

**科学研究費助成事業 研究成果報告書**

平成 28 年 6 月 22 日現在

機関番号：10105

研究種目：若手研究(B)

研究期間：2012～2015

課題番号：24700290

研究課題名(和文)多様大規模化する生命科学データのための柔軟な統計解析手法の開発

研究課題名(英文)Flexible statistical analysis for huge and multiple data in life science

## 研究代表者

茅野 光範 (Kayano, Mitsunori)

帯広畜産大学・畜産学部・講師

研究者番号：20590095

交付決定額(研究期間全体)：(直接経費) 3,500,000円

研究成果の概要(和文)：本研究では、以下の三つを目標に、生体生命情報学や獣医畜産分野における、柔軟で高精度な統計解析手法を提案し、その応用例を示した。目標 [1]「多種多様大規模生命科学データのための柔軟な統計解析手法の構築」、[2]「動植物の分子・個体・集団レベルのシステム解明を目指す統計解析手法の構築」、[3]「高速な統計解析手法の提案」。[1]と[3]では、各種オミックスにおける高速な検定や統合解析を提案し、診断マーカー検出への応用研究を行った。[2]では、乳牛を中心として、分子レベル(乳成分や各種オミックス等)、個体レベル(診療記録等)、集団(農場)レベルのデータを基に大規模な統計解析例を提示した。

研究成果の概要(英文)：This study proposed flexible and high-performance statistical methods in bioinformatics and showed applications in medical, veterinary science and animal husbandry. Three objectives in this study were: A) constructing flexible statistical methods for multiple big data in life science (particularly, omics data), B) proposal of statistical methods and application to elucidate systems in animals and plants at molecular-, individual-, population-level, and C) constructing fast statistical methods. For A) and C), this study proposed fast hypothesis testing, high-performance statistical methods on each type of omics data. Their applications included finding diseases biomarkers. For B), we showed statistical applications in this field, particularly for dairy cows. Three types of focus in this study were 1) molecular level (milk composition, blood component, and each type of omics data), 2) individual level (medical records in cows, shape of crop), 3) population level (dairy herd etc).

研究分野：バイオインフォマティクス、統計学

キーワード：バイオインフォマティクス オミックス 獣医学 畜産学

## 1. 研究開始当初の背景

(1) 今日と同様に、研究開始当初もヒトやマウス、植物(シロイヌナズナ等)のオミックスデータ(ゲノム、遺伝子、タンパク、代謝産物のデータ)の計測と蓄積が盛んに行われていた。例えば、従来は少数個体のデータしか取れなかったような、アレイデータも容易に(比較的安価で)取得可能になっていた。これにより、複数個体の数万遺伝子の時系列データも入手可能になった。さらに、次世代シーケンサーによるDNA配列情報のデータが膨大に蓄えられつつあった。網羅的な代謝産物の濃度データも蓄えられつつあった。

(2) 家畜等についても、オミックスデータ(特にゲノムや遺伝子のデータ)が計測され始めていた。また、これらや農作物についての個体~集団レベルのデータが膨大に蓄積されていた。例えば、研究代表者が所属する帯広畜産大学やその周辺組織では、網羅的な統計解析が可能な大規模データがあった。農作物についても個体レベルの有益なデータが蓄積されていた。

## 2. 研究の目的

(1) 「多種多様な大規模生命科学データのための柔軟な統計解析手法の構築」(各種オミックスデータの統計解析手法の提案)

(2) 「動植物の分子・個体・集団レベルのシステム解明を目指す統計解析手法の構築」(獣医畜産データの統計解析)

(3) 「高速な統計解析手法(検定等)の提案」

(1) では、各種オミックスにおける交互作用を捉えるための高速な検定や統合解析、診断マーカー検出への応用を提案する。オミックスデータを、各オミックスのレベルで、従来の相関解析や時系列解析に加えて新たな統計手法を提案し適用すると共に、様々なオミックスのレベルの情報を取り入れた解析を行うための統計手法を提案する。

(2) では、牛を中心として、分子レベルのデータ(乳成分や血液成分、各種オミックス等)と、個体レベルのデータ(診療記録や農作物の形状等)、集団(農場)レベルの情報を用いて、獣医畜産分野における広範囲な研究を行う。

(3)では、目的(1)、(2)におけるそれぞれの場面で、高速、かつ、出来るかぎり小サンプルで推定可能な統計解析手法を提案する。

## 3. 研究の方法

(1) オミックス解析のための統計手法としては、相関係数、仮説検定、ロジスティック回帰を中心に、それらを発展・融合して、新たな統計手法を提案する。データは、基本的に、遺伝子発現データベース(GEO)や、先行研究で用いられているデータを用いる。老年性疾患のバイオマーカー検出においては、国立研究開発法人 長寿医療研究センターのオリジナルデータを用いる。

(2) 獣医畜産データの統計解析においては、基本的に、既存手法(ロジスティック回帰や混合効果モデル、主成分分析)を適用する。必要に応じて、ベイズモデルの構築や他の方法と融合した方法を提案する。データは、共同研究者から提供して頂く。

## 4. 研究成果

(1) 2つのサンプルグループ(健常者と患者等のケース・コントロール)間で、共発現パターンに差がある(機能が異なる)遺伝子を検出するための既存の統計手法をまとめ、問題点を議論した(発表論文)。この問題点を克服した高速な仮説検定を提案し、大規模な発現データベースに基づいて手法を検証することを考えている(論文投稿準備中)。検定が高速であることによって、大規模なデータから共発現パターンに差がある遺伝子を網羅的に検出することが可能となる。

(2) 時系列遺伝子発現データに基づいて、時間に依存するマーカー遺伝子を検出するための回帰モデルを提案し、多発性硬化症の発現データに適用した(発表論文、学会発表)。また、時系列で観測されたオミックスデータ(遺伝子や代謝産物のデータ)に対して、サンプルグループ間で発現量や濃度に差がある遺伝子や代謝産物を検出するための仮説検定を発表した(学会発表、)。

(3) 代謝ネットワーク推定のために、遺伝子発現量と、酵素や代謝産物の濃度を統合して解析可能で(統合オミックス解析)、より高精度でネットワーク推定が可能な統計手法を提案した。具体的には、一次偏相関係数に基づいた方法を提案し、その有効性を数値実験、理論、実データによって検証した(発表論文)。

(4) 老年性疾患(特に認知症)のバイオマーカーを検出するための新たな統計手法を提案した。従来のt検定による単一分子の検出ではなく、分子間の交互作用を加味した統計手法を提案した。より高精度で簡便な認知症の血液検査に繋がる成果であると捉えている[公表前であるため詳細は割愛する]。

(論文投稿中1件、論文投稿準備中1件、発表論文、学会発表、産業財産権)

(5) 乳牛のケトosis(出産後に頻発する疾患)を乳量と乳成分から検出するためにロジスティック回帰を適用した。従来のように、単一の乳成分を用いるだけでなく、他の乳成分や乳量を組み合わせることで、より高い精度で判別が可能となることを示した(発表論文)

(6)大規模代謝プロファイルテスト(血液検査)データを用いて、過去18年間に渡る北海道の乳牛約5万頭の状態の推移を調べた。獣医畜産分野の網羅的大規模データ解析の一例で、この分野のトップジャーナルに採択された(発表論文)

(7)子牛の分娩事故のリスク要因を明らかにするために混合効果ロジスティック回帰を適用した。北海道における大規模データ(120万件の事故データ)の解析例で、個体レベルのデータを説明変数として、農家レベルの情報を混合効果項としてモデルに入れた。獣医学分野のトップジャーナルに採択された(発表論文)

(8)獣医畜産分野における分子～個体レベルの研究として、他にも、イヌのゲノム解析(解析を継続中、学会発表)、結核菌ゲノムに対する統計的仮説検定の適用(発表論文)、牛の繁殖に関わる分子レベルのデータへの回帰モデルの適用(発表論文、)を行った。

(9)畜産分野における多種データの統計解析として、長尺農作物(長芋)切断を効率化するための統計モデルを提案した。このモデルにより、従来は手作業で行われていた農作物切断作業を自動化(かつ高精度化)することが可能になった。この分野の学術雑誌への論文掲載に向けて準備中である(論文投稿中1件、論文投稿準備中1件、学会発表、産業財産権)

#### 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10件)

Mitsunori Kayano, Hidetoshi Matsui, Rui Yamaguchi, Seiya Imoto, Satoru Miyano, "Gene set differential analysis of time course expression profiles via sparse estimation in functional logistic model with application to time-dependent biomarker detection", *Biostatistics*, 17(2) 235-248, 2016, 査読有, doi: 10.1093/biostatistics/kxv037

Mitsunori Kayano, Mutsuyo Kadohira, Mark Steavenson, "Risk factors for stillbirths and mortality during the first 24 hours of life on dairy farms", *Preventive Veterinary Medicine*, 127(1), 50-55, 2016, 査読有 doi:10.1016/j.prevetmed.2016.03.012  
Hiroto Takahashi, Shingo Haneda, Mitsunori Kayano, Motozumi Matsui, "Differences in progesterone concentrations and mRNA expressions of progesterone receptors in bovine endometrial tissue between the uterine horns ipsilateral and contralateral to the corpus luteum", *Journal of Veterinary Medical Science*, 印刷中, 査読有, doi:10.1292/jvms.15-0366

Mitsunori Kayano, Tomoko Kataoka, "Screening for ketosis using multiple logistic regression based on milk yield and composition", *Journal of Veterinary Medical Science*, 77(11), 1473-1478, 2015, 査読有, doi:10.1292/jvms.14-0691

Mitsunori Kayano, Katsuya Kida, "Identifying alterations in metabolic profiles of dairy cows over the past two decades in Japan using principal component analysis", *Journal of Dairy Science*, 98(12), 8764-8774, 2015, 査読有, doi:10.3168/jds.2015-9791

Kayo Okumura, 他4名(計5名中4番目), "Construction of a virtual *Mycobacterium tuberculosis* consensus genome and its application to data from a next generation sequencer", *BMC Genomics*, 16, 218, 2015, 査読有, doi:10.1186/s12864-015-1368-9

Miura R., Haneda S., Kayano M., Matsui M., Development of the first follicular wave dominant follicle on the ovary ipsilateral to the corpus luteum is associated with decreased conception rate in dairy cattle, *Journal of Dairy Science*, 98(1): 318-321, 2015.

Mitsunori Kayano, Motoki Shiga, Hiroshi Mamitsuka, "Detecting differentially coexpressed genes from labeled expression data: a brief review", *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 11(1), 154-167, 2014, 査読有,

doi: 10.1109/TCBB.2013.2297921  
茅野光範, 新飯田俊平, "インフォマテイクス解析によるバイオマーカーの検出", *アンチエイジング医学*, 10(1),

2014, 査読無,  
[http://www.m-review.co.jp/magazine/detail/J0038\\_1001](http://www.m-review.co.jp/magazine/detail/J0038_1001)  
Mitsunori Kayano, Seiya Imoto, Rui Yamaguchi, Satoru Miyano,  
“Multi-omics approach for estimating metabolic networks using low-order partial correlation”,  
*Journal of Computational Biology*,  
20(8), 571-582, 2013, 査読有,  
doi: 10.1089/cmb.2013.0043

〔学会発表〕(計 7 件)

Kayano, M., Hirayama, A., Sumi, Y., Bundo, M., Sakurai, T., Soga, T., Niida, S. and Takikawa, O., Blood biomarkers for Alzheimer's disease revealed by capillary electrophoresis/mass spectrometry (CE/MS)-based metabolomics, *Metabolomics 2014*, 鶴岡ワシントンホテル(山形県鶴岡市), 2014 年 6 月 25 日

根本誓哉、大嶋なつみ、茅野光範、高木しのぶ、小林菜々恵、横澤幸美、姜興起、鈴木宏志、性格関連遺伝子多型のロジスティック回帰分析による盲導犬適性予測, 第 6 回日本身体障害者補助犬学会学術大会, 日本大学(千葉県松戸市), 2013 年 10 月 26 日.

千葉大基、八谷満、姜興起、茅野光範、ナガイモの種イモ同時切断技術の開発、農業食料工学会第 72 回年次大会、帯広畜産大学(北海道帯広市), 2013 年 9 月 11 日.

Kayano, M., Kawano, S., Efficiently discriminating typical time series patterns in omics data based on functional t-test, *Joint Conference on Informatics in Biology, Medicine and Pharmacology*, Tower Hall Funabori (東京都江戸川区), 2012 年 10 月 3 日.

茅野光範\*, 川野秀一, 時系列オミックスデータのための関数 t-検定, 第 7 回メタボロームシンポジウム, 慶応義塾大学鶴岡メタボロームキャンパス(山形県鶴岡市), 2012 年 10 月 11 日.

姜興起, 茅野光範, 八谷満, 千葉大基, 種イモ切断自動化の設計を支援する長芋形状推定方法の開発, 2012 年度統計関連学会連合大会, 北海道大学(北海道札幌市), 2012 年 9 月 11 日.

茅野光範, 松井秀俊, 山口類, 井元清哉, 宮野悟, 関数 t-検定と関数ロジスティック回帰の L1 正則化による遺伝子選択, 2012 年度統計関連学会連合大会, 北海道大学(北海道札幌市), 2012 年 9 月 11 日.

〔産業財産権〕  
出願状況(計 2 件)

名称: 軽度認知症を検査する方法  
発明者: 新飯田俊平、滝川修、  
檜垣小百合、松本健治、茅野光範  
種類: 特許  
番号: 特許願 2015-173901  
出願年月日: 平成 27 年 9 月 3 日  
国内外の別: 国内

名称: 長尺農作物の切断調整装置  
発明者: 八谷満、千葉大基、姜興起、  
茅野光範、五十嵐正和  
種類: 特許  
番号: 特許願 2013-022921、  
特許開 2014-151394  
出願年月日: 平成 27 年 2 月 8 日  
国内外の別: 国内

〔その他〕

ホームページ等

研究代表者ホームページ

<http://www.obihiro.ac.jp/~kayano/index.html>

6. 研究組織

(1) 研究代表者

茅野 光範 (KAYANO, Mitsunori)

帯広畜産大学・畜産学部・講師

研究者番号: 20590095