

2022年5月10日

医療法人徳洲会 野崎徳洲会病院附属研究所

学習過程を使って細胞は遺伝子発現を制御する

【本研究成果のポイント】

- ・細胞による遺伝子発現制御の分子機構は、実は、AIの深層強化学習の過程と類似している。
- ・2つの因子(遺伝子)の量は、競合的増幅と誤差依存性減衰の確率過程を繰り返すことで、自律的に最適比率に至る、つまり、学習能を有する動的平衡状態となる。
- ・この学習過程をもつ遺伝子発現制御モデルにより、ヒト全遺伝子の様々な分化過程を再現可能。
- ・学習能を持つ動的平衡状態を表現する「生物学的慣性の法則」を、生命の原理として、数式で提示。
- ・重症コロナ病棟で有名な野崎徳洲会病院の、斬新な基礎研究が、Scientific Reports 誌に掲載された。

【概要】

細胞の中では、1万以上の遺伝子の発現量が、状況に応じて正確に調整されている。しかし、このような複雑な調整を可能とする原理は、未解明であった。野崎徳洲会病院附属研究所の山口智之主任研究員は、細胞の中で、人工知能(AI)の深層強化学習に類似した学習過程が働いていると想定することで、全遺伝子の制御が可能となることを理論的に示した。この学習に必要な素過程は、ゲノム構造の変化やストレス応答など、既知の分子機構で説明可能であった。自律的に最適化するという「学習能」を有するためには、細胞が中味の分子を更新しながらもその比率を一定に保つという「動的平衡」が重要であった。今回、これを細胞の持つ原理と考え、「生物学的慣性の法則」として提唱した。原理に基づくことで、工学的な医療設計が、将来可能となるかもしれない。このような斬新な理論研究の成果が、重症コロナ病棟で有名になった野崎徳洲会病院の附属研究所から発表され、Nature 系列の国際科学誌 *Scientific Reports* に、5月9日、オンラインで公開された。

【研究の背景】

細胞は非常に複雑なシステムであり、1万種類以上の遺伝子の発現量を状況に応じて、精密に制御している。遺伝子発現制御の分子機構については、近年の解析技術の進歩もあり、膨大な量のデータが蓄積され続けている。分子生物学は、細胞を「精密な機械」と見なすことで、その部品である各分子・遺伝子の役割を解明してきた。このような個別現象についての多数の研究成果にも関わらず、これほどまでに複雑な系の適切な制御がどうして可能なのかは、全くわかっていない。500個程度の最小限の遺伝子を持つ細菌の同一環境下で短期間の過程について、全遺伝子発現制御のモデル化に成功した論文が、2022年の *Cell* 誌に掲載されるような状況である。

一方で、最新の機械は著しい進歩を遂げ、学習能を持つようになった。Alpha Go の成果に代表される、深層強化学習を利用した人工知能(AI)が注目されている(用語説明1)。複雑な状況の中で適切に動作できるためには、「精密に作り込んだ機械」よりも「自律的学習能を持つ機械」が優れている事が明らかになっている。多くの想定外の状況に適宜対応する必要のある生物・細胞は、「精密な機械」ではなく「学習能を持つ機械」と見立てるべきではないかと考えた。

AIと生物学の融合研究は、AI技術で膨大な生物学データを解析する研究に、ほぼ限定されていた。生物を「学習能を持つ機械」と見立てる研究は少ない。因果関係を不明瞭化する学習過程や確率過程を、積極的に取り入れる生物学研究は、特に、遺伝子発現制御機構の分野では、これまで敬遠されてきた。

【研究の内容】

最初に、抽象化した単純なモデルで、「学習能」に必要な最小限の過程を明らかにしようと試みた。2つの因子の数が、増加と減少の確率過程を繰り返すうちに、自動的に目標比率に近づくことが可能かを、シミュレーションで調べた。いくつかの増加・減少方式を試した結果、小さな加算項を含む競合的増幅による増加と、目標比率と現比率の誤差に依存する減衰による減少を想定した場合、2つの因子の比率は自動的に、目標比率に近づくことが分かった。これを学習的ペア制御過程と呼ぶこととした(用語説明2)。

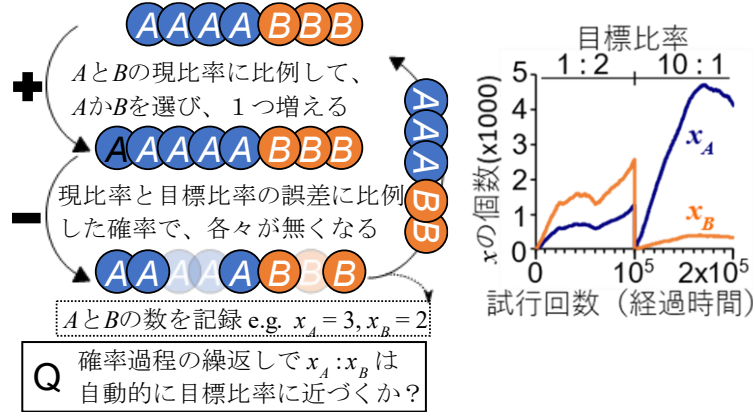


図1 学習的ペア制御過程のシミュレーション

二種類の因子AとBの数 x_A と x_B が増加減少の確率過程を繰り返す。 $(x_A+1) : (x_B+1)$ の比率で選ばれた方が1つ増える競合的増幅と、目標比率と現比率の間の二乗誤差に比例する減衰の確率過程を繰り返すと想定した。右図シミュレーション結果は、 $x_A : x_B$ の目標比率を1:2から10:1に変えた場合でも、各時点の目標比率に自律的に近づくことを示す。

多数の遺伝子の発現量を制御する場合は、この学習的ペア制御過程が階層構造になっていると想定する。つまり、関連する2つの遺伝子群の比率には、各状況でストレスの少ない正しい比率があり、学習的ペア制御過程が働く想定する。さらに、各群の内部にある2つの遺伝子の間の比率も別の学習的ペア制御過程で自律的に最適化される。このような階層構造を想定すると、1万以上の因子の比率であっても目標比率に自律的に近づく。実際に、初期胚から胚盤胞に至るまでの全遺伝子発現パターンの変化を再現できた。その精度は、実測される1細胞毎のばらつきと同程度であった。さらに、全く同一のモデルで、初期値を造血幹細胞の遺伝子発現パターンで設定し、目標値を末梢血細胞で設定すると、血液細胞分化の遺伝子発現パターンの変化を再現することもできた。

このモデルで必要であったのは、どの遺伝子(群)同士がペアとなるかを予め決める全ヒト細胞に共通のペアの階層構造と、その各ペアの活性化度を最適化する学習的ペア制御過程という単一の規則だけである。その上で、各遺伝子群の発現量比が、「不適切な場合はその比率が不安定化する」と想定するだけで、1万以上の全遺伝子発現パターンが自律的に適切な比率に近づく。

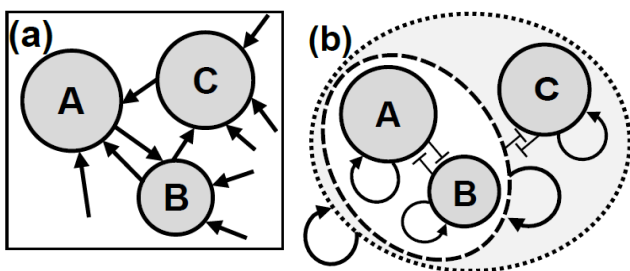


図2 遺伝子発現制御のモデル

従来モデル(a)では各遺伝子は別の遺伝子により制御されるため、数が増える毎に、複雑になる。今回提案するモデル(b)では、学習的ペア制御過程が階層構造を形成している。2種類の集団間の競合的増幅を想定するだけである故に、外側に同様の構造を追加するなど、容易に因子数を増やせる。

今回提案する、学習能を持つ遺伝子発現制御のモデルは、従来モデルとは大きく異なる。階層構造を追加するだけで、同じシステムで因子の数を増やせる。この単純さは、モデル化が容易なだけでなく、現実の進化や組織化でも有利となる。各因子は、増幅や減衰という自分自身の量に比例する変化量で増減する。そのため、原則的には現比率を維持する。実際、「動的平衡」と言われるように、細胞に限らず生物は、「内部の構成要素を入れ替えながらも、その成分比率を維持する」ことを特徴とする。このような動的平衡状態を、図2aのような複雑な制御ネットワークの結果と考えるのではなく、生物に共通する基本的動作原理と考えた。より重要なことに、このこと自体(現在の比率を再現できること)が、学習能に必須である。

各成分が、競合的増幅と誤差依存性減衰と小加算項（ノイズ項）で経時に変化するということは、次の数式で表せる。

$$\frac{dx_i}{dt} = A \frac{x_i}{\sum x_j} - E_{i(x, T)} x_i + B_{i(x)}$$

これを生物学的慣性の法則と呼び、生命の原理として提唱する。

より簡単な言葉で説明すると、以下になる。各個それぞれは、競争しながら勝手に増え、うまく行かなければ減るだけで、時々間違いもする。それを繰り返していくうちに、うまくいく状態へと自動的に近づいていく。うまく行っている間は、惰性で同じことを続けていれば良い。これが、多数のものを柔軟に制御するためには、優れた方法だ。

従来モデルのような機械的システムやあるべき比率を強制する方式は、権威主義的システムに似て、柔軟な対応ができない。個々の自由な増幅を保障する方式は、自由主義的システムに似ており、こちらの方がむしろうまく行く。自由の重要性の理論的根拠をも示唆している。

確率的ゆらぎ（ノイズ）や増幅という、生物を「精密な機械」と見なす分子生物学では不都合な真実として、軽視されてきた過程が、「学習能する機械」では不可欠な役割を担っている。今後は、全ヒト細胞で唯一共通のペアの階層構造を具体的に明らかにすることが目標となる。これが解明できれば、生物を、工学のように、数式・理論に基づいて操作することが可能となる新時代がくるだろう。

【用語の解説】

1. 人工知能(AI)の深層強化学習；2015年開発された囲碁ソフト Alpha Go がプロ棋士に勝利したことを契機に注目を集め、画像認識や言語処理等に活用されている技術。深層学習は深い階層を持つニューラルネットワークで、多次元情報を重みづけに従って処理を重ねていき、低次元の出力をだす。出力と正解との誤差を基に、逆伝播法によって重みづけの値の変更に反映させることで、処理の精度を上げる。強化学習は、確率的な試行錯誤を繰り返しながら精度を上げていく学習方法。
2. 学習的ペア制御過程の具体例：2つの因子の具体例としては、同様の転写制御領域を持つ2つの類縁遺伝子 A と B のヒストンアセチル化の数（クロマチン構造の開放度を表す値）を考える。ヒストンのアセチル化によりクロマチン構造が開いた状態になり、転写因子が接着しやすくなる。同様の制御領域を持つ2つの遺伝子の発現量の場合、発現量の比率は、このクロマチンの開いている比率に比例すると想定できる。つまり、シミュレーション結果は、自律的に正しい発現量比率に至るシステムであることを示す。

学習能に必要な次の素過程は、次のような既知の遺伝子発現制御の分子機構で説明できる。

1. 競合：1つの転写因子はどちらか一方の遺伝子座を選択して結合する。
2. 増幅：開いている遺伝子座に接着しやすく、接着した転写因子はクロマチンを開くヒストンアセチル化酵素を引き連れてくるため、転写因子の接着した遺伝子座はさらに開く。
3. 小さな加算項；体細胞は全ゲノムを持ち、偶然その遺伝子に転写因子が接着する可能性はある。
4. 減衰：個々のアセチル基がはずれる確率は同じである。
5. 誤差依存性減衰：発現量比が不適切な場合は、ストレス応答が誘導され、クロマチンが閉じる。

提示した抽象的モデルは、2つの因子の数が自律的に最適化する場合なら、普遍的に成立する過程であり、2つの細胞種の数の比率にも応用できる。その場合、増幅は「各細胞の一定確率での分裂増殖」であり、誤差依存性減衰は「不適切な比率の場合に細胞死が誘導されること」を指す。樹木の枝の太さ、分枝する神経回路にも同じ、学習的ペア制御過程が働いている。

【論文情報】

- ・掲載雑誌：Scientific Reports
- ・論文題目： Learning processes in hierarchical pairs regulate entire gene expression in cells
- ・著者：山口 智之 Yamaguchi Tomoyuki
- ・所属： 野崎徳洲会病院附属研究所 Research Institute, Nozaki Tokushukai Hospital
- ・DOI： 10.1038/s41598-022-10998-z

【お問い合わせ先】

山口 智之

Yamaguchi Tomoyuki

野崎徳洲会病院附属研究所

免疫システム研究部門 主任研究員

〒574-0074 大阪府大東市谷川 2-10-50

Tel: 072-874-1641

Fax: 072-818-3723

t.yamaguchi@tokushukai.jp