

機関番号：55502

研究種目：若手研究 (B)

研究期間：2009～2010

課題番号：21700336

研究課題名 (和文) 癌の医療方針決定統合システムの開発に関する研究

研究課題名 (英文) **Research of the development of a medical decision-making system for cancer**

研究代表者

北風 裕教 (KITAKAZE HIRONORI)

大島商船高等専門学校・商船学科・講師

研究者番号：70342558

研究成果の概要 (和文)：セルアレイから取得される遺伝子産物の細胞内分布情報を、可視化アルゴリズムを用いることで、がんを分子レベルで特徴化・分類できる情報処理システムの開発を行った。システムは分類データベース部、可視化部、分類比較部から構成される。医師が直感的に細胞内分布情報を把握できる模式化を行ったことが情報圧縮につながった。また、この情報を利用することで各がん細胞とタンパク質の比較表を作成でき、この比較表から分類を可能にした。

研究成果の概要 (英文)：By using a visualization algorithm based on the information of sub cellular distribution of gene product, obtained from the cell array, a data processing system which can classify characteristics of cancer at molecular level is developed. The system is composed of database part, visualization part and comparison part for classification. Making the information of the sub cellular distribution schematic by which doctors can grasp that information intuitively, led to compression of the information. Then, by using this information, a comparison table of each cancer cell and protein could be created and we made classification of cancer possible by this comparison table.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2009年度	1,700,000	510,000	2,210,000
2010年度	800,000	240,000	1,040,000
総計	2,500,000	750,000	3,250,000

研究分野：総合領域

科研費の分科・細目：情報学・生体生命情報学

キーワード：ニューロインフォマティクス、バイオインフォマティクス、システム生物学

## 1. 研究開始当初の背景

(1) 遺伝子、タンパク質形成、代謝、シグナル伝達などの分子間ネットワークにおいて、シミュレーションツールを用いて脆弱部位

を特定する研究が進められているが、ネットワークにフィードバックを持つ場合、脆弱部位を推定することは困難であり、知能情報を用いた新しい推定法の提案が重要な

課題となっている。我々の研究グループでは、がん細胞の脆弱部位の自動推定法に着手しているが、実システムとして十分利用できるまでには至ってはいなかった。がん細胞の生命パスウェイからがんの重要部位を推定することができるシステム開発が求められた。

(2) がんの分類は、治療方法を決定する点から、重要である。現在までの手法では、病理部門において、顕微鏡を用いて目視で行われている。がんの分類ができない存在も知られており、目視による分類には限界があった。がんの発生は、遺伝子やその産物である分子の異常によるものである。がんを分子レベルで分類できる手法があれば、より正確ながん分類が可能になり、医師が納得できる表現で可視化できれば、有効な治療方法の決定につながる。しかし、未だに分子レベルの有効な分類方法および可視化法は確立されていない。

(3) がん細胞情報はレーザスキャニングサイトメータ (LSC) を利用して取得され、分析システムによって分類が行われるが、分析用の細胞数が数百万細胞となるので、容易に結果を統合し、医療支援を行う必要がある。しかし、統合できるデータベースの開発はおこなわれてはいない。また、ネットワークを介して遠隔地でも作業を行う事は実現できていない。

## 2. 研究の目的

(1) 分子間ネットワークにおける分子状態の推移の様子を分析できるハイブリッド関数ペトリネット (HFPN) を用いて、分子間の反応経路 (生命パスウェイ) において細胞システム維持に影響を及ぼす重要な経路 (生命パスウェイ重要部位) を自動で推定できる手法を提案し、推定率の向上に努める。

(2) 遺伝子産物の情報を多次的に同時に取得できるセルアレイの技術を利用し、セルアレイからの多次元多量なデータから情報科学的手法によって規則性を導き出し、さらにそのデータを利用したシステム生物学的手法を開発することで、がんの分類に役立てる。

(3) 取得したがん細胞情報及び分類分析結果、可視化の結果などを統合し、Web を介して遠隔地でも一連の作業を行う事ができる医療方針支援を行うための分類データベースの構築を行う。

## 3. 研究の方法

(1) 細胞内の様々な反応経路 (生命パスウェイ) を HFPN 記述によりモデリングできるツ

ール Cell Illustrator (CI) を用いて、生命パスウェイ重要部位を自動で推定するための手法について調べる。生命パスウェイに分岐やフィードバックループがある場合の生命パスウェイ重要部位の推定が非常に困難であることから、マウスの概日リズムを利用して、推定率を得るためのシミュレーション実験を行うことでシステムの有効性について検討した。

従来手法では、ユーザがシミュレーションツール (CI) を利用して HFPN 表記により生命パスウェイを作成した後、生物実験におけるロックアウトの作業を CI 画面上でアークの削除により実現する。この作業は CI において手作業で行い、続けて HFPN の実行処理を行う。HFPN の実行は計算機によって自動的に行われ、トークンの時間変化がグラフとして出力される。この出力結果から、ユーザは目視により重要部位の判定を行う。ここでの判定は、ユーザがこれまでの経験による判断によって決定される。この作業を全てのアークにおいて繰り返す事により、生命パスウェイにおける重要部位を確認することができる。この手法は、ネットワーク全体の中から重要部位を確認することができるが、一連の作業に要するシミュレーション時間と、重要部位の確認に要する時間は生命パスウェイが大きくなるに伴い増加するため、逐次実行を繰り返す必要の体自動推定法が求められる。

そこで本研究では次に占める RNN を用いた重要部位の自動推定手法を提案し、これらの問題点の改善に努める。

(2) スライドグラスに直径 2mm の円 (ウェル) を 50 個整列させて印刷したセルアレイへがん細胞をセットし、LSC を用いてタンパク質情報を抽出する。LSC は、スライド上の細胞や連続した粒子の範囲に強力な単色のレーザを光ビームで照射し、蛍光色素を励起して細胞から光を散乱させるシステムであり、前方散乱と細胞のレーザ冷機によって発光された蛍光はフォトセンターより検出でき、電子アナログ信号に変換され、細胞面積、周囲の長さ、細胞内のタンパク質量やその凝集度などのデータを取得することができる。得られた結果から、細胞面積とタンパク質量、その凝集度情報に着目し、散布図を作成することで、タンパク質の分布の傾きを知る事ができ、増減しつつも細胞内にまんべんなく分散している場合や、量は多くないが細胞内の特定の箇所分布しているなどの特徴を、抽出することができる。

本研究では、いくつかのタンパク質のセルアレイを準備し、得られる散布図パターンの組み合わせによって、がん細胞を特徴化して分類する手法を開発する。また、これらの結

果から、タンパク質の種類と細胞種類の表を作成して、自己組織化マップ (SOM) による自動分類学習を行う。パターン判別は、その分布状況による直感的な結果として得ることができる。また、これ以外の特徴的なパターンを見出すために、取得した散布図へ細胞面積・外周情報を取り入れ、濃度パターンの統一化を行うことができる可視化支援システムの開発を試みる。

(3) LSC から取得したデータはがん細胞の配置や各セルのデータ量が統一されていない問題、レーザ照射時間が安定せず利用できないセルがある問題などあり、容易に利用できるデータへ変換を行う必要がある。また、分析を行うためには、データ準備やデータ加工、事前にプログラム読み込みを行う必要があり、準備時間に1ヶ月以上かかる。その問題を解決するために、がん細胞データから必要な情報を容易に準備する事ができるデータベース(以下DB)システムの構築を行う。その後、実際にがん細胞データを全てDBシステムに蓄積し不具合等の確認を行う。さらに、ネットワークを利用して離れた場所からでも研究者がデータを参照でき、暗号化することでセキュアな通信ができるWebサーバの構築とサーバサイドプログラミングを行う。

#### 4. 研究成果

(1) 多くの処理時間を要するHFPNから学習によるシステム形成後には、想起処理によりそれ程の処理時間を要さないニューラルネットワーク(NN)を用いて重要部位を推定する。HFPNではトークンの時間変化情報が重要部位の判定に必要となる。そこで、NNには時間を記憶可能なリカレントニューラルネットワーク(RNN)を用いる。HFPNからRNNへネットワークの変換を行い、通時的逆伝播法(BPTT法)の学習により、自動で連続的に重要部位を推定するシステムを作成した。この手法は、ネットワークの移行のために、RNN学習時間を要するが、連続的にアークの削除が行えるため、推定結果を得るまでの時間を改善できた。

生命モデル(マウスの概日リズム)を用いた反応推定実験結果を表1に示す。

表1 RNNによる重要部位の推定結果

	重要部位	重要ではない部位	合計
HFPNの基準指標数	23	18	41
RNN手法の的中率	87% (20)	77.8% (14)	82.9% (34)
RNN手法の誤り率	13.0% (3)	22.2% (4)	17.1% (7)

自動推定では、約83%の的中率を得ることができ、がん細胞の生命パスウェイの重要部位を推定できる可能性を示すことができた。

(2) がん細胞の分類を自己組織化マップ(SOM)により行った。扁平上皮癌(A1~A5)、大腸癌:腺癌(B1~B5)、軟骨肉腫(C1~C5)、乳癌(D1~D5)、肺癌(E1~E5)、子宮頸癌:扁平上皮癌(F1~F5)、肺癌:扁平上皮癌(G1~G5)、乳癌(H1~H5)、肺癌:腺癌/前立腺がん(腺癌)(I1~I5)、副腎癌(J1~J5)の10種類のがん細胞のデータをSOMを用いて学習した。タンパク質p27の発現における結果を図1に示す。また、Her2の発現における結果を図2に示す。

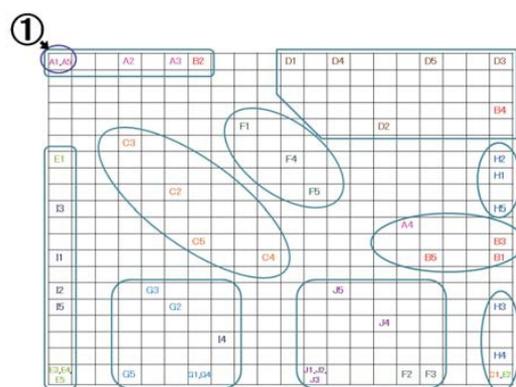


図1 タンパク質 p27 の結果 (上)

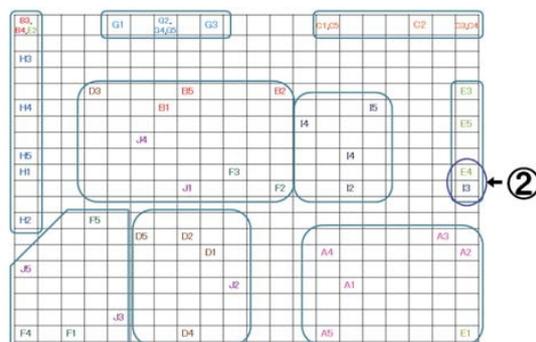


図2 タンパク質 Her2 の結果 (下)

SOMは、各データを類似性に応じて分離するため、類似しているものはより近くに、類似していないものはより遠くに位置するという特性を持っている。そのため、図1で示した①では、同じがん細胞が近くに位置している。また、図2で示した②では、異なるがん細胞が隣接して位置している。つまり、同じがん細胞であっても異なるがん細胞であっても類似していれば近くに位置する。

図1で示した①のタンパク質p27における扁平上皮癌(A1)の散布図を図3の上図に、扁平上皮癌(A5)の散布図を図3の下図に示

す。また、図2で示した②のタンパク質 Her2 における肺癌(E4)の散布図を図4の上図に、肺癌[腺癌]／前立腺癌[腺癌](I3)の散布図を図4の下図に示す。

図3と図4の各々の散布図を比較すると、どちらのタンパク質も類似していることから、図3においては同じ癌で同じタンパク質であることから類似した結果であることは明らかである。しかし、図4では、異なるがん細胞であったにも関わらず散布図による特徴が類似しており、SOM を用いてがん細胞を分類することができる可能性を示した。

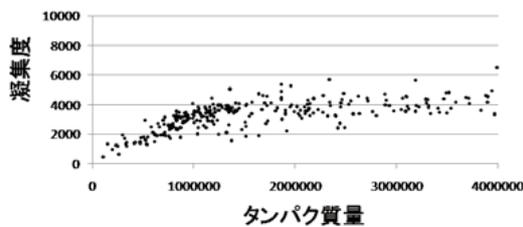
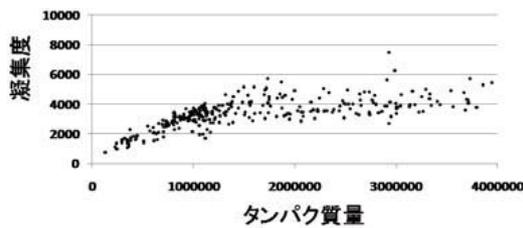


図3 タンパク質p27 における扁平上皮癌 (A1) と扁平上皮癌 (A5) の散布図

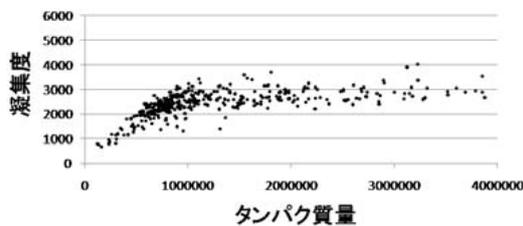
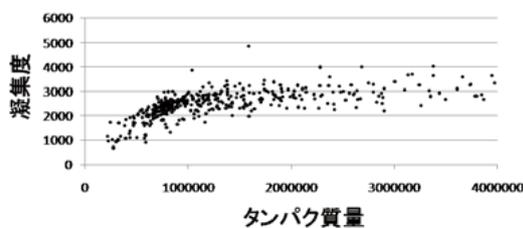


図4 タンパク質Her2における肺癌 (E4) と肺癌[腺癌]／前立腺癌[腺癌](I3)の散布図

SOM アルゴリズムは非線形学習であるため、初期値や条件によっては分類結果が異なる。また、大まかな分類は行うことはできるが、

どのような条件が起因したのか医師に説明をすることができない。そこで、医師が直感的にがん細胞情報を判断でき、分類作業を円滑に行うための可視化支援システムの開発を行った。

図5に可視化支援システムからの出力結果を示す。可視化は、LSC から得られた散布図と可視化支援システムから得られたモデル図を同時に出力することによって実現する。

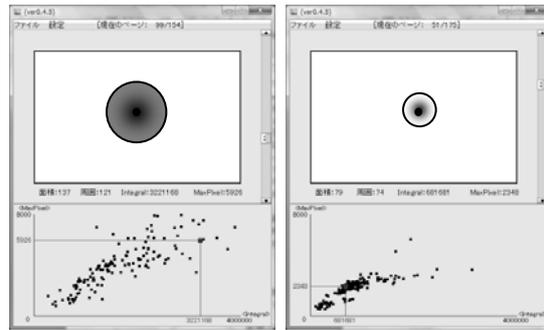


図5 タンパク質:Her2 における比較 (左: LK-2 癌, 右: PC-3 癌)

がん細胞は、通常の細胞に比べて変形した形状をしており、統一されてはいないが、可視化では全てを統一して円で表現する。LSC から抽出したがん細胞の面積・外周情報は、可視化システムでは、円の大きさで表し、タンパク質の凝集度は、円の中心の色で表現する。色が明るい色であれば、凝集度が高く、暗い色であれば凝集度が低いことを意味する。また、タンパク質量は、円内部の色の広がり度合いで表現する。タンパク質 Her2 において LK-2 癌、PC-3 癌の例を示すが、各々のがん細胞の特徴を直感的に理解できる。同じタンパクでも癌により細胞のサイズに異なりがあり、タンパク質の分布においても違いがあることも、このシステムを利用すれば容易に判断できるようになった。

医師による利用効果について現在調査中である。

(3) Web ブラウザ上でがん細胞とタンパク質の名前の登録し、がん細胞のデータファイルをサーバにアップロードをすることで、サーバ側ではファイルのデータとそのファイル情報から SQL 文を自動生成して DB へ登録を行うことができるようになる。この登録された DB からは、条件(タンパク質の名前、がん細胞の名前、採集日の期間)に応じたデータを抽出可能である。そのデータは csv ファイルとしてダウンロードでき、データを有効活用できる。

がん細胞分類データベースシステムを利用してがん細胞を登録する画面を図6に示す。

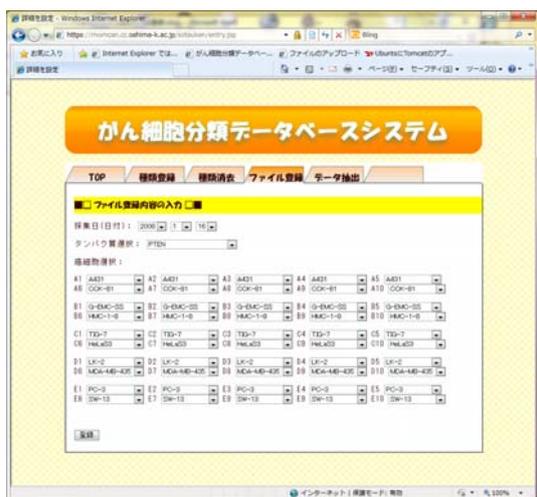


図6 データベースへの登録画面

従来はデータを開いて必要なデータを取り出す作業を、実験の都度行っていたが、このシステムの利用でデータをDBへ入れておくと、離れた場所にいる人でも登録データを共有でき、必要なデータを容易に取り出すことが可能で、分析効率はかなり向上した。データベースの抽出実験の結果を表2に示す。

表2 データベースのデータ抽出実験の結果

145,949個のデータベース	時間
1種類のがん細胞 1種類のタンパク質データ (2,817個のデータの抽出)	約5秒
1種類のがん細胞 3種類のタンパク質データ (6,547個のデータの抽出)	約18秒

## 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1件)

- ① 北風裕教, Cell Illustratorを用いた細胞システム維持に影響する生命パスウェイ重要部位の推定、Informatics World、査読無、No.17、2009、pp.4-7

[学会発表] (計4件)

- ① 北風裕教、井川翔平、伊藤剛、池田信彦、松野浩嗣、がん細胞内のタンパク質情報を用いた可視化による分析支援法、電気・情報関連学会中国支部第61回連合大会、2010、p.117

- ② 下村佳史、池田信彦、北風裕教、松野浩嗣、

タンパク質の量と凝集度に基づくSOMによるがん細胞分類システムの動作検証、電気・情報関連学会中国支部第60回連合大会、2009、pp.561-562

- ③ 立花麻耶、北風裕教、池田信彦、松野浩嗣、タンパク質情報を用いたSOMによるがん細胞分類の実験と評価、電気・情報関連学会中国支部第60回連合大会、2009、pp.563-564

- ④ Hironori Kitakaze、Prediction Method of Essential Points in a Biological Pathway for Cell System Stability by using Recurrent Neural Network、The 2009 International Workshop in "Computational and Integrative Biology"(CIB)、2009、p.90

[図書] (計0件)

[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

<http://www.oshima-k.ac.jp/kitakaze/index.php>

## 6. 研究組織

(1) 研究代表者

北風 裕教 (KITAKAZE HIRONORI)

大島商船高等専門学校・商船学科・講師

研究者番号：70342558

(2) 研究分担者

なし

(3) 連携研究者

なし