

令和元年6月14日現在

機関番号：82657

研究種目：基盤研究(C) (一般)

研究期間：2016～2018

課題番号：16K00416

研究課題名(和文) パーソナルゲノム時代におけるRNA-seqを活用したプロテオミクス解析手法の確立

研究課題名(英文) Establishment of proteomics analysis methods using RNA-seq in the personal genomic era

研究代表者

河野 信 (Kawano, Shin)

大学共同利用機関法人情報・システム研究機構(機構本部施設等)・データサイエンス共同利用基盤施設・客員准教授

研究者番号：40470075

交付決定額(研究期間全体)：(直接経費) 3,400,000円

研究成果の概要(和文)：本研究課題では、質量分析計を用いて得られたヒトの個人のプロテオームやがんのプロテオームなど、変異を含むプロテオームデータの効果的な解析方法について検討した。次世代シーケンサーを使ってDNA・RNA配列から変異を検出して作成した、変異を考慮したタンパク質配列データベースを利用してプロテオームデータを解析することにより、これまで既存の方法では検出することができなかった、サンプル特有の変異を含むペプチドを検出することが可能になった。

研究成果の学術的意義や社会的意義

さまざまなゲノムの変異が病気に関係しているが、実際に生体内で作用して機能を担っているのは発現しているタンパク質である。変異を含むタンパク質を検出可能にすることは、病気に関係するバイオマーカーの発見や創薬につながる。今後、次世代シーケンサーならびに質量分析計の測定精度がますます向上していくと予想され、“個人”のマルチオミクス情報を組み合わせた解析は、precision medicineを実現する上で必要不可欠なものになると期待される。

研究成果の概要(英文)：In this research project, I examined effective analysis methods of proteome data including mutations, such as human personal proteome and cancer proteome obtained using a mass spectrometer. By analyzing proteome data using a protein sequence database considering mutations, which was created by detecting mutations from DNA / RNA sequences using a next-generation sequencer, it became possible to detect peptides containing sample-specific mutations which has not been possible to detect using conventional methods.

研究分野：バイオインフォマティクス

キーワード：プロテオーム パーソナルゲノム

## 様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

### 1. 研究開始当初の背景

これまで質量分析計を用いたヒトプロテオーム研究は、得られた質量スペクトルをヒトタンパク質の標準的な配列を集めたリファレンス配列データベースに対して検索して配列を決定していたため、変異を含むサンプルでの精密な配列同定は困難であった。ヒトにおいては個人ごとに DNA 配列の違いに起因するリファレンス配列とのアミノ酸配列の違いが存在し、特にがん細胞においては正常細胞に対して多量の変異が蓄積していることが知られている。そのため、リファレンス配列に対してではなく、リファレンス配列に対して変異や挿入・欠失、スプライシングバリエーションなどを含んだ個別のデータベース(バリエーションデータベース)に対して検索を行う必要がある。

近年の次世代シーケンサ(NGS)技術の発展によって、個人や個々の細胞のゲノム配列やゲノム配列から実際に転写されているトランスクリプトーム配列を容易に測定することが可能になった。そこで本研究課題では、NGS から得られたゲノム/トランスクリプトーム配列から質量スペクトル検索用のバリエーションデータベースを作成し、より高精度なプロテオーム同定法ならびに定量比較法の開発を検討することとした。

### 2. 研究の目的

現在の質量分析を用いたプロテオーム解析では UniProt 等のリファレンス配列に対して検索を行うため、個々のサンプルに変異などがあつた場合、正しく解析できない。本課題では、バリエーションを含むサンプルから、RNA-seq のデータを利用して質量分析によるプロテオームデータを効率的に解析する手法を開発する。また、性質の異なるトランスクリプトームの定量データとプロテオームの定量データから、生物学的に意味のある比較定量を可能にするための方法論を構築する。さらに開発した手法を用いて、新規サンプルの解析を行う。これらを達成することによって、今後研究の加速が予想される個人のマルチオミクス情報を活用したバイオマーカー探索や創薬ターゲット研究などへの進展が期待される。

### 3. 研究の方法

本研究課題では、次世代シーケンサと質量分析装置から得られるデータを組み合わせて、これまで同定が困難であったバリエーションを含むペプチド配列を同定・比較定量するための新規解析手法を開発する。

まず、既存手法との比較を行った。公共データベースに登録されている、いくつかのヒトの同一サンプルから取得された RNA-seq データとペプチドマスデータを収集した。RNA-seq データから SNV やスプライシングバリエーションを考慮した mRNA 配列を作成する。これを従来の遺伝子モデルにそつた形で翻訳、もしくは 6 フレームを機械的に翻訳して、バリエーションを含むアミノ酸配列のデータベースを作成した。このバリエーションデータベースと UniProt の Proteome データセットや Ensembl など、従来からペプチド同定に使われてきたデータセットを利用して、ペプチドマスデータからペプチド同定を行い、結果を比較した。また、RNA-seq のデータからは発現している mRNA の定量が可能であり、ペプチドマスのデータからもスペクトラルカウント法や emPAI 法などによって発現タンパク質の定量が可能であることから、両方法を用いた発現量比較の方法について検討した。さらに、市販されている実際のサンプルから得られたデータを使って、本手法の有効性について確認を行った。

### 4. 研究成果

初年度には、次世代シーケンサより測定された RNA-seq データを利用して変異を含むデータベースをサンプルごとに個別に作成し、これに対して質量スペクトルのデータベース検索を行い、リファレンスデータベースを利用するこれまでの手法と比較した。具体的には、1) UniProt や Ensembl、IPI などのこれまでペプチド同定に用いられてきたデータベース(リファレンスデータベース)、2) 次世代シーケンサから出力されたがんゲノムの RNA-seq データをアセンブリして得た変異等を含むタンパク質アミノ酸配列データベース(バリエーションデータベース)それぞれに対してマススペクトルの検索を行い、発現しているタンパク質の同定を行った。その結果、リファレンス配列に対して検索を行うよりも RNA-seq データから得たデータベースに対して検索したほうがタンパク質の同定精度が向上することが明らかとなった。

RNA-seq のデータからは配列の情報が得られるのに加えてシーケンスタグの出現数から mRNA の発現量を見積もることが可能である。また、質量分析データからもスペクトラルカウント法、emPAI 法などを用いることで発現タンパク質の定量が可能である。そこで次年度には、RNA-seq から得られる mRNA の発現量と、質量スペクトルから得られるタンパク質の発現量を計算し、トランスクリプトーム階層・プロテオーム階層における定量値を比較する方法について検討を行い、単純な構造を持つタンパク質については、比較定量が可能であることを確認した。一方で、RNA-seq より得られた RNA タグが複数の mRNA にマッピングされてしまう場合があること、質量分析でも同様にマススペクトルから同定されたペプチドが複数のタンパク質配列にマッピングされてしまう場合があること、さらにスプライシングバリエーションにより 1 つの遺伝子から複数のプロテオフォームが作られる場合があり、得られたペプチドがどのプロテオフォーム由来か区別できない場合があるなど、容易には比較できない場合も確認できた。また、ペプチドにリン酸化や糖鎖付加などの翻訳後修飾があつた場合、マススペクトル上では異なるペプチドとし

て検出されるため、これを補正する必要があった。また、RNA-seq 法と質量分析法の結果の解像度の違いも問題となった。RNA-seq 法は測定方法にもよるが、比較的全体を網羅した結果が得られるのに対し、質量分析法では装置の解像度の問題から、多量に発現しているタンパク質しか検出できないという問題があり、発現量の少ないタンパク質の定量については課題が残った。

最終年度には、これまで開発してきた手法を実際のがん細胞データ（市販の肺腺がん細胞を使用。プロテオームデータは京都大学薬学研究科にて取得、RNA-seq のデータは DBKERO (<http://kero.hgc.jp/>) から取得) に応用し、解析を行った。解析結果は日本発の proteome データベース jPOSTdb (<https://globe.jpostdb.org/>) から公開する予定である。

## 5. 主な発表論文等

〔雑誌論文〕(計 4 件)

1. Yuki Moriya, Shin Kawano, Shujiro Okuda, Yu Watanabe, Masaki Matsumoto, Tomoyo Takami, Daiki Kobayashi, Yoshinori Yamanouchi, Norie Araki, Akiyasu C. Yoshizawa, Tsuyoshi Tabata, Mio Iwasaki, Naoyuki Sugiyama, Satoshi Tanaka, Susumu Goto, and Yasushi Ishihama, The jPOST environment: an integrated proteomics data repository and database, *Nucleic Acids Research*, 47, D1218-D1224 (2019). 査読あり、DOI: 10.1093/nar/gky899
2. Eric W. Deutsch, Sandra Orchard, Pierre-Alain Binz, Wout Bittremieux, Martin Eisenacher, Henning Hermjakob, Shin Kawano, Henry Lam, Gerhard Mayer, Gerben Menschaert, Yasset Perez-Riverol, Reza M Salek, David L. Tabb, Stefan Tenzer, Juan Antonio Vizcaino, Mathias Walzer, and Andrew R. Jones, The Proteomics Standards Initiative: Fifteen Years of Progress and Future Work, *Journal of Proteome Research*, 16, 4288-4298 (2017). 査読あり、DOI: 10.1021/acs.jproteome.7b00370
3. Shujiro Okuda, Yu Watanabe, Yuki Moriya, Shin Kawano, Tadashi Yamamoto, Masaki Matsumoto, Tomoyo Takami, Daiki Kobayashi, Norie Araki, Akiyasu C. Yoshizawa, Tsuyoshi Tabata, Naoyuki Sugiyama, Susumu Goto, and Yasushi Ishihama, jPOSTrepo: an international standard data repository for proteomes, *Nucleic Acid Research*, 45, D1107-D1111 (2017). 査読あり、DOI: 10.1093/nar/gkw1080
4. Eric W. Deutsch, Attila Csordas, Zhi Sun, Andrew Jarnuczak, Yasset Perez-Riverol, Tobias Ternent, David S. Campbell, Manuel Bernal-Llinares, Shujiro Okuda, Shin Kawano, Robert L. Moritz, Jeremy J. Carver, Mingxun Wang, Yasushi Ishihama, Nuno Bandeira, Henning Hermjakob, and Juan Antonio Vizcaino, The ProteomeXchange Consortium in 2017: supporting the cultural change in proteomics public data deposition, *Nucleic Acid Research*, 45, D1100-D1106 (2017). 査読あり、DOI: 10.1093/nar/gkw936

〔学会発表〕(計 49 件)

1. 河野信、プロテオーム研究のデータジャーナル *Journal of Proteome Data and Method* 創刊に向けて、2018 年度第 3 回 J-STAGE セミナー、2019 年
2. 河野信、HUPO-PSI 最新報告、第 3 回質量分析インフォマティクス研究会ワークショップ、2018 年
3. 河野信、Yasset Perez Riverol, Tobias Ternent, 守屋勇樹, Eric Deutsch, Michel Dumontier, Juan Antonio Vizcaino, Henning Hermjakob, 五斗進、各種レポジトリに登録されたメタデータを収集した OmicsDI の RDF 化、トーゴの日シンポジウム 2017、2017 年
4. Shin Kawano, Yasset Perez Riverol, Tobias Ternent, Yuki Moriya, Eric Deutsch, Michel Dumontier, Juan Antonio Vizcaino, Henning Hermjakob, and Susumu Goto, OmicsDI RDF, 16th Human Proteome Organization World Congress, 2017
5. 河野信、プロテオームデータの標準化とデータベースの世界動向、日本プロテオーム学会 2017 年大会、2017 年
6. Shin Kawano, Yuki Moriya, Shujiro Okuda, Yu Watanabe, Tadashi Yamamoto, Masaki Matsumoto, Tomoyo Takami, Daiki Kobayashi, Norie Araki, Akiyasu C. Yoshizawa, Tsuyoshi Tabata, Susumu Goto, and Yasushi Ishihama, jPOST: Proteome Database Project in Japan, HUPO-PSI meeting 2017, 2017
7. 河野信、生命システムを俯瞰するための質量分析情報解析技術とデータベースの活用、第 39 回日本分子生物学会年会、2016 年
8. 河野信、守屋勇樹、Tobias Ternent, Juan Antonio Vizcaino, Eric Deutsch、プロテオームメタデータの RDF 化、トーゴの日シンポジウム 2016、2016 年

9. Shin Kawano, Yuki Moriya, Tobias Ternent, Juan Antonio Vizcaino, Eric Deutsch, Implementation of flexible search for proteomics metadata, 15th Human Proteome Organization World Congress, 2016
10. 河野信、荒木令江、プロテオームインフォマティクス&システムズバイオロジー研究への招待、日本プロテオーム学会 2016 年大会、2016 年

〔その他〕

ホームページ等

Japan Proteome Standard Repository/Database (jPOST): <https://jpostdb.org/>

jPOSTrepo: <https://repository.jpostdb.org/>

jPOSTdb: <https://globe.jpostdb.org/>

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、研究者個人に帰属されます。