

平成 21 年 6 月 26 日現在

研究種目：基盤研究（B）
 研究期間：2007-2008
 課題番号：19300100
 研究課題名（和文） 臨床オミックスデータに基づく疾患メカニズム関連因子探索のための統計的方法論の開発
 研究課題名（英文） Development of statistical methodology for detecting factors of disease mechanism based on clinical omics data
 研究代表者
 松浦 正明 (MATSUURA MASAOKI)
 財団法人癌研究会 癌研究所 がんゲノム研究部・部長
 研究者番号：40173794

研究成果の概要：

マイクロアレイデータと臨床情報を用いて疾患のメカニズムに関連する候補遺伝子を探索するために、生物学的情報を加味した遺伝子評価法を開発し、複数の予測モデルを組み合わせた論理和予測モデルを用いて予測性能の検証も行った。また、各患者で 50 万箇所以上におよぶ塩基多型（SNP）データに対して、特定の複数 SNP を組合わせた場合にのみ、表現型と関連する SNP の組を探索するための高速な統計的手法の開発に成功し、特許申請を行った。

交付額

（金額単位：円）

	直接経費	間接経費	合計
19 年度	7,900,000	2,370,000	10,270,000
20 年度	6,700,000	2,010,000	8,710,000
年度			
年度			
年度			
総 計	14,600,000	4,380,000	18,980,000

研究分野：総合領域

科研費の分科・細目：情報学・統計科学

キーワード：統計数学、ゲノム、マイクロアレイ、プロテオーム、生体生命情報学

1. 研究開始当初の背景

SNP 情報を取り扱うゲノミクス、マイクロアレイデータを取り扱うトランスクリプトミクスや、タンパク発現データを取り扱うプロテオミクスなどから得られる網羅的分子情報（オミックスデータ）に基づいて、統計学的に相関の強いバイオマーカーを探索することが世界的に進められてきた。しかしながら、従来の統計学的手法ではバイオマーカーの探索や、それを用いた予測システムの確立、さらには疾患メカニズムに関連する因子の探索を行うことは困難であった。疾患のメカニズムに関連する遺伝子を同定すること

ができれば、ゲノム創薬や新たなテーラーメイド医療の確立に寄与することができる。このような研究はバイオインフォマティクスの分野で開発が進められているが、十分な進展は認められていなかった。疾患メカニズムに関連する遺伝子などを統計学的手法で探索するには、統計学分野の研究者だけでなく、医学・生物学分野の研究者との共同研究体制を整える必要があり、そのような研究組織による研究が期待されていた。

2. 研究の目的

癌研究で蓄積してきた癌に関する貴重

なオミックスデータ解析の経験を十分活用し、医学・生物学分野の研究者との共同研究を進めることにより、ゲノムや遺伝子発現情報などデータを用いて、抗癌剤の治療効果や副作用などを規定している疾患のメカニズムに関連する因子を探索するための新しい解析システムを構築することを目的とする。

具体的には、以下の3点に対する統計的方法論の開発を行う。

(1)質量分析計によって得られた網羅的タンパク発現データを用いて、抗がん剤の治療効果に関連するタンパクの同定を行う。

(2)マイクロアレイデータによって得られた網羅的遺伝子発現データを用いて、癌患者の予後を規定するメカニズムに関連する遺伝子の同定を行う。

(3)50万以上 SNP データを用いて、複数の遺伝子が同時に表現型に影響を与えるエピスタシス効果を有する遺伝子の同定を行う。

3. 研究の方法

(1)網羅的タンパク発現データを用いた抗がん剤の治療効果に関連するタンパクの同定法の開発では、タンパク発現スペクトルの中から抗がん剤治療効果に関連するタンパク（ピーク）を検出する際に、Fushiki et al.(2006)が開発したコモンピーク法を改良し、治療効果ラベルごとにコモンピークを抽出する方法を検討した。タンパク発現情報を解析データに変換する際には、タンパク発現強度を示す intensity に対して、連続量と2値化した両方の場合のデータを作成し、それぞれの予測結果の良好度を比較した。なお、治療効果を予測する方法として、機械学習理論に基づく AdaBoost 法などを検討した。

(2)マイクロアレイによる遺伝子発現データを用いて、癌患者の予後を規定するメカニズムに関連する遺伝子を検出する方法として、Web で広く公開されている生物学的情報である Gene Ontology データを用いた。表現型と強く関連する生物学的機能を示すタームを統計学的に抽出するためモンテカルロ有意確率をスコアとして算出し、各遺伝子を評価するための手法を確立した。

(3)膨大な数の SNP データの中から、複数の遺伝子が同時に表現型に影響を与えるエピスタシス効果を有する遺伝子の同定する方法では、高速計算を可能にするため、必要最小限度の演算回数を実現するためのオッズ比統計量を用いたスコアを開発した。

実際の解析で用いるデータとしては、研究代表者が所属する財団法人癌研究会ゲノムセンターで収集したデータならびに一般に公開されているオープンデータを用いた。なお、SNP 情報の解析においては、インフォー

ムドコンセントが取得、倫理委員会で認められたデータを使用した。

4. 研究成果

(1)平成19年度は、以前より進めてきた研究を進展させ、質量分析計から得られるタンパク発現データを用いて、乳がんに対する抗がん剤パクリタキセル単剤投与の治療効果に関連するタンパクを同定し、治療効果予測を行うための方法論を確立した。表現型と関連する重要なタンパク候補の選定法として、独自に改良したコモンピーク法を適用し、抗がん剤の治療効果に関連するタンパク発現情報を集約した。各患者のタンパク発現を2値化することにより生成されたデータに対してAdaBoost法を適用することにより、抗がん剤の臨床効果予測を行なうことに成功した。本方法は recursive SVM よりも良い予測性能を示し、これらの結果については、Cancer Informatics誌に投稿し論文採択された。

(2)平成19年度の第2段階の研究として、臨床オミックスデータの中でも世界的にも多くの施設でデータが取得されているマイクロアレイによるトランスクリプトームデータに対し、疾病関連因子の探索理論の構築を行った。

疾病関連因子の探索理論の構築においては、癌を多因子遺伝子疾患のモデルとし、疾患制御因子を探索・同定するためのプロセスとして、

パスウェイの最少単位としての2変量の交互作用項を考慮したロジスティック回帰モデルの中から、2段階機械学習法により表現型との関連が強いモデルを絞り込み、次に

表現型に強く関連している生物学的機能を探索するための手法として、Webで広く公開されている生物学的情報であるGene Ontologyデータを用い、表現型と強く関連する生物学的機能を示すタームを統計学的に抽出するためモンテカルロ有意確率を算出し、遺伝子抽出のためのスコアとして検討した。さらに

開発したシステムに対し臨床情報を有する実データを用いて解析・検証を行った。

神経芽細胞腫のデータを利用し、患者の予後予測とメカニズムの推測を試みたところ、モデルのパリテーション成績である誤判別率は0.12の値を示し、従来法よりも小数の遺伝子を用いたモデルであっても高精度な予測が可能であった。また予後を規定するメカニズムの推測を試みた結果、細胞増殖、Gタンパクシグナル、解糖系、神経生

理プロセス、カテコールアミン代謝等が予後に影響するものと推測された。本結果を2008年のバイオコンピューティングの国際シンポジウム(ハワイ)で発表し、さらに論文発表を行った。

(3)平成20年度は、臨床オミックスデータの中のゲノムのSNPデータ、マイクロアレイデータ、プロテオームデータを取り扱い、個別化医療の確立に向けて重要な課題である治療効果・副作用予測の問題に関連させて研究を進めた。

疾患の易罹患性や薬剤の感受性に関連するゲノムの変異としての一塩基多型であるSNPデータに基づき、抗がん剤の副作用と関連する遺伝子の同定結果に関してアメリカ癌学会で報告を行った。

さらに、50万種のSNPデータの中から、複数の遺伝子が同時に表現型に影響を与えるエピスタシス効果を有する遺伝子の同定法の開発を行った。表現型のモデルの一つとして、乳がんの抗がん剤であるパクリタキセル単剤投与に対する副作用の原因遺伝子検出のための研究を行った。SNPデータは、癌研病院においてインフォームドコンセントが取れた患者に対して、一人当たり90万SNPものジェノタイプを調べたゲノムワイドなデータを使用した。ゲノムワイドな遺伝子変異データの組み合わせは膨大な数に昇り、これらの中から、複数の遺伝子が表現型の原因となる場合のエピスタシス効果を探索するための高速な方法論を開発した。本研究に関しては、特許申請を行なった後に、多重比較に関する国際シンポジウムで報告を行った。

一方、マイクロアレイデータを用いた解析では、培養細胞株に各種化合物を作用させて変化する遺伝子をデータベース化したコネクティビティマップを利用した解析結果についても癌学会で報告した。また、イメージング・マス・スペクトロメトリーとして急速に進展しているプロテオームデータについても検討を行い、質量分析計からのプロテオームデータ解析の統計学的問題点と打開策案として、昨年度本研究で検討したコモンピーク法に関して統計学会で報告した。

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 6件)

松浦正明、ファーマコゲノミクス、バイオインフォマティクス、バイオスタティクス臨床医薬 23, 295-325, (2007)

査読無し

Enomoto Y, Ushijima M, Miyata S, Matsuura M, Ohtaki M. Inference on biological mechanisms using an integrated phenotype prediction model. Hiroshima Journal of Medical Sciences 57, 7-15, (2008). 査読有り

Oku Y, Matsuura M (15人中10番目), Miyata S (11番目) et al. Identification of the molecular mechanisms for dedifferentiation at the invasion front of colorectal cancer by a gene expression analysis. Clinical Cancer Research 22, 7215-7222, (2008). 査読有り

Osako T, Ushijima M (8人中2番目), Matsuura M (7番目) et al. Predictive factors for efficacy of capecitabine in heavily pretreated patients with metastatic breast cancer. Cancer Chemother Pharmacol. 63, 865-871, (2008). 査読有り

Isomura M (20人中1番目) et al. IL12RB2 gene and ABCA1 gene are associated with susceptibility to radiation dermatitis. Clinical Cancer Research, 14, 6683-6689, (2008). 査読有り

Fujisawa H. and Eguchi S. Robust parameter estimation with a small bias against heavy contamination. Journal of Multivariate Analysis, 99, 2053-2081, (2008). 査読有り

[学会発表](計 15件)

牛嶋大, 宮田敏, 松浦正明. 質量分析計データを用いた判別解析とバイオマーカー探索. 統計関連学会連合大会 2007年9月7日. 神戸大学.

藤澤洋徳, 堀内陽子, 春島嘉章, 江口真透, 望月孝子, 坂口隆之, 倉田のり. SNPを考慮した遺伝子発現データ解析. 統計関連学会連合大会. 2007年9月7日. 神戸大学.

栗木哲, 春島嘉章, 藤澤洋徳, 倉田のり. 遺伝子座間の相互作用による生殖的隔離障壁の検出と多重性調整. 統計関連学会連合大会. 2007年9月7日. 神戸大学.

松浦正明. Bioinformatics for cancer research. 第65回日本癌学会学術総会. 2007年10月4日. 横浜パシフィコ.

牛嶋大, 宮田敏, 長崎光一, 三木義男, 野田哲生, 松浦正明. Omicsデータを用いた治療効果予測の高精度化. 第65回日本癌学会学術総会. 2007年10月4日. 横浜パシフィコ.

榎本友美, 宮田敏, 牛嶋大, 松浦正明. 多

変量解析手法を用いた表現型判定システムの構築法及び表現型規定メカニズムの推測法の検討. 第 65 回日本癌学会学術総会. 2007 年 10 月 4 日. 横浜パシフィコ.

宮田 敏、松浦正明. 多変量確率分布を用いたエクソナーレイデータの解析. 第 65 回日本癌学会学術総会. 2007 年 10 月 4 日. 横浜パシフィコ.

Enomoto Y, Ushijima M, Miyata S, Matsuura M. Inference on biological mechanisms using an integrated phenotype prediction model. The Pacific Symposium on Biocomputing (PSB) 2008, 2008.1.4., Hawaii (USA)

Fujisawa H, Horiuchi Y, Harushima Y, Eguchi S, Mochizuki T, Sakaguchi T, Kurata N. A unified method for detecting single feature polymorphisms and gene expression level differences. Pacific Symposium on Biocomputing. 2008.1.6. Hawaii (USA).

Isomura M, Matsuura M, Noda T, Miki Y. Genome-wide survey of genetic polymorphisms associated with cytotoxicity of paclitaxel using HapMap cell lines derived from Japanese population. Annual meeting of American Association for Cancer Research. 2008.4.13. San Diego (USA).

Isomura M, Matsuura M, Usijima M, Miyata S., et al. Genome-wide survey of genetic polymorphisms associated with cytotoxicity of paclitaxel using HapMap cell lines derived from Japanese population. Second JCA-AACR Special Joint Conference. The Latest Advances in Breast Cancer Research. 2008.6.14. Awaji Yumebutai International Conference Center, (Japan)

牛嶋大、宮田敏、松浦正明. プロテオームデータ解析の現状および統計的諸問題. 統計関連学会連合大会. 2008 年 9 月 9 日. 慶應義塾大学.

宮田敏、牛嶋大、松浦正明. コネクティブイティマップを用いた潜在的に有効な抗がん剤の遺伝子発現印章に基づく識別. 第 67 回日本癌学会総会 2008 年 10 月 29 日. 名古屋国際会議場.

磯村 実、松浦正明、野田哲生、三木義男. Genome-wide survey of genetic polymorphisms associated with cytotoxicity of paclitaxel using HapMap cell lines 第 67 回日本癌学会総会 2008 年 10 月 29 日. 名古屋国際会議場.

Ushijima M, Isomura M, Miki Y, Matsuura M. A simple method for detecting epistasis in genome-wide association

studies. 2009 年 3 月 27 日 Tokyo University of Science (Japan)

〔図書〕(計 1 件)

牛嶋 大、宮田 敏、松浦正明、

株式会社メディカルドゥ、東京

マイクロアレイデータ解析法, 遺伝子医学 MOOK10 号「DNA チップ/マイクロアレイ臨床応用の実際 - 基礎・最新技術、臨床・創薬応用への実際から今後の展開・問題まで - 」 2008 年、401-408.

〔産業財産権〕

出願状況(計 1 件)

名称: 網羅的ゲノムワイド SNP 情報に基づくエピスタシス効果の探索アルゴリズム

発明者: 松浦正明、牛嶋大、磯村実、他

権利者: 財団法人癌研究会、
NEC ソフト株式会社

種類: 特許

番号: 特願 2009-070753

出願年月日: 2009 年 3 月 23 日

国内外の別: 国内

6. 研究組織

(1) 研究代表者

松浦 正明 (MATSUURA MASAOKI)

財団法人癌研究会 癌研究所

がんゲノム研究部・部長

研究者番号: 40173794

(2) 研究分担者

江口 真透 (EGUCHI SHINTO)

統計数理研究所・数理推測研究系・教授

研究者番号: 10168776

藤澤 洋徳 (FUJISAWA HIRONORI)

統計数理研究所・数理推測研究系・準教授

研究者番号: 00301177

星川 裕 (HOSHIKAWA YUTAKA)

財団法人癌研究会・癌研究所がんゲノム研究部・研究員

研究者番号: 80280626

磯村 実 (ISOMURA MINORU)

財団法人癌研究会・ゲノムセンター・研究員

研究者番号: 40272497

宮田 敏 (MIYATA SATOSHI)

財団法人癌研究会・ゲノムセンター・研究員

研究者番号: 60360343

牛嶋 大 (USHIJIMA MASARU)

財団法人癌研究会・ゲノムセンター・研究員

研究者番号 : 60328565

(3) 連携研究者