誌に掲載されている会議録約22,400件より、発表数が多い乳癌研究ならびに肺癌研究に関する会議録計624件をテストサンプルとした。

(2) データの電子化

対象年度の日本病理学会会誌は冊子版のみであったため、ScanSnapSV600（富士通株式会社、東京、日本）を用いてテキストファイル化した。

(3) データ解析

テキスト化した会議録より466件をMeCab (Social Insight, ユーザーローカル、東京、日本)を用いて解析した。抽出された単語間の関係はSocial Insight (ユーザーローカル)を用いて、ワードクラウドならびに共起ネットワークで解析した。

4. 研究成果

(1) 発表数の変遷

日本病理学会総会における肺癌研究、乳癌研究の発表数はそれぞれ10年で1.2倍、1.6倍、15年で2.0倍、2.6倍と顕著な増加が示された。これらより抽出された単語数は、肺癌計119,969語、乳癌計59,242語であり、抄録1件あたりの単語数は肺癌423語、乳癌363語であった。

(2) 用語の出現頻度による特徴解析

抽出された用語の出現頻度分布をワードクラウドで解析すると、各年度の学会で注目度の高かった解析方法や解析の目的が癌種によって異なる状況が浮き出された（図1）。

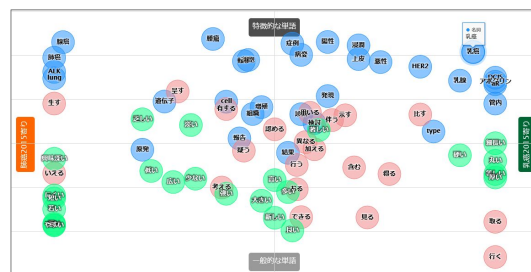


図1-A. 2015年に日本病理学会総会で発表された乳癌研究と肺癌研究の特徴分布

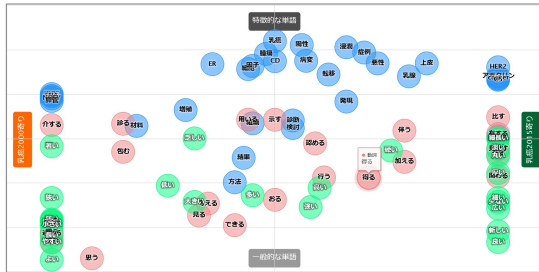


図 1-B . 2000 年と 2015 年に日本病理学会総会で発表された乳癌研究の特徴分布

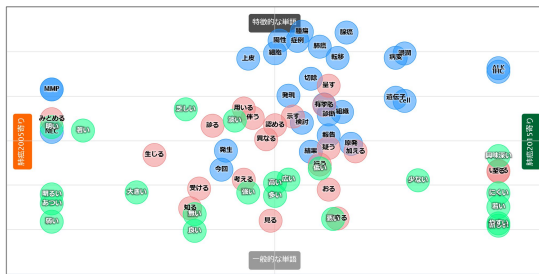


図 1-C . 2005 年と 2015 年に日本病理学会総会で発表された肺癌研究の特徴分布

図 1. 単語の出現頻度による特徴解析

今回のサンプル数では、出現頻度の上位にくる単語は、解析年度毎の病理診断や癌治療に必須の分子や状態を示していた。例えば、2015 年の例では、乳癌では、癌の浸潤性や治療選択に必要な癌の type (型別) や HER2 という治療標的分子の重要度が高い。一方、肺癌では、特に腺癌という種類の肺癌における ALK という治療標的分子の重要度が高い。また、乳癌と比較して、病理診断に遺伝子情報が必要となっていることがわかる (図 1-A)。

経年変化では、乳癌のホルモン療法に必要な ER (エストロゲン受容体) が重視されていた 2000 年と比較して、分子標的治療が進んだ 2015 年では、治療標的分子の着目度が高い (図 1-B)。肺癌でも癌の浸潤性や、神経内分泌癌についての検討から、特に最も患者数の多い腺癌という種類の肺癌の遺伝子情報が大規模に調査されて明らかとなり、分子標的治療が主流となった 2015 年ではその特徴が見られた (図 1-C)。

(3) 用語間の関係解析

次に、共起ネットワークにより、出現頻度の高い用語間の関係を解析した。これらの結果から、各年度で主流の概念と関連付けが低いものの周辺に出現する新たな概念が、癌種また、年度別に見出された (図 2)。

乳癌では、2015 年には治療標的分子である HER2 が乳癌における主要な用語となっていることが示された (図 2-A, 図 2-B)。乳癌が乳癌一種類として表現されている一方で、肺癌は、肺癌というよりもその下位概念として分類される腺癌や扁平上皮癌という用語それぞれに異なる用語の関係性が見られることが示された (図 2-C、図 2-D)。

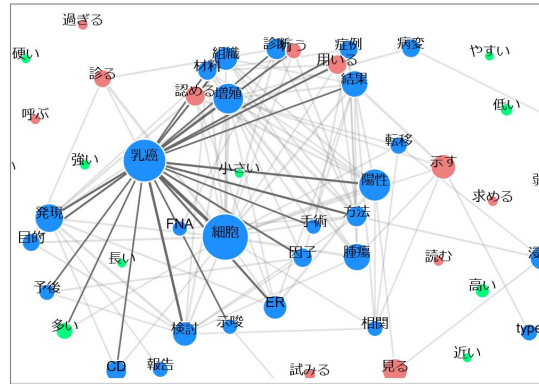


図 2-A: 2000 年乳癌研究

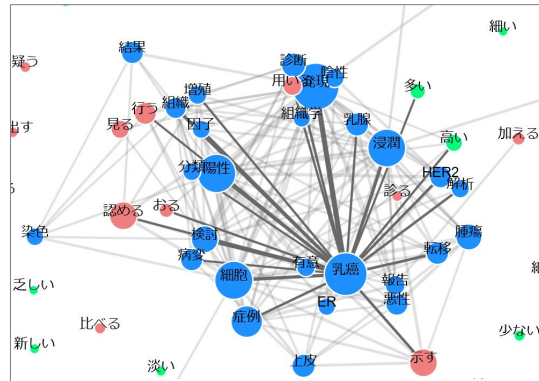


図 2-B: 2015 年乳癌研究

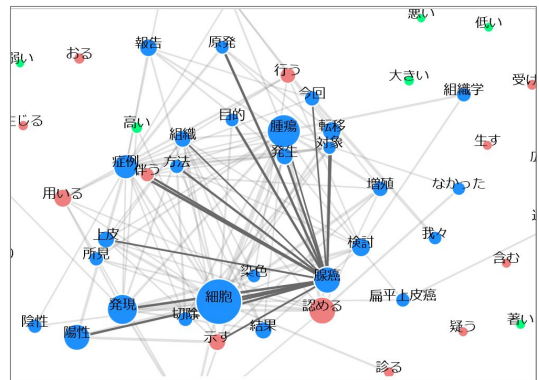


図 2-C: 2000 年肺癌研究

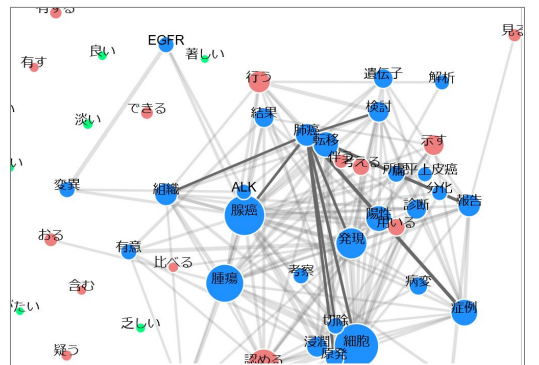


図 2-D: 2015 年肺癌研究

図 2. 共起ネットワーク解析

(4) 考察

病理学や病理診断が形態学を基盤としていることから、形態情報としての病理学情報は集約とデータベース化が進んでいる。しか

しながら、近年の分子生物学の目覚ましい発展を基盤とした疾患概念の変遷や、治療法の発展に随時呼応するためには、主要なキーワードをもとに、病理学周辺の分野へと拡張した関連付けが必要である。本研究では、時事的な速報性に富みながらも一過性となっている会議録情報を電子化して蓄積し、これまでの画像情報ではなく、文字情報を解析することを試みた。この結果、年度毎の疾患の病理診断や治療を取り巻く重要事項が特徴的に捉えられ、継続的に更新可能となることが期待された。今後さらに、抽出された分子や薬剤および疾患名などについての各種情報の関連付けをより具体的に行い、実用性の向上を図る必要があると考える。

<引用文献>

Johnston DJ, Costello SP, Dervan PA, O'Shea DG. Development and preliminary evaluation of the VPS ReplaySuite: a virtual double-headed microscope for pathology. BMC Med Inform Decis Mak. 2005;5:10.

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 0 件)

[学会発表](計 5 件)

中西陽子、佐藤 永、原 一貴、病理情報のデータベース化. 第 34 回医学情報サービス研究大会大阪枚方大会, 関西医科大学(大阪府), 2017 年 8 月 26 日(掲載確定)

中西陽子、病理診断支援システムの構築. 第 14 回バイオアカデミックフォーラム, 東京ビッグサイト(東京都), 2017 年 6 月 29 日

中西陽子、関 利美、廣谷ゆかり、大荷澄江、楠美嘉晃、大幸英至、吉田行弘、徳橋泰明、山館周恒、杉谷雅彦、骨軟部腫瘍組織を用いた融合遺伝子検査自動化の検討. 第 47 回日本臨床検査自動化学会, パシフィコ横浜(神奈川県), 2015 年 10 月 8 日

中西陽子、根本則道、逸見明博、シンポジウム 1 分子組織細胞化学の新展開 病理組織・細胞診検体からの分子形態学的解析 診断・治療への展開 第 47 回日本臨床分子形態学会総会・学術集会, 長崎大学(長崎県), 2015 年 9 月 18 日

中西陽子、関 利美、廣谷ゆかり、鈴木淳子、坂本麻菜美、楠美嘉晃、逸見明博、増田しのぶ、杉谷雅彦、根本則道、細胞検査士会 ヤング committee ワークショップ LBC 検体の標本作製と応用 遺伝

子解析への応用「呼吸器検体の LBC と肺癌分子診断への応用」. 第 56 回日本臨床細胞学会総会, くにびきメッセ・松江テルサ(島根県), 2015 年 6 月 13 日

[図書](計 2 件)

中西陽子 他、丸善出版、D.病理診断技術 7 コンパニオン診断技術、2017、271-278

中西陽子 他、中西印刷、レーザーマイクロダイセクション法の基礎と応用、2015、pp.199-209

[産業財産権]

出願状況(計 0 件)

名称:
発明者:
権利者:
種類:
番号:
出願年月日:
国内外の別:

取得状況(計 0 件)

名称:
発明者:
権利者:
種類:
番号:
取得年月日:
国内外の別:

[その他]
ホームページ等

6. 研究組織

(1)研究代表者

中西 陽子(NAKANISHI, Yoko)
日本大学・医学部・助教
研究者番号: 9 0 3 6 6 5 9 2

(2)研究分担者

()

研究者番号:

(3)連携研究者

()

研究者番号:

(4)研究協力者

佐藤 永(SATO, Ei)
原 一貴(HARA, Kazuki)