

令和 5 年 5 月 14 日現在

機関番号：14603

研究種目：基盤研究(B)（一般）

研究期間：2020～2022

課題番号：20H04283

研究課題名（和文）呼気成分による複数疾病のAI診断と生物学的基盤に関する研究

研究課題名（英文）AI diagnosis of multiple diseases by exhaled gas components and their biological interpretation

研究代表者

作村 諭一（Sakumura, Yuichi）

奈良先端科学技術大学院大学・データ駆動型サイエンス創造センター・教授

研究者番号：50324968

交付決定額（研究期間全体）：（直接経費） 13,700,000円

研究成果の概要（和文）：本研究では、呼気内の揮発性有機化合物（VOC）を利用した疾患診断の性能評価とその生物学的な解釈を行った。扱った疾患は肺癌、歯周炎、NASH、肝がんであり、VOCに対して適切な前処理を適用することで、簡素な機械学習法で極めて高精度の診断が可能であることが分かった。また、肝臓内のシトクロムP450が代謝するVOCと腸内細菌が分泌するVOCの組み合わせが、これらの疾患の診断に有効であることが分かった。また、がん細胞内の同時観測できない分子濃度を同時観測化する技術開発に取り組み、擬似的な同時観測データへの変換手法を開発した。

研究成果の学術的意義や社会的意義

呼気は血液成分を反映するため血液検査と同等の生体情報を持っている可能性がある。しかし多くの研究で共通したバイオマーカーが明確ではなく、その生物学医学的な解釈もされていない。そのため、呼気による血中アルコール濃度の推定が実用化されているにも関わらず、呼気VOCによる疾患診断の実用化が進んでいない。本研究は、特定のVOCではなく、複数のVOCの特定の組み合わせが疾患診断の情報を持っていること、それらのVOCは特定の酵素と腸内細菌が共通ファクターとなっていることを明らかにし、臓器の機能レベルを表すバイオマーカーとして生物学医学的に解釈できることを示した。

研究成果の概要（英文）：In this study, we evaluated the performance of disease diagnosis using volatile organic compounds (VOCs) in exhaled breath and their biological interpretation. The diseases examined were lung cancer, periodontitis, NASH, and liver cancer. We discovered that by applying appropriate preprocessing to VOCs, a simple machine learning method can provide extremely accurate diagnoses. Moreover, the combination of VOCs metabolized by cytochrome P450 in the liver and those secreted by intestinal bacteria proved to be effective in diagnosing these diseases. Additionally, we focused on developing technology to convert the concentration of molecules that cannot be simultaneously observed in cancer cells into simultaneous observations. As a result, we have devised a method to transform the data into pseudo-simultaneous observation data.

研究分野：データ駆動型生物学

キーワード：呼気揮発性有機化合物 疾患診断 機械学習

1. 研究開始当初の背景

病気は早期に診断され治療されることが望ましい。そのためには、診断が気軽に行えるシステム作りが必要である。医師の負担を軽減するために、機器による自動化が必要である。そのような背景のもと、非侵襲検査による診断技術の開発が盛んである。警察によるアルコール検査が呼気で行われるように、血中の揮発性有機化合物 (VOC) 成分は肺を介して呼気として排出される。これまで本研究グループは、健康・肺がん・歯周病・非アルコール性脂肪肝炎 (NASH)・肝がんの呼気成分と AI (機械学習) を用いた診断技術開発を行ってきた (図 1)。しかし、実用化には課題が残る。

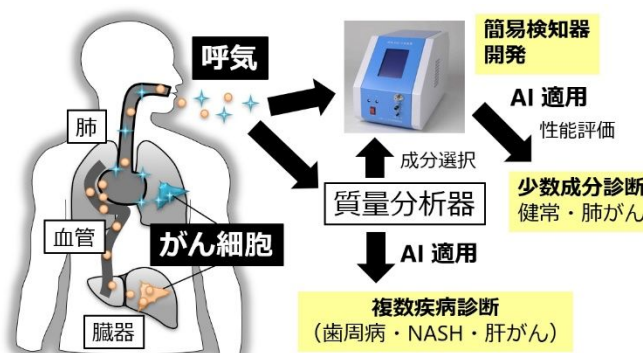


図 1：呼気成分と AI による簡易検査器開発と複数疾病診断。

2. 研究の目的

本研究の目的は、上記課題に含まれる医学・産業・情報科学・基礎生物学にまたがる学際的な課題の解決によって、呼気 VOC による診断の有効性を示し普及の足がかりとすることである。実用面においては、例えば診断の偽陽性が多ければ逆に医師の負担を増やしかねない。ゆえに性能向上だけでなく、基礎生物学に基づいた検証結果を添えることで、利用する医師の信頼を得る。我が国が得意とするセンサ開発をヘルスケアへのアプリケーションとして展開することは本研究の大局的・俯瞰的な目的である。また、簡単に小型な検知器開発が進めば体内データの経時計測が可能となる。その結果、例えば一日中の変動から健康状態を予測するなど、新しい医学研究や予防医療法が期待できる。血压検査のような気軽に病気の早期発見やリスク判定が可能になれば、当分野の学術的なレベルアップに大きく寄与できる。本研究課題の期間 (2020-2022 年度) のほとんど全ての期間は、新型コロナウイルスの流行が重なったため、当初予定していた呼気採取がほとんど不可能であった。ゆえに、データ解析、細胞の代謝物の検討を中心的な目的とした。

3. 研究の方法

(1) 呼気 VOC による健康・歯周病・NASH・肝がんの疾患診断の精度向上と生物学的解釈

倫理委員会の承認 (広島大学, E-922) を得て、被験者からの書面によるインフォームドコンセントを得た後、健康者 101 人、NASH 患者 67 人、肝がん患者 95 人、歯周炎患者 29 人から呼気をサンプルした。サンプルされた呼気に付随する口腔内診査や臨床情報等、個人を識別できる情報を削除し、独自の記号を付与することにより呼気サンプルを匿名化し、個人を特定を不可能にした。生体データは個体差が大きいいため、その排除を目的とした前処理が不可欠である。本研究では各 VOC を疾患に関連する VOC の総和に対する比率を用いた。これにより、肺機能の個体差と疾患に関係しない VOC による影響を排除した。疾患に関連する VOC を機械学習による診断精度に基づいて判断した。本研究では、教師あり機械学習で VOC 濃度比から健康と疾患を診断する学習器を構築した。健康者と疾患、または疾患同士の 2 クラス分類では非線形 support vector machine (SVM)、健康者と複数の疾患の多クラス分類ではランダムフォレストを採用した。精度評価は 10 分割交差検証法で行った。

(2) がん細胞の代謝物の解析

正常肝細胞 (Hc3176 細胞株) に 0.2M パルミチン酸を 18 時間作用することにより脂肪沈着を誘導した脂肪化肝細胞 (パルミチン酸-Hc3176 細胞) の培養上清中に排出された揮発性のガス成分を GC/MS を用いて解析し、正常細胞と比較検討した。一方、肝がん細胞の分泌成分解析には、HepG2 細胞と HuH7 株の培養上清中の排出ガス成分を調べ、正常肝細胞株と比較した。測定については専門業者に依頼し、細胞 12 検体について、CE-TOFMS のカチオンモード、アニオンモードによる測定を実施、代謝物質の同定を行った。

(3) がん細胞内の分子活性の時系列解析

VOC や代謝物について、複数の種類を同時に計測しているが、スナップショット、すなわち、ある 1 時刻の計測であるため、時間変化を追うことができない。しかし、疾患は生体の時間的な変化であり、将来的に VOC や細胞の代謝物の時間経過が必要になることが予想される。計測対象の種類を複数にすることと、時間経過を見ることはトレード・オフの関係にあり、実験的に解決することは難しい。そこで、個別に時間経過を観測したデータを変換し、擬似的に同時観測時系列にする手法開発を行った、研究対象の表現型に関して同一の時系列を参照信号として同時観測化した。

4. 研究成果

(1) 呼気 VOC による健康・歯周病・NASH・肝

がんの疾患診断の精度向上と生物学的解釈

全体的な傾向を見るため、16 種の VOC をサンプルごとに比率に変換し、健康状態別にサンプルの分布を調べた (図 2)。

16 種の VOC は疾患別に異なる傾向を示し、酵素 P450 関連、腸内細菌由来、その他の VOC で健康コントロールの濃度比率が小さく、

NASH・肝がんでは大き

かった。xylene のみ逆の

関係が見られた (図 2A)。腸内細菌由来の VOC も、健康者に比べ疾患群で増加傾向にあり、酢酸や Acetoin など一部の VOC は NASH と肝がん異なる傾向だった (図 2B)。これらの結果は、肝機能低下に伴う呼気中 VOC 濃度の上昇、肝疾患患者の呼気中 Limonene 増加、肝疾患患者の特定腸内細菌の増加といった先行研究と一致する。さらに、他の VOC 比率でも疾患別に異なるパターンが見られた (図 2C)。これら VOC 比率は疾患に関する情報を持っている可能性が高い。

診断精度を VOC の 3 つのカテゴリの組み合わせに分けて比較した (図 3)。P450 カテゴリの VOC の組み合わせだけでも精度が高く、腸内細菌 (Gut) カテゴリを追加した条件 (P450-Gut)

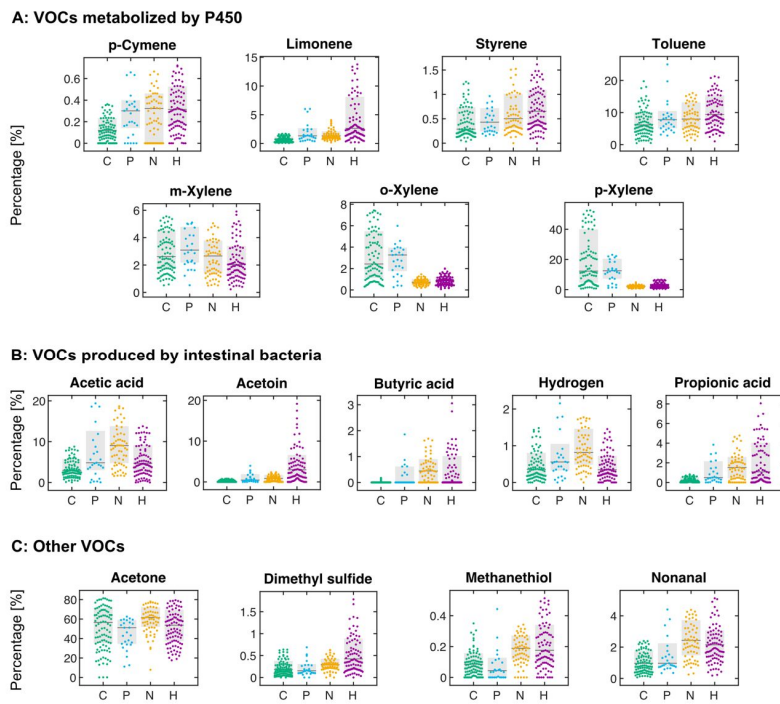


図 2 : 比率に変換した際の VOC の分布。

では更に精度が向上した。Others カテゴリを含めた 3 カテゴリ条件ではそれほど精度が向上しなかった。この結果は、健常者と比較して肝疾患患者は肝機能が低下し、腸内細菌叢の組成が変化するという疫学的知見と一致する。

P450-Gut 条件の多くでは NASH と肝がんの診断精度が 90% を超え、歯周炎の最高精度は P450-Gut-others 条件の組み合わせであった。

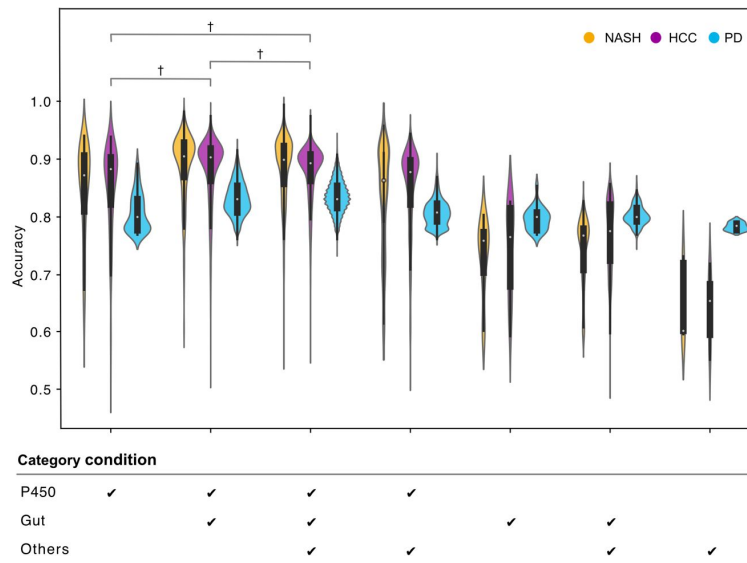


図 3：呼気 VOC の 3 つのグループの組み合わせにおける診断性能分布。

(2) がん細胞の代謝物の解析

がん細胞と正常細胞の資料について表 1 にまとめた。表 1 の細胞から、合計 110 種類の代謝物が同定された。各代謝物について、内部標準物質を用いて濃度を算出した。代謝物について主成分分析をかけ、2次元で細胞の分布を確認したところ(図 4)、Control と Fat 細胞は近く、肝がん細胞は異なるクラスターを形成した。階層的クラスタリングで正常細胞とがん細胞を比較した場合も、代謝物の濃度が大きく変わることがわかった。また、多くの物質が統計的に有意な差があることを示した。現在、診断に有効な呼気 VOC との関係性を調査中である。

表 1：代謝物を解析した細胞

試料名	試料量 ($\times 10^6$ cells)	群名
HC-3716-1	2.24	Control
HC-3716-2	2.24	
HC-3716-3	2.24	
HC-3716Pal-1	2.08	Fat
HC-3717Pal-2	2.08	
HC-3718Pal-3	2.08	
Hu7-1	1.96	Liver-cancer1
Hu7-2	1.96	
Hu7-4	3.65	
HLF-1	2.64	Liver-cancer2
HLF-2	2.64	
HLF-3	2.64	

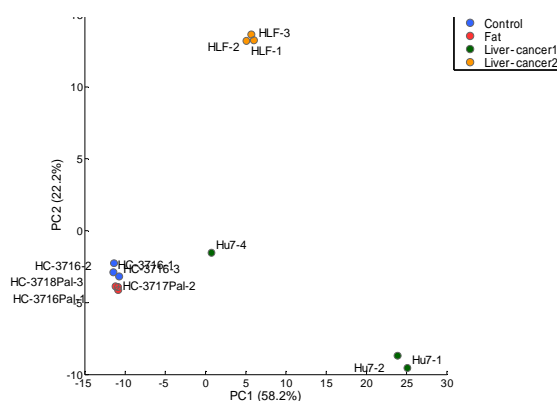


図 4：110 種の代謝物の主成分における細胞の分布

(3) がん細胞内の分子活性の時系列解析

個別に時間変化を観測した例として、がん細胞の運動のデータを利用した。細胞内の分子はそれぞれが簡素な役割しか持たないが、複数の分子が協調することで細胞運動を含む高度な機能を実現する。Rho GTPase の Cdc42、Rac1、RhoA が細胞運動に「関係する」ことは、これまで多くの細胞生物学の研究で報告されてきた。しかし、これらの分子がどのように協調して細胞運動を制御するのか未解明であった。ゆえに、これら分子の同時観測時系列が必要であるが、一般的に各細胞で 1 種類の分子活性度を計測し、複数種類の活性度を同一細胞で同時に計測することは技術的に極めて困難である。本研究ではデータ前処理法 motion-

本研究ではデータ前処理法 motion-

triggered average (MTA) を開発し、個別の細胞で計測された 3 つの Rho GTPase の活性度時系列を擬似的に同時計測と見なせる時系列データに変換した。Rho GTPase の活性度を計測したイメージングデータから、ある細胞エッジにおける 1 種類の Rho GTPase 活性度とそのエッジの変形速度の時系列が得られる。また、どの細胞エッジも拡張と収縮の連続で変化する。したがって、短時間であれば Rho GTPase の種類に関係なく類似したエッジ速度時系列をほとんどの細胞から見つけることが可能である。本研究では各 Rho GTPase ごとに MTA を適用し、エッジが同じ動きをしたときの活性度時系列を収集し、データを平均化した (図 5)。これにより、エッジの特定の速度パターンと同時に発生した 3 つの Rho GTPase 活性度パターンを抽出できた。これらの活性度パターンは特定の速度パターンを生み出す最も典型的な時系列であり、擬似的に同時計測時系列と見なせる。抽出された 3 種の活性度時系列は、個々の Rho GTPase に関する先行研究の報告と一致する性質を持っており、細胞エッジの速度時系列に対してそれぞれの分子が固有の活性度時系列を持つことが分かった。

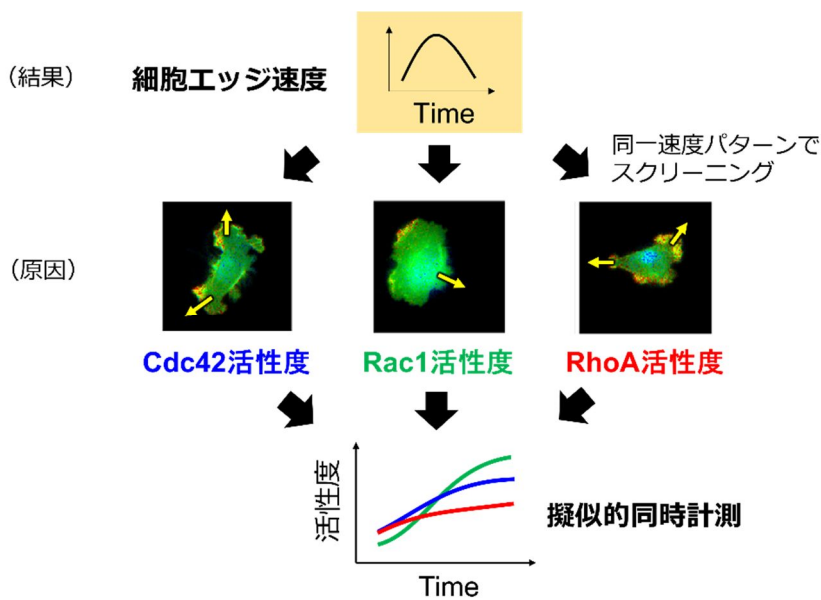


図 5 : MTA の概要。個別観測された HT1080 細胞の 3 種の Rho GTPase を、細胞エッジの動きを参照シグナルとして同時観測化した。

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件 / うち国際共著 0件 / うちオープンアクセス 1件）

1. 著者名 Katsuyuki Kunida Nobuhiro Takagi Kazuhiro Aoki Kazushi Ikeda Takeshi Nakamura Yuichi Sakumura	4. 巻 42
2. 論文標題 Decoding cellular deformation from pseudo-simultaneously observed Rho GTPase activities	5. 発行年 2023年
3. 雑誌名 Cell Reports	6. 最初と最後の頁 1-16
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.celrep.2023.112071	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 Toshio Itoh, Yutaro Koyama, Yuichi Sakumura, Takafumi Akamatsu, Akihiro Tsuruta, Yoshitake Masuda, Woosuck Shin	4. 巻 387
2. 論文標題 Discrimination of volatile organic compounds using a sensor array via a rapid method based on linear discriminant analysis	5. 発行年 2023年
3. 雑誌名 Sensors and Actuators B: Chemical	6. 最初と最後の頁 1-12
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.snb.2023.133803	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 -

〔学会発表〕 計2件（うち招待講演 1件 / うち国際学会 2件）

1. 発表者名 Katsuyuki Kunida & Yuichi Sakumura
2. 発表標題 Decoding cellular deformation from pseudo simultaneously observed Rho GTPase activities
3. 学会等名 International Conference on Systems Biology（国際学会）
4. 発表年 2022年

1. 発表者名 Yuichi Sakumura
2. 発表標題 Quantitative conversion from Rho GTPase signals to cellular deformation by pseudo-simultaneous measurements
3. 学会等名 The 2nd International Conference on Genomics, Nanotech, and Bioengineering（招待講演）（国際学会）
4. 発表年 2022年

〔図書〕 計1件

1. 著者名 作村諭一, 伊藤敏雄	4. 発行年 2020年
2. 出版社 技術情報協会	5. 総ページ数 13
3. 書名 においのセンシング、分析とその可視化、数値化	

〔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	宮内 睦美 (Miyuchi Mutsumi) (50169265)	広島大学・医系科学研究科(歯)・教授 (15401)	
研究分担者	清水 淳市 (Shimizu Junichi) (80796889)	愛知県がんセンター(研究所)・分子診断TR分野・研究員 (83901)	
研究分担者	申 ウソク (Shin Woosuck) (10357246)	国立研究開発法人産業技術総合研究所・材料・化学領域・副研究部門長 (82626)	
研究分担者	伊藤 敏雄 (Ito Toshio) (90377888)	国立研究開発法人産業技術総合研究所・材料・化学領域・主任研究員 (82626)	
研究分担者	増田 佳丈 (Masuda Yoshitake) (20324460)	国立研究開発法人産業技術総合研究所・材料・化学領域・研究グループ長 (82626)	

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	赤松 貴文 (Akamatsu Takafumi) (60635316)	国立研究開発法人産業技術総合研究所・材料・化学領域・主任研究員 (82626)	
研究分担者	樋田 豊明 (Hida Toyoaki) (80250249)	愛知県がんセンター（研究所）・分子腫瘍学分野・研究員 (83901)	

	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
研究協力者	佐藤 一雄 (Sato Kasuzo)		
研究協力者	池田 和司 (Ikeda Kazushi)		

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