

令和 4 年 10 月 20 日現在

機関番号：21601

研究種目：挑戦的研究(萌芽)

研究期間：2017～2018

課題番号：17K19687

研究課題名(和文) Ab initio法を駆使した造血器腫瘍の病態解析

研究課題名(英文) Ab initio analysis on the pathophysiology of hematopoietic neoplasms

研究代表者

森 努(Mori, Tsutomu)

福島県立医科大学・看護学部・准教授

研究者番号：60244373

交付決定額(研究期間全体)：(直接経費) 5,000,000円

研究成果の概要(和文)：遺伝子は40億年も進化してきたが故に、その深遠な力は容易に人知の及ぶものではない。しかし幾多の疾患が遺伝子と関わる以上、遺伝子の幅広い機能を素早く的確に知る技術があれば医学進歩を大いに促すだろう。私たちは、遺伝子間相互作用に由来する特性に注目することにより、遺伝子の持つ力を詳細にかつ容易に算出できるab initio遺伝子軌道法(以下ab initio法)を開発した。本研究プロジェクトにおいては、造血器腫瘍に関わる遺伝子群にab initio法を適用し、それらの機能解析を通じて病態解析への応用を図った。

研究成果の学術的意義や社会的意義

これまで複雑系である生物系に対して、理論計算の有効性は限られていると考えられてきた。しかし計算能力の劇的向上により状況は変化した。私たちはab initio法という、過去の実験データを用いない計算法を用いることで生命科学研究に役立つツールとしての基盤を作ることができた。ab initio法の適用範囲は極めて広く、これまで実験上の制約で研究が進まなかった分野の発展にも寄与する可能性がある。例えばnon-coding RNA研究など、実験手法や既存データが乏しい状況で仮説を立てるにあたり、極めて重要な指針を提供するだろう。本研究が契機として、理論計算を融合した生命科学の創設に結びつくこと期待される。

研究成果の概要(英文)：Genes have evolved for four billion years. Now they possess profound functions, which is far beyond human intelligence. Indeed, even after the completion of the human genome project, the etiology of many diseases remains unknown. This indicates that the functional analysis of human genome is still not complete. Since genes are involved in many diseases, if there is a technique to understand their broad functions rapidly and accurately, then it must provide a substantial improvement in medicine. We developed an informatic method called “ab initio GO method” that can perform the functional analysis of a gene of interest within only several minutes, which took more than 10 years by the previous wet experiments. Here we utilized this method to address the pathophysiology of hematopoietic neoplasms.

研究分野：生物情報物理学

キーワード：生物物理学 情報理論 血液内科学 バイオインフォマティクス

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

1. 研究開始当初の背景

(1) 生命は情報の存在様式である。遺伝子の存在目的は、生物の fitness に貢献することであり、その情報自体の存在確率に対して、数億年間に及ぶ長期的な影響を与える。遺伝子は遺伝子産物の構造をコードし、それを通じて物質・エネルギー代謝を支配し、形態と機能を制御する。そして 40 億年の進化の末、複雑な body plan を構築し、中枢神経系と免疫系に代表される高次機能系を産み出すなど、極めて深遠な機能を獲得するに至った。この過程で中心的な役割を演じてきたのは情報の進化である。

(2) 進化するのは遺伝子だけではない。遺伝子間相互作用も淘汰圧を受けることによって進化の経過を辿る。相互作用の進化の結果、遺伝子間に情報交換が発生し、それは生体系のエントロピーを制御することにより、確率的事象に影響を及ぼす。遺伝子の持つ情報量は莫大であり、その情報が生物の存在確率を左右するのであるから、遺伝子は生体および、他の多くの遺伝子に対して、影響力を行使すると予想される。実際、遺伝子間に相互作用が生じることは、分子生物学では周知の事実である。そもそも遺伝子の複雑な機能は、遺伝子単独で発揮し得るものではなく、他の多くの遺伝子との共同作業により初めて果たし得るものである。細胞内外の複雑な条件に応じて適時適切な共同作業をする能力もまた、進化の過程で獲得されたものに違いない。すなわち遺伝情報は分子構造だけでなく、遺伝子間の相互関係をもコードするのである。

(3) 遺伝子は 40 億年も進化してきたが故に、その深遠な力は容易に人知の及ぶものではない。しかし幾多の疾患が遺伝子と関わる以上、遺伝子の幅広い機能を素早く的確に知る技術があれば医学進歩を大いに促すだろう。私たちは、遺伝子間相互作用に由来する特性に注目することにより、遺伝子の持つ力を詳細にかつ容易に算出できる *ab initio* 遺伝子軌道法(以下 *ab initio* 法)を開発した。本研究プロジェクトにおいては、造血器腫瘍に関わる遺伝子群に *ab initio* 法を適用し、それらの機能解析を通じて病態解析への応用を図った。

2. 研究の目的

(1) ヒトゲノム計画が終了して 20 年も経過したが、原因不明で治療法もない疾患は未だ数多く残されている。これは染色体上の遺伝子機能解析が未完成であることを示唆している。しかし従来の“wet”な実験法では、遺伝子機能の決定に 10 年余も掛かる。私たちは無限次元の情報量を正確に扱う理論を世界に先駆けて開発し、これを応用して、実験操作を必要としない遺伝子機能解析法である *ab initio* 法を開発した(特許第 6820621 号)。これは遺伝子概念を根底から見直し、情報理論により遺伝子特性を数値化し、機能計算に結びつける手法である。結果、実験を一切要せず、遺伝子機能を詳細に、しかも“wet”実験法より圧倒的な高速で解析することが可能となった。この技術で、従来約 10 年掛かった新規遺伝子の多角的な機能解析・相互作用解析・疾患関連性予測が、わずか数分で行えるようになった。さらに分子間相互作用とパスウェイ解析、さらに疾患関連性まで予測でき、疾患マーカーや治療標的の特定にも応用できる。

(2) 同時に私たちは、NIH の TCGA データベースを用い、腫瘍の発症・進展・予後に関わる遺伝子を発見するプログラム STAIC を開発した。これを *ab initio* 法と統合的に活用すれば、新規な腫瘍遺伝子の発見から機能解析まで、従来よりも遙かに効率的に実行できる。本研究に先立っては、*ab initio* 法の精度を検証する目的で、造血器腫瘍に関わる遺伝子群に適用して予備的検討を行った。この過程では、長らく発見が待たれていた RUNX1 のユビキチンリガーゼの絞り込みに成功するなど、予想以上の有用性を強く認識するに至った。本研究では *ab initio* 法の具体的な応用を進めるため、造血器腫瘍の発症に関わる造血系遺伝子 RUNX1, ASXL1, EZH2, HMGA2 の生理機能と病態関連性を *ab initio* 法により予測し、それらを wet 実験により検証することから、造血器腫瘍の発症メカニズムを明らかにするとともに、dry/wet を融合する生命科学的手法の確立を目指した。

3. 研究の方法

(1) 様々な生理現象が多くの分子の複雑な相互作用によって制御されていることが明らかになり、従来の単一もしくは少数の分子に着目した研究手法に限界があることは、今や誰の目にも明白である。複雑系である生命現象を理解し、疾患の予防、治療に役立てるためには、今後は情報理論を駆使した新しい解析法の開発が必須である。遺伝子発現パターンをもとにパスウェイ解析を行う手法としては、Gene Set Enrichment Analysis (GSEA) などいくつかのプログラムが開発され広く使用されているが、これらの方法と *ab initio* 法を比較すると下記のような相違がある。

従来法は、機能解析の完了した遺伝子を組み合わせることで人為的に定めた「パスウェイ」が、実験系において関与する確率を算出するものである。個々の遺伝子機能(特に新規遺伝子の機能)を計算することはできない。また、パスウェイの選択には、しばしば科学的根拠が欠落していた。

従来は確率論と情報理論の関係が不明であった。私たちは「情報」と「確率」を結び新たな定理(情報と確立の等価性原理)を見だし、*ab initio* 法の基礎としている。その結果、統計的有意性を持って情報を扱うことに世界で初めて成功した。

無限次元の確率変数を取り扱う情報理論を、大規模グラフ理論(ネットワーク解析)に応用し

たのは、ab initio 法が世界初である。従来のパスウェイ解析は、これら両方の理論には立脚しておらず、得られる結果の解釈は客観性を欠くものとなる。

以上 ~ の理由から、私たちが開発した ab initio 法では、従来法よりも遥かに正確に分子機能を予測できると考えられる。本研究は、ab initio 法を医学・生命科学研究に応用する第一歩となる重要な研究である。

(2) 科学進歩への貢献度を考えてみよう。膨大な情報が溢れる informatics は、何らかの本質的貢献を成し遂げただろうか？否、単にデータ量を増やすことができただけである。方法論的な制約もあり、従来の informatics では、学問自体の追求は常に後回しにされてきた。結果的に幾多の方法論が並立する一方、大規模なデータの奥に隠された科学的真実を提示するには至らなかったのである。

私たちは、一見何気ない数値データの中に、本質的に重要な情報が含まれることに気づいた。永い進化の過程で遺伝子が過酷なふるいに掛けられてきた足跡こそが、根源的な価値を持つ情報なのである。ワトソン=クリック・モデルによれば、これは遺伝子産物の構造情報に帰着する。しかし私たちは、遺伝子機能を定める最も重要な情報が、分子構造ではなくネットワーク構造としてコードされていることを見出した。すなわち遺伝子のネットワーク情報量 $H(N)$ を算出することから、分子構造よりネットワーク構造に依拠する情報量の方が遙かに巨大であることが判明した。したがって、ネットワーク構造を正しい方法で解析することが肝要なのである。私たちは遺伝子ネットワークの構造的特性を、数学・物理学を援用しながら明らかにするとともに、刻まれた進化の足跡を解読する手法を開発した。

4. 研究成果

【成果の概要】

ここでは ab initio 法の理論面だけに絞って、研究経過のアウトラインを記載することとする。

【多次元情報理論の体系化】

遺伝子が発現する情報は極めて多彩であり、情報理論を適用するに際しては、複雑な情報を一元的に扱うことを可能とするような、数理的手段を確立する必要がある。しかるに遺伝情報は量的変化・構造変化・分解制御・局在変化・複合体形成・クロマチン制御など、多彩な様式で同時に表出され、その変化も連続的あるいは離散的とバラバラであるため、情報の総体を統一的指標で把握することは、従来極めて困難であった。この難問を解決し、複雑系である生命現象を理解して、疾患の予防・治療に役立てるためには、情報理論の大幅な拡張が必要である。そのため私たちは、遺伝子の発現する情報を、極めて高次元の確率変数 (highly multi-dimensional random variable) として扱うこととした。

【遺伝子情報量】

まず遺伝子 X が発現する情報を m 次元確率変数と捉える。各次元ごとに取り得る状態が n 通りであって、各状態はその状態を取るか取らないかの何れかとする。このとき遺伝情報は並列に同時表出されることを考慮すると、ある時刻 t において遺伝情報が取り得る全ての状態数は 2^m と計算できる。実際は m, n とともに大きな数であるため、遺伝子が取り得る状態数は莫大な数となる。これは生体の複雑さを考えれば納得できる規模であろう。ここで留意すべきは、ゲノムが非常に複雑な情報を発現していながら、それがコードするのは混沌ではなく、逆にカオスに見紛うほどの複雑な生体系に対して、必要十分な秩序を付与していることである。そのメカニズムこそ遺伝情報の本質であるに相違ない。

【遺伝子の最大情報量】

遺伝子 X が発現しうる最大情報量はゲノム上の塩基配列により規定される。ゲノムサイズを 30 億塩基対とし、タンパクをコードする遺伝子の数を 2 万とすると、各遺伝子に割り当てられる平均領域長は約 150kb である。この見積もりは、従来は意味不明なジャンクとされてきた遺伝子間領域にも、間接的であれ情報に影響を与える構造が含まれていると想定した。実際、近年の GWAS の結果は、この想定の上しさを示唆している。すると各遺伝子が発現しうる最大情報量は、塩基配列の順列を考慮することにより、 $41.5E5 = 9.97 \times 10^{90,308}$ 通りと算出される。この情報量が如何に巨大であるかは、物理化学で用いられるアボガド口数が高々 6.02×10^{23} でしかないことから明らかであろう。すなわち物理系では各構成要素が区別されないため、それが産生される情報量の上限となってしまう。それに対して生物系の場合、ヌクレオチドやアミノ酸など構成要素の種類は極めて少数にも関わらず、順列が情報源となるために、莫大な情報量が産み出されるのである。

しかしこの最大情報量の中で、生理的意味合いを科学的手法で人間によって実際に証明できるのは、極めて少ない場合に留まるに違いない。それにも関わらず、高等な生物ほど、より広大な遺伝子間領域が介在するという良く知られる知見は、遺伝子間に存在する情報が、生物進化の過程では意外に大きな貢献をしてきたことを暗示するのではなからうか。ともあれ、遺伝子のコードする莫大な情報が、如何にして複雑な生命体を成立させるかを探ることが、われわれに与えられた最初の課題である。

【多次元確率変数】

ある時刻 t に遺伝子 X が状態 (m, n) を取る状態確率を $p(m, n)$ とすると、その自己情報量 $I(m, n)$ は $I(m, n) = -\log[p(m, n)]$ である。このとき遺伝子の発現する全情報量 $H(X)$ は、 $H(X) = -\sum p(m, n) \log[p(m, n)]$ と表現される。ここで従来の情報理論では、ON/OFF のごとく低次元の情報を扱っており、遺伝子情報のような極めて高次元な情報を扱うことはできなかった。そこで私たちは、個々の遺伝子が発現する高次元情報が及ぼす生体への影響を定量的に扱うために、多次元情報理論を体系化することとした。

【遺伝子と生体の相互情報量】

私たちは遺伝子が生体に与える影響を定量的に示す指標として、相互情報量 (mutual information, MI) が適切であることを見出した。遺伝子が生体に与える影響は MI であり、その分だけ生体エントロピーが減少する。これは、遺伝子情報によって生体に秩序がもたらされることを意味しており、その結果として生命が維持され、子孫を残すために必要な機能が付与されるのである。ここで大切なのは、従来強調されてきた、生体エントロピー減少のメカニズムの主体を成すのは遺伝子情報であることである。言葉を換えれば、生体の状態を規定して混沌を回避するための情報を発現することこそが、遺伝子の持つ主要な機能と言える。私たちは遺伝子から発現する情報が生体に及ぼす影響を表現するような、一般性を持つ数式を提案した。

【遺伝子間相互情報量】

上述の遺伝子と生体間の相互情報量を論じる際には、まず遺伝子間相互情報量を念頭に置く必要がある。なぜなら個々の遺伝子情報が発現する過程においては、複数の遺伝子相互の関係が重要だからである。生体構築の主体は幾多の遺伝子自体なのであるから、ある遺伝子が生体に及ぼす影響の主たる要素は、その遺伝子が他の遺伝子に与える影響だと見なせるだろう。この時、遺伝子間相互作用の様式もまた MI である。私たちは遺伝子間 MI を表現する数式を見出し、それを体系的に扱う一連の方法を開発した (特許第 6820621 号)。

ここで遺伝子間 MI について考察を付け加える。従来、遺伝子間相互作用は wet 実験の所見に基づいて、分子間相互作用として各論的に議論されてきたことを思い起こして欲しい。この場合、相互作用の種類が異なれば、当然のごとく伝達される情報も異なると考えられてきた。しかし確率論に軸足を置き、シャノンが開発した情報理論を適用すれば、非常に幅広い異なる形式の情報交換様式を、ただ一種類の指標 (すなわち MI) で表現することが可能となるのである。

この理論に基づくと、遺伝子間 MI の大きさが指数分布に従うことが導かれる。私たちは下記の *ab initio* 法の計算から、大多数の遺伝子間 MI は実際に指数分布に従っていることを見出した。すなわち情報理論という総論的な見地に立つことで、遺伝子同士が交換する情報は、物理化学的な性質よりむしろ数理的性質が本質的なのだと、捉え直すことができるだろう。

【多次元相互情報量の計算】

上述の通り遺伝子情報は無限に近いほど大きな次元を持つ。したがって遺伝子間相互情報量は極めて高次元な相互情報量である。ところが、従来は多次元相互情報量の扱いそのものが行われてこなかったし、計算に用いるべきデータもなかった。そこで私たちは、遺伝子間の高次元な相互情報量を、正確かつ容易に計算する手段を確立することとした。この方法は知財としての申請を準備中である。結果的に確立された方法から遺伝子間 MI が算出できるが、これは遺伝子間の機能的近接度を表現することが数理解析により導かれ、実際その通りであることが多くの実例によって裏付けられた。

【遺伝子軌道】

私たちは情報距離 (information metric) に基づく空間を考え、これを情報空間と名付けた。ヒトゲノムがコードする約 2 万の遺伝子は、1 万次元の情報空間に存在する情報因子 (情報を持つ因子を *informaton* と命名した) と見なすことができる。このとき、各 *informaton* の形成する情報力学的な作用を、情報軌道 (information orbital) として表現することができる。これは個々の遺伝子が他の遺伝子に及ぼす影響の大きさを示し、遺伝子間協調性を定量的に示す指標である。この情報軌道の数式は、物理化学で汎用される分子軌道 (molecular orbital, MO) と同様のスレーター型関数である。

【*ab initio* 遺伝子軌道法】

情報軌道は遺伝子の示す影響力であり、多次元の情報空間において情報軌道の「重ね合わせ」を行うことができる。私たちは、できた合成軌道を遺伝子軌道 (genetic orbital, GO) と命名した。これは分子軌道法 (MO 法) の場合に、各原子軌道 (atomic orbital) の重ね合わせによって分子軌道を求めるのと同様の手法である。形成された GO は、ある遺伝子が他の遺伝子に及ぼす影響力の総和である。その生物学的意味合いは、ある遺伝子が他の多くの遺伝子を適時適切に「利用」することによって、その遺伝子本来の機能を効果的に発揮させることにある。すなわち GO は、遺伝子が発揮するネットワーク機能であり、拡張された遺伝情報と捉えることができるだろう。そのアナロジーとして、ちょうど人間が発揮する社会的機能を考えれば良い。私たち人間は社会

的生物であり、他の多くの人たちに多少なりとも影響を与え、彼らを動かすことによって社会的機能を発揮しつつ生きているのである。遺伝子も生体内でネットワーク機能を発揮し、広範な生物学的効果に關与することになる。

重要なのは G0, すなわち MI に由来するネットワーク機能の大きさを計算すると、遺伝子産物であるタンパクなどの分子構造が発揮する情報量よりも遙かに巨大であることが明確に判ることである。いわゆるオーダーの違いではなく、完全に比較にならないほどの格差なのである。さらに私たちは集団遺伝学的手法を用い、ネットワーク機能の進化過程を導くことができた。すなわち、遺伝情報の主体はネットワーク構造としてコードされるのであって、分子構造ではない。このネットワーク機能を素早く計算する方法が、ab initio 遺伝子軌道法である。われわれの計算では、造血系悪性腫瘍における遺伝子間相互作用を、正しく算出することができた。

【今後の展望】

これまで、複雑系である生物系に対して、理論的計算法の持つ有効性は限られていると考えられてきた。しかし計算能力の劇的な向上により状況が変化した。本研究は ab initio 法という、過去の実験データを用いない計算法を用いることによって今後の生命科学研究に役立つツールとしての基盤を作ることができた。ab initio 法の適用範囲は極めて広く、これまでは実験上の制約で研究が進まなかった分野の発展にも、大きく寄与する可能性がある。例えば non-coding RNA 研究など、実験ツールや既存のデータが乏しい状況で仮説を立てるにあたり、極めて重要な指針を提供するだろう。本研究を契機として、理論計算学を融合した、新しい生命科学の研究分野を創設したいと希望している。

5. 主な発表論文等

〔雑誌論文〕(計3件)

Ueda K, Kazuhiko I, Tsutomu M, et al. Hmga2 collaborates with JAK2V617F in the development of myeloproliferative neoplasms. Blood Advances, 査読あり, vol 1, 2017, 1001-1005

DOI: 10.1182/bloodadvances.2017004457

Sato N, Haga J, Mori T, et al. Ex vivo Pretreatment of Islets with Mitomycin C: Reduction in Immunogenic Potential of Islets by Suppressing Secretion of Multiple Chemotactic Factors. Cell transplantation, vol 26, 2017, 1392-1404

DOI: 10.1177/0963689717721233

大原正志, 河村隆, 森努, 錫谷達夫, ヒト大腸癌培養細胞における短鎖脂肪酸の抗腫瘍効果と作用機序 39 巻 2 号 107-112 頁

〔産業財産権〕

取得状況(計2件)

名称: 相互依存性の特定方法

発明者: 森 努, 河村 隆

権利者: 公立大学法人福島県立医科大学

種類: 特許

番号: 特許第 6820621 号

出願年月日: 2018 年 3 月 30 日

取得年月日: 2021 年 1 月 7 日

国内外の別: 国内

名称: 肺癌の予後バイオマーカー

発明者: 鈴木弘行, 森努, 武藤哲史

権利者: 福島県立医科大学

種類: 特許

番号: 特許第 7114112 号

出願年月日: 2019 年 7 月 24 日

取得年月日: 2022 年 7 月 29 日

国内外の別: 国内

〔その他〕

ホームページ等

<http://www.nirf.jp>

6. 研究組織

(1) 研究分担者

研究分担者氏名: 北村 俊雄

ローマ字氏名：KITAMURA, Toshio

所属研究機関名：東京大学

部局名：医科学研究所

職名：教授

研究者番号：20282527

研究分担者氏名：合山 進

ローマ字氏名：GOYAMA, Susumu

所属研究機関名：東京大学

部局名：医科学研究所

職名：准教授

研究者番号：80431849

研究分担者氏名：池田 和彦

ローマ字氏名：IKEDA, Kazuhiko

所属研究機関名：福島県立医科大学

部局名：医学部

職名：教授

研究者番号：90381392

(2)研究協力者

研究協力者氏名：河村 隆

ローマ字氏名：KAWAMURA, Takashi

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。