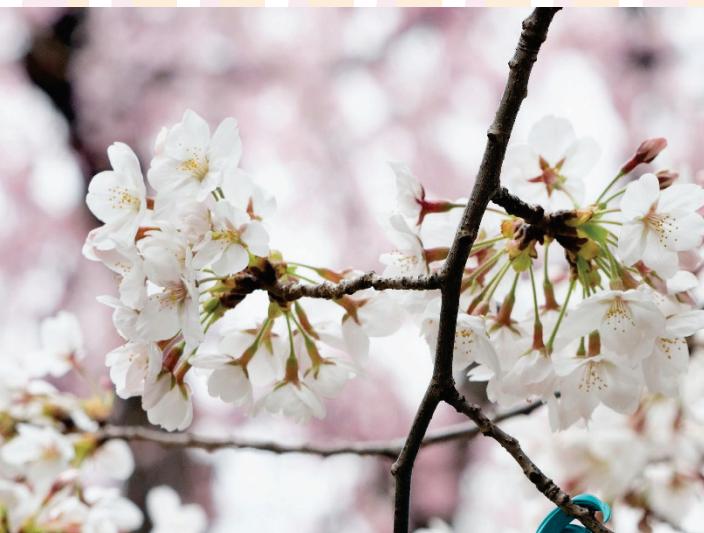




システム癌  
新次元



DIMENSIONAL SYSTEMS UNDERSTANDING  
CONQUERING CANCER THROUGH NEO-

がんシステムの新次元俯瞰と攻略

文部科学省科学研究費補助金 新学術領域研究

(研究領域提案型)(複合領域 : 4701)

Newsletter

No. 20

2019年4月

## 公募研究

# がんの統合的解明を目指した生体情報の階層的ネットワークに対する深層学習の応用

公募研究 A02-2-18 研究代表者

国立がん研究センター研究所がん分子修飾制御学分野 分野長 浜本 隆二



21世紀に迎えたポストゲノミクスの時代において、遺伝子やタンパク質に関するゲノムのみならず、トランスクリプトーム、プロテオーム、メタボロームなどの大量のオミックスデータが得られるようになり、多くの分子間ネットワークが階層的に織りなす生体情報の全容が少しづつ明らかとなっていました。がんを始めとする多くの疾患はこの複雑な生命システムの破綻を起点としており、がんの全体像を解明するためには、分子間のネットワーク構造が生み出す特性、生体情報の階層性を横断的に解析し、システムレベルでの深い理解が欠かせません。しかしながら、生命を構築する各階層の分子間ネットワークは、正負両方のフィードバックループや経路間のクロストーク、階層間相互作用などが複雑に入り組んだ本質的な複雑性を有し、そのモデリングは容易ではありません。また、次世代シークエンサーの登場によって加速的にデータが蓄積されていく一方で、ノイズの増加への対応、組み合わせによる仮説の増加による相対的な検出力の低下、マルチモーダル性を扱う解析の本質的な難しさなど、統計科学上の課題が浮き彫りとなっております。すなわち、システムの理解という観点から考えると、生体情報のネットワークが生み出す特性、階層性を横断的に解析し、全体像を把握することは未だに困難な課題であるといえます。

そこで、我々は「50年来の技術的ブレーカスルー」とも言われる深層学習を中核技術として用いることで、がんの生体時空間にわたるシステム的統合理解への橋渡し研究を現在行っております。人間や生物

の脳神経系にヒントを得た情報処理メカニズムであるニューラルネットワークを多層化した深層学習は、画期的な技術として機械学習の分野を超えて、その社会的なインパクトにおいても大きな注目を集めています。深層学習は画像や音声といった比較的高度な知的タスクで既に人間を上回る性能を示すに至り、生命情報処理の分野においてもその応用を目的にした世界的な研究開発競争が激化しております。申請者らは、平成28年度戦略的創造研究推進事業(CREST)に採択された研究課題「人工知能を用いた統合的ながん医療システムの開発」において、国立がん研究センターを中心に人工知能技術を用いた新しいがん医療システムの開発に先駆的に取り組み、既に内視鏡画像からの腫瘍自動診断などに成功しております。しかしながら、多くの生体情報のように、因子間のネットワーク構造そのものや、階層間の相互作用に対して直接的に深層学習を適応することは未だ困難な課題でありました。このような経緯に基づき、本新学術領域「システム癌新次元」では多層オミックスデータに対して深層学習を用いた発見的なアプローチを追及し、生体情報の階層的なネットワーク構造に内包される高次元空間内の相関関係を低次元に圧縮し、科学上意味のある結果を導き出すための方法論の構築に取り組んでおります。

我々は初めに、多層オミックスデータにおける各階層内の分子細胞生物学的な特性に応じた情報表現と、その潜在表現を抽出するための深層学習器の設計について取り組んでおります。ゲノム情報や分

子間ネットワークに対しては、概して高次元かつ疎(スパース)な情報表現が必要とされることが多いです。すなわち、モデルにおけるメモリ効率の高さや、生体分子をその機能を反映した潜在空間に上手く特徴付けるような低次元表現への符号化が必要であります。特に、自然言語処理や画像に対する深層学習では、「フランス」「パリ」「東京」=「日本」、「眼鏡を付けた男性」「眼鏡を付けていない男性」「眼鏡を付けていない女性」=「眼鏡を付けた女性」といったような、人間にも了解可能な意味空間の抽出が得られております。生体分子においても潜在空間における距離から機能的な類似性を推定することが出来れば、目的に応じた新規因子の発見の契機になる可能性があります。

また今後の研究で重要視している点は、多層オミックスデータにおける階層間の関係性のモデリングについての検討です。機械学習や深層学習を用いた多層オミックスデータ解析に関する先行研究の多くは、ゲノム、エピゲノム、トランスクリプトームや臨床情報などの各階層の情報を同列なデータとして扱っているものが多いです。これに対して我々は、生物学的に異なる次元のデータについて、それぞれの階層における特性や経験的知識に基づいたモデリングの有用性を検証していく方針であります。特に、表現型からの距離や一定の因果性、生物学的実験におけるデータの一定性などについての考察に基づいたアプローチを模索していきます(図1、図2)。

本研究の特色ですが、がんの本態解明を目的として、分子間のネットワーク構造が生み出す特性や生体情報の階層性を横断的に解析するため深層学習を用いた方法論の開発を主軸としていることが挙げられます。これは、システム生物学が広く用いるモデリングを中心とした構成的なアプローチとは対照的に、深

層学習によるデータ中心の発見的なアプローチとして位置づけられ、当該研究領域の様々な分野と相互開発的に関わることが出来ると考えております。また、本研究課題は申請者らが所属する国立がん研究センター研究所がん分子修飾制御学分野内のメンバー(客員研究員、外来研究員を含む)で遂行しており、基

礎医学、臨床医学及び情報科学それぞれを専門とした研究員が密接に連携することによって、がんの生物学的理理解に根差したモデリングと、出力に対する解釈性の追求を両輪とした研究の遂行を目指していきます。

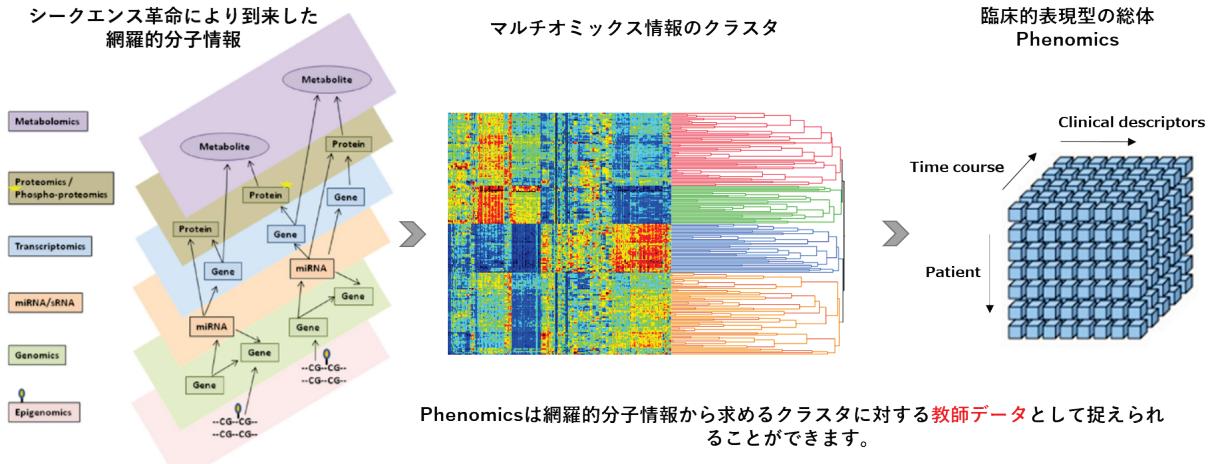


図1: がんマルチオミックス解析におけるPhenomicsの必要性

意味のある治療を選択するための単位として疾患を個別化・層別化するためには、ゲノム情報等の網羅的分子情報(マルチオミックス情報)から求めた個々のクラスタが、どのような臨床的転帰(疾患の実際の振る舞い)に影響するのかを求める必要があります。

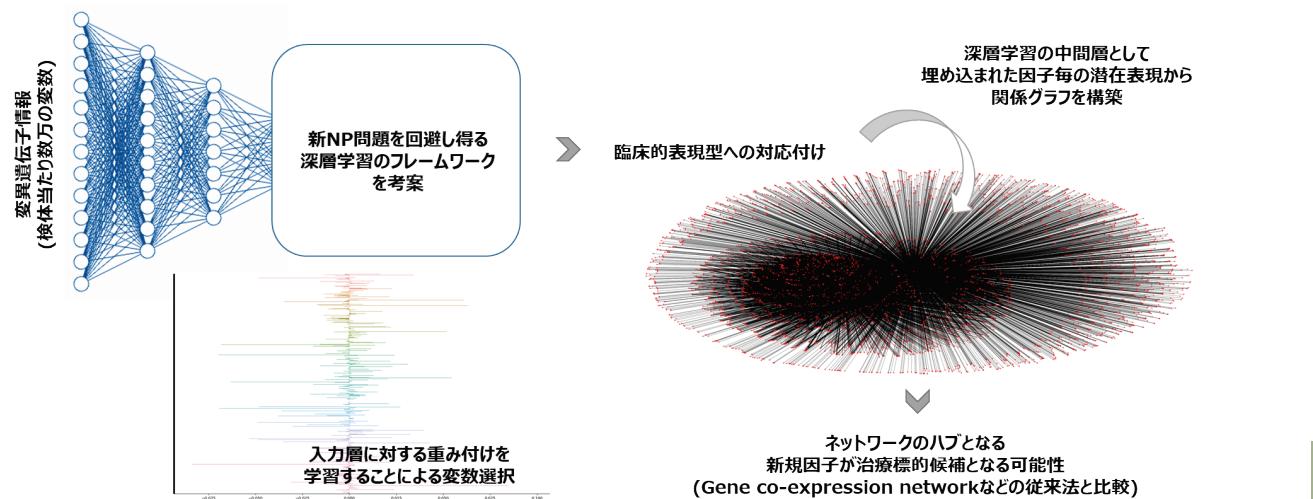


図2: 細胞内分子情報に対する解析の戦略

ゲノム情報(DNA上の変異部位、RNA-seqによる発現遺伝子などが格納されたデータ)は一般的に数万の単位の変数を有し、サンプル数nに対してパラメータ数pが圧倒的に多くなる新NP問題が課題となります。こうしたデータに深層学習を当てはめようとする際には、互いに隣り合う変数間の独立が仮定されるために通常の畳み込み層を使うことが出来ず、一層目のパラメータ数が過大となる問題がありました。

## 公募研究

# BRCA関連遺伝子の変異による乳がんの発がん機構の解明



公募研究 A02-3-18 研究代表者  
京都大学医学研究科 准教授 笹沼 博之

乳がんは、がんのなかでも、日本女性がかかる割合(罹患率)がトップであり、その罹患率は年々増加しています。日本の全乳がんの患者さんの中で、5-10%の患者さんが遺伝性乳がん卵巣がん症候群の原因遺伝子であるBRCA変異を持っています。BRCA1遺伝子は、1994年に本新学術領域班員でいらっしゃる三木義男先生により発見され、その翌年、BRCA2遺伝子が発見されました。BRCA1/2の機能欠損を持つと、乳がん発症のリスクが5-10倍増加すること、また若年発症の傾向があることも知られています。近年、BRCA1変異を持つことがわかったアメリカの人気女優が

乳がんリスクを減らすために、30歳代で予防的切除術を受けたことを記憶されている方も多いかと思います。

BRCA1/2遺伝子発見を端緒に、BRCA遺伝子の研究が精力的に世界中で行われてきました。最も有名なBRCA1/2機能は、DNA修復です。ゲノムDNAは、いろいろな要因により1細胞あたり1日で数十万損傷を受けています。ゲノム中にこうしたDNA損傷があると、ゲノム複製時に変異を生みます。例えば、ゲノムDNAに一本鎖DNA切断があるところが、DNA複製されるとDNA二重鎖切断に変換され、DNA複製が停止してしまいます(図1、①→③)。

BRCA1/2は、DNA複製停止からの回復する時に使われるDNA修復機構に重要な働きをしています(図1、③→④)。近年、がん治療で注目を集めているPARP阻害剤は、一本鎖DNA切断修復に必要であるPARP酵素の活性を阻害します(図1、①→②)。PARP酵素阻害によって蓄積する一本鎖DNA切断は、DNA複製時にDNA二重鎖切断へと変換されます(図1、①→③)。BRCA遺伝子が正常であれば、変換されたDNA二重鎖切断は、速やかに修復されます(図1、③→④)。一方、BRCA機能が减弱したがん細胞において、それらは修復されずに死滅します(図1、③の右矢印)。

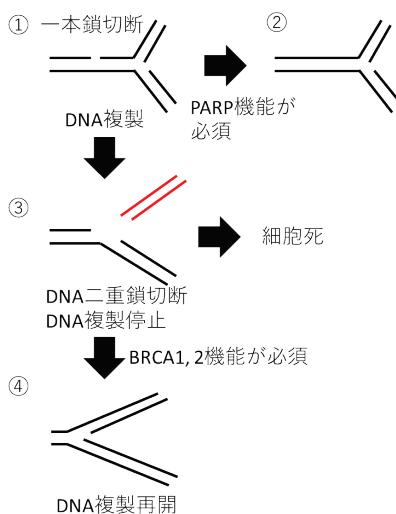


図1: BRCA1の既知の機能

DNA二重鎖にニック(一本鎖DNA切断)が入っているところに、DNA複製が起こる(①)と、DNA二重鎖切断(③)になります。発生したDNA二重鎖切断を直すには、BRCA1,2の機能が必要です(③→④)。

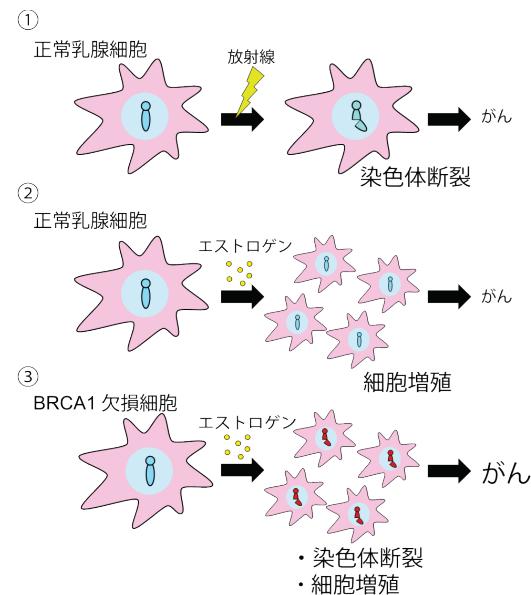


図2: 新しいエストロゲンの変異原性

エストロゲンは、細胞増殖を促進することは良く知られている(②)。BRCA1遺伝子の機能が損なわれると、エストロゲンの細胞増殖促進作用に加え、放射線によってできるDNA損傷(①)のような変異原性の作用が、がん発生を強力に促進する(③)。

なぜ、BRCA遺伝子の機能が損なわれるか、乳がんや卵巣がんが選択的に引き起こされるのでしょうか？性ホルモンであるエストロゲンとの関係が指摘されていますが、明確な答えは得られていません。私たちは、エストロゲンによって生じるDNA損傷を修復するために重要な役割を果たしていることを見つけました。この機能は、すでに解明されている、上述したDNA複製時におけるBRCA1の機能(図1、③→④)とは別のものです。エストロゲンは、エストロゲン受容体を活性化して細胞増殖を促すことは良く知られています(図2の②)。エストロゲン受容体陽性の乳がんでは、エストロゲンによる腫瘍増殖抑制を目的に抗エストロゲン薬(タモキシフェン)を使用したホルモン治療が行われます。我々の発見によると、エストロゲンには2つの側面があります(図2)。良く知られた細胞増殖効果に加えて、DNA損傷を作る変異原性です(図2の③)。

では、このエストロゲンの変異原性とはどういうものでしょうか。活性化したエストロゲン受容体は、転写因子としてエストロ

ゲン応答遺伝子の転写制御領域に結合します。この時にトポイソメラーゼ2型(Top2)酵素が一時的にDNA二重鎖を切断します。我々のグループでは、BRCA1遺伝子の機能が失われた細胞とマウスでは、エストロゲンによってトポイソメラーゼ2(Top2)酵素依存的に切断されたDNA切斷が修復できないことを見つけました。BRCA1が正常な細胞では、エストロゲンのDNA切斷は観察されないですが、BRCA1遺伝子の機能が失われてしまうと、エストロゲンがまるで放射線によってできるDNA二重鎖切斷のような振る舞いをします(図2の①)。私たちが見つけたエストロゲンの変異原性とは、Top2酵素によるDNA二重鎖切斷です。エストロゲンの細胞増殖と変異原性は、細胞のがん化を強く促進するのではないかと考えています(図2の③)。BRCA遺伝子の機能が損なわれると、エストロゲンの変異原性効果が強くなり、エストロゲン受容体が発現している乳がんや卵巣がんが選択的に引き起こされるのではないかと考えています。

私たちのグループでは、今までの知見を

元に、現在、BRCA1欠損細胞を使ってエストロゲンによる細胞のがん化の初期プロセスを解析しようとしています。私たちのBRCA1欠損による発がんモデルを図3に示します。このモデルを検証するために、現在、エストロゲンによってTop2が作るDNA切斷の位置の正確な決定を行なっています(図3の③)。この正確な位置決定がなされれば、その遺伝子座に絞って乳がん患者のゲノムデータベースを使って、ゲノム変異解析をスタートしたいと考えています。またBRCA1遺伝子正常、欠損細胞のエストロゲンに応答した遺伝子発現を比較することも考えています。特にBRCA1欠損によってエストロゲンによって強く発現誘導がかかる癌遺伝子中心に解析を進めることを計画しています。エストロゲンに応答した「Top2による(持続性)DNA切斷」と「遺伝子発現」の因果関係は、現段階では明らかではありません。班員の皆様のご意見を伺いながら、私たちが提案する新しい発がんモデルを証明していきたいと考えています。

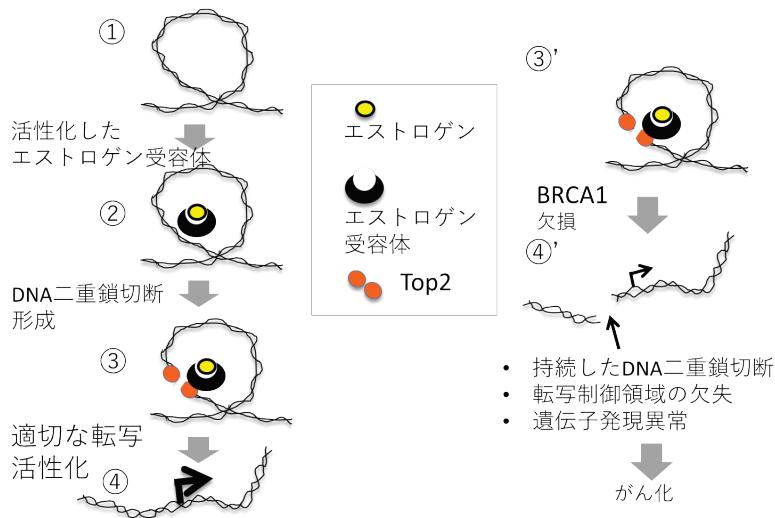


図3: BRCA1欠損による発がんモデル

エストロゲンと結合したエストロゲン受容体は、エストロゲン応答遺伝子の転写制御領域に結合する(②)。この時、トポイソメラーゼ2も一緒に結合し、DNA二重鎖を切断する(③)。この切斷は、正常細胞では速やかに修復され、エストロゲンに応答した適切な転写活性化が行われる(④)。一方、BRCA1が欠損すると、切斷は修復されず、持続したDNA二重鎖切斷となる(右側、③' → ④')。持続したDNA二重鎖切斷は、不正確な修復により転座や塩基欠失を招く。あるいはエストロゲンに応答した遺伝子発現制御を異常にする。

## Information



Human Genome Center

### 東京大学医科学研究所 ヒトゲノム解析センタースーパーコンピュータ

システム癌新次元で用いられるデータ解析ソフトウェアは、ヒトゲノム解析センターのスーパーコンピュータで動いています。スーパーコンピュータはどなたでもご利用になれます（有償）。



大容量のディスク装置



高密度の計算機

[supcom.hgc.jp](http://supcom.hgc.jp)



システム癌  
新次元

<http://neosystemscancer.hgc.jp/>