

計算機で生き物を科学する

計算工学専攻 瀬々 潤 研究室

瀬々 潤 准教授 1976年埼玉県生まれ。東京大学大学院新領域創成科学研究科情報生命科学専攻にて博士号取得。2011年より、大学院情報理工学研究科計算工学専攻准教授。



バイオインフォマティクスという、計算機の上で生命科学を扱う分野がある。測定技術と計算機の発展により生まれた新しい分野である。大量の情報を高速で処理することができるという計算機の特長を活かすことで、生き物を新たな視点から見ることができるようになるのだ。今回は、瀬々先生の研究を生命科学的アプローチから紹介する。

バイオインフォマティクスとは

生命科学の研究をしている研究室と聞くと、どのような場所を思い浮かべるだろうか。多くの人は、生き物や化学薬品を使って実験をしている研究室というイメージをもっているだろう。

瀬々研究室も生命科学を扱う研究室である。ところが、この研究室にはコンピュータがずらりと並んでいる。瀬々研究室では、計算機を用いて生き物を理解する、バイオインフォマティクスという分野の研究を行なっている。最近注目を集めている分野だ。

DNAの二重らせん構造が提唱された1953年から約半世紀、生命科学の中で最も発展したのは分子生物学という分野である。分子生物学では、生き物を機械のように捉え、生き物を構成するさまざまな要素を、その部品のように分解した上で調べる。具体的には、個体を組織、細胞、細胞小器官、分子と、どんどん細かく分解していき、生き

物を構成するさまざまなパーツの仕組みを解明しようとするものである。

1990年代に入ると、測定技術の精度が上がり、生き物をよりミクロな視点で見えていくことができるようになった。そして、生き物の体内にはどのような物質が存在するのか、あるいはその物質がどのようにはたらいているのか、といったことについて大量の測定データを得られるようになった。さらに、計算機のめざましい技術発展により、得られたデータをより大量に、かつ高速に処理できるようになった。

これらのことをきっかけに、1990年代後半、生命科学を計算機の上で扱おうとするバイオインフォマティクスという分野が生まれ、研究が盛んに行われるようになった。

先ほど述べたように、分子生物学では生き物をたくさんの細かいパーツに分解して考える。その一方で、バイオインフォマティクスでは、パーツごとに得られるデータを処理し、パーツ同士の関

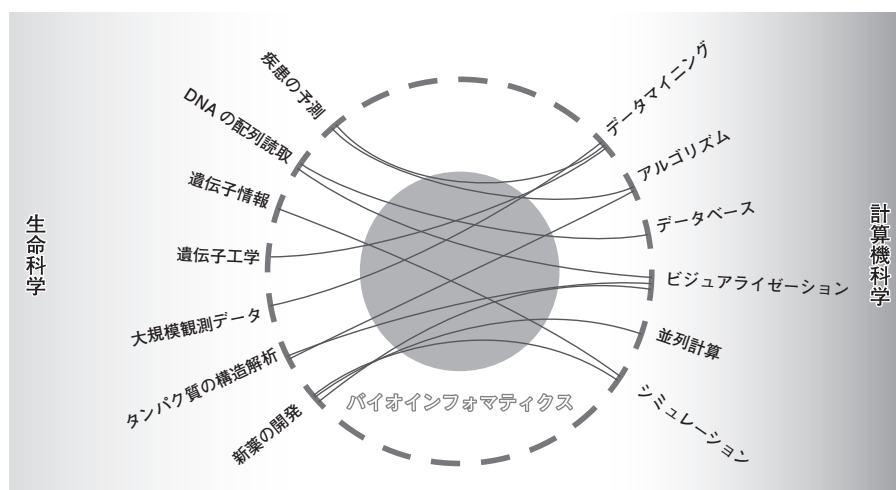


図1 バイオインフォマティクス

生命科学と計算機科学のさまざまな分野がつながった幅広い研究がなされている。

係を導出することで、個々のパーツを組み上げていき、個体全体を解明しようとしている。しかし、分解したパーツをただ寄せ集めただけでは生命の息吹は生まれない。パーツ同士の相互関係が成り立っているからこそ、生き物が形作られているのだ。膨大な種類のパーツについて、どのパーツの組み合わせが生命現象に必要なのかを手で一つひとつ実験を繰り返して調べていくことには、とてつもない時間と労力が必要になる。そこで、瀬々先生は、数え上げや繰り返しを得意としている計算機を使い、適切なプログラムを組み立てて実行することで、これを格段に速く行おうとしている。

このように、計算機を用いることにより、従来の分子生物学の、一つひとつのパーツについて深く探っていく見方だけでなく、生き物を形作るパーツの相互関係に着目した、新しい見方もできるようになると先生は考えている。

バイオインフォマティクスは、生命科学という分野に計算機科学という全く異なった分野の手法を取り入れたため、2つの大きな分野をまたいでおり、図のようにそれぞれの分野の広い範囲にカバーしている(図1)。そのため、両方の知識が必要になる横断的な分野である。

次に、バイオインフォマティクスによって具体的にできるようになったことを、ゲノム配列解析を例にとって紹介していく。

解析技術の進歩と計算機の利用

1991年、ヒトの身体的设计図であるDNAの総体、ヒトゲノムを解読しようというヒトゲノム計画が世界中の研究チームの協力によって始まり、12年後の2003年に解読が終了した。この計画により、ヒトゲノムは約30億の塩基対から構成されていることが判明した。

これをデータにするとどれくらいの大きさになるのか考えてみよう。コンピュータの中では、すべての情報が2進法、すなわち0と1の2種類の数字の列で表されている。一方ゲノムはアデニン、シトシン、グアニン、チミンという4種類の塩基の配列である。よって、この4種類の塩基をデータとして扱うには、2進法を用いて、00, 01, 10, 11をそれぞれの塩基に当てはめればよい。1文字分の情報量は1bitなので、1塩基は2bitの情報量をもつことになる。

ここから計算すると、ヒトゲノムの情報量は約60億bitとなる。8bitは1byteであり、100万byteは約1MBであるから、ヒトゲノムの情報量は約750MBとなり、CD1枚分にあたる。ヒトゲノムは私たちの細胞一つひとつが持っているもので、私たちの体を構成する小さな細胞一つひとつに、CD1枚分ものデータが格納されているということになる。

ゲノム解読が始まった年である1985年においては、生き物のゲノムを構成する膨大な塩基配列を読み取っていく工程を、一文字一文字手作業で地道に行っていた。この手法では、1つの研究室で1日あたり1000文字ほどしか塩基配列を読むことができず、ヒトゲノムをすべて解読するためには8000年もかかってしまう。

ところが、1990年代前半にシークエンサという機械が登場し、それまでの地道な手作業が自動化された。現在、ゲノム解読の技術のめざましい発展とその高速化により、1つの研究室で1日に300億文字以上もの塩基配列を読むことができるようになった。

これにより、たくさんの人のゲノムの解読や、ヒト以外の生き物のゲノムの解読が容易にできるようになった。そして、塩基配列のデータが大量に得られるようになった。

この大量のデータを処理するために計算機が必要となり、1990年代後半にバイオインフォマティクスという分野が生まれた。バイオインフォマティクスは、塩基配列のデータを計算機を用いて解析し、利用可能な形に変えていくことで、近年の生命科学研究に大きく寄与している。

例えば、医学では、ある病気の患者と健康な人のゲノムの塩基配列を比較することで、病気を引き起こす遺伝子を特定することができるようになった。また、進化生物学では、ヒトと別の生き物の塩基配列を比較することで、生き物同士の遺伝的な近さを解明し、進化の過程や共通の祖先を知ることができるようになった。

さらに、塩基配列のデータを応用し、商業的に利用する試みも進んでいる。実際に、顧客の唾液からゲノムを取り出して塩基配列を調べ、さまざまな病気へのかかりやすさや自分の先祖のルーツを割り出すサービス、脂肪の燃焼度合いを調べて一人ひとりに合ったダイエットプランの提案をするサービスなどが行われている。

このように、ゲノムの塩基配列のデータから、有用な情報を取り出すことが可能になる。瀬々先生は、データから情報を取り出すための技術の開発、そしてそれを利用した解析を行なっている。では、瀬々先生の具体的な研究を見ていこう。

パスウェイと発現量

先生の研究をイメージしやすくするために、がんという病気を例にとって説明する。

がんの原因は、がん細胞である。がん細胞は、細菌やウイルスのように外から身体の中に入ってくるものではなく、自分の体内にある普通の細胞が変異したものである。

私たちの身体の中では、毎日数千億個もの細胞が自身の遺伝子をコピーし、その遺伝情報をもとに新しい細胞を作り出している。私たちが文字を書き写すときに書き間違えをしてしまうことがあるように、たくさんのコピーが行われると細胞が遺伝子の一部をコピーし間違えてしまうことがある。細胞分裂を司る遺伝子のコピーを間違えてしまうと、この遺伝子が異常を起し、際限なく分裂し続ける細胞ができてしまうことがある。これが、がん細胞である。

がん細胞は、本来その細胞が体内で担うべき役割を果たさずに、正常な細胞の栄養分を奪うことで無秩序に増殖し、生き物を弱らせていく。

がんを治すための研究は世界中で何十年にもわたって続けられてきた。そして、どの遺伝子の異常ががんを引き起こすかも次々と分かってきている。それにもかかわらず、がんを根本的に治す方法はまだ確立されていない。なぜ、がんは治せないのだろうか。

その理由を知る一つの鍵がパスウェイにある。一つひとつの細胞内ではたくさんの化学反応が連鎖的に起こっており、遺伝子の発現が化学反応の制御を行なっている。これらの化学反応を信号ととらえ、遺伝子間で信号が伝わる様子を模式的に表したものがパスウェイである。

本来、パスウェイには、ある遺伝子 α の発現量(コラム参照)が多くなると、別の遺伝子 β がそのことを感知してその発現を抑える、フィードバックという機構が備わっている。ところが、遺伝子 β が異常を起してしまうと、遺伝子 α の発現を抑えるものがなくなり、発現が止まらなくなってしまう。つまり信号が永遠に出続けてしまうのだ。

がん細胞では、細胞の分裂を抑える命令を担う

遺伝子が異常を起こしているので、分裂するための信号が絶えず出続けている状態になってしまっている。このために、がん細胞は無限に増殖し続けてしまうのだ。

異常な信号をピンポイントに止める薬を作めることは可能である。しかし、実際にはこれを止めても、がん細胞はネットワーク上に別のパスウェイを見つけ、そのパスウェイを経由して異常信号を送るようになってしまう。さらに、止められたパスウェイの先からも信号を送ることができるようにパスウェイを作り変えてしまう。

したがって、がんを治すにはひとつのパスウェイを止めるだけでは不十分であり、パスウェイ全体のつながりを把握した上で対処しなければならないのである。

網目のように張り巡らされたパスウェイのネットワークは電車の路線図にたとえることができる(図2)。路線図には、どの路線を使ってどこへ行くことができるという情報が記されている。しかし、何時何分に電車がどこにいるのかといった、電車の具体的な時間や場所の情報は記されていない。これらの情報を知るには、時刻表が必要となる。パスウェイにおいては、遺伝子の発現によって合成されたタンパク質の発現量がこの時刻表のヒントとなり、いつ、どこで、どの遺伝子がはたらいっているのかという情報を与えてくれる。瀬々

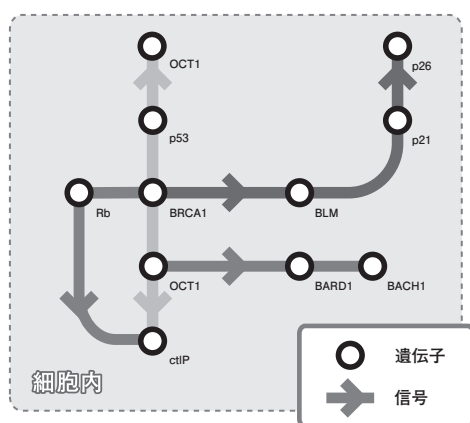


図2 パスウェイ

細胞内では、まるで駅の間を走る電車のように、たくさんの遺伝子の中で信号が伝わっていく。

先生は、遺伝子の細胞内における発現量を比較するための手法作りをしている。

例として、細胞の酸化ストレスを調べる研究について説明しよう。酸化ストレスが加わると、細胞内で酸化反応が次々と起こり、細胞内器官が傷つく。また、酸化ストレスは、私たちが試験を前にして緊張しているときに細胞に加わるストレスに似ている。私たちがストレスを感じたとき細胞の中では何が起きているのかが、バイオインフォマティクスによってわかるようになるのだ。

まず同じゲノムをもった細胞を2つ用意して、ひとつはそのまま、もうひとつには過酸化水素水を加えて酸化ストレスを与え、2つの細胞内でのさまざまな遺伝子の発現状況の違いを調べる。

このときの正常な細胞における発現量を x 軸に、酸化ストレスを与えた細胞における発現量を y 軸に取ったグラフが以下の図である(図3)。各点がそれぞれの遺伝子の発現量を表している。もし、2つの細胞の遺伝子の発現状況が全く同じであれば、すべての点が直線 $y=x$ 上にくるはずである。そこからずれている点があれば、正常な細胞と酸化ストレスを与えた細胞との間で遺伝子の発現量に差があるということになる。 $y>x$ の範囲にある点は、酸化ストレスを与えると発現量が増加する遺伝子を表し、 $y<x$ の範囲にある点は、酸化ストレスを与えると発現量が減少する遺伝子を表す。

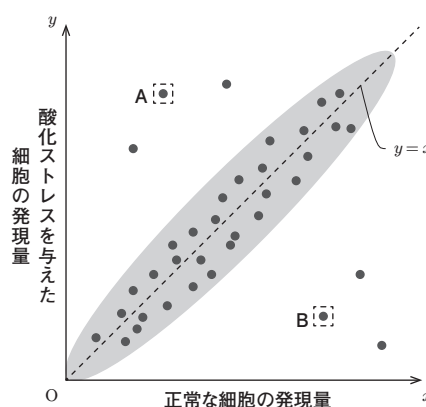


図3 細胞の遺伝子発現量の比較

軸は細胞内の遺伝子の発現状況を、点は各遺伝子を表している。
A: ストレスにより活性化した遺伝子
B: ストレスにより発現が抑えられた遺伝子

このようにして、細胞が酸化ストレスを受けたときにどのような遺伝子が活性化、または不活性化するかを知ることができる。

この例においては、酸化ストレスを与えた細胞と何もしていない細胞という2種類の細胞の比較であった。先生はこの手法をさらに発展させて、より多くの対象を同時に比較する多次元的な解析を行なっている。

多次元的な解析では、扱う次元の数が増えるので、グラフの軸の数が多くなる。2次元的な解析とは異なり、各点が各細胞を表し、各軸が各遺伝子の発現量の大小を表す。

解析の対象とする遺伝子は数千個から数万個の規模になり、人の手で調べることは到底できない。そこで、計算機の長所を生かし、数え上げや繰り返しを用いることで、膨大な数の遺伝子の発現量を細胞間で比較することが可能になる。先生は現在、この比較を速く処理するためのプログラムの開発を進めている。

この解析の技術は、さまざまな場面での応用が期待できる。例えば、ある病気の患者の細胞と、健康な人の細胞について、各遺伝子の発現量を比較する。その結果、患者の細胞でのみ発現量が極端に多かったり少なかったりする遺伝子があれば、その遺伝子はその病気と関係がある可能性が高いということがわかる。

将来的には、得られた病気ごとの遺伝子の発現量のデータと、受診者の遺伝子の発現量のデータを比較することにより、その人がその病気にかかっているかどうかを病院で診断することができるようになるのである。

また、先生は遺伝子の発現の量の情報とパスウェイの情報を結びつけていく研究も行なっている。発現量の情報(時刻表)をパスウェイのネットワーク(路線図)と合わせていくことで、同時にはたらくパスウェイを見つけようとしている。

発現量を調べてみると、一見何の関係もなさそうに見える2つのパスウェイが、同時に活発にはたらくしているということがある。まるで、路線図には載っていないトンネルが存在していて、路線図上では遠く離れている2つの駅がつながっているかのようである。

このような、同時にはたらくパスウェイを調べる研究は、将来的に薬の副作用の予測につながると先生は考えている。薬の副作用とは、薬が本来ターゲットにしている場所とは別の場所にも影響を与え、害を及ぼしてしまうことである。同時にはたらくパスウェイを見つけることができれば、あるパスウェイを薬によって人工的に調節したときに、他のどのパスウェイが同時に影響を受けるのかを知ることができる。このことによって、従来は動物実験などによって調べていた副作用を、薬を実際に生き物に与えることなく予測することができるようになるだろう。

バイオインフォマティクスの展望

ここまでで紹介した内容以外にも、先生は生き物から得られるさまざまなデータを用いた解析のための技術づくりを行なっている。

また、バイオインフォマティクスは幅広い分野を扱っていることから、先生の研究室は他の研究室との共同研究も盛んに行なっている。先生がコンピュータを使った解析を行うにあたり、実験系の研究者からデータの解析や、そのための技術の開発を頼まれることもあれば、実験系の研究者に実験をしてもらって必要なデータを取ってもらうこともある。実験系の研究者との協力が先生の研究にはとても大切である。

先生がバイオインフォマティクスというテーマを選んだ理由は、その分野の新しさと、できることの多さにある。まだ手つかずの、新しい発見の可能性を秘めたデータは日々生み出されている。

執筆者より

私は生命科学を専攻していますが、瀬々研究室を取材して、バイオインフォマティクスについて教えて頂いたことで、生命科学だけにとらわれない横断的な分野の面白さを知り、視野が広がりました。瀬々先生はとても気さくな方で、記事の内容にとどまらない、興味深いお話をたくさんして下さいました。本当にありがとうございました。

(畑中 恵菜)

発現

生き物の体内で絶えず起っているさまざまな生命現象はいずれもタンパク質が担っている。すべての細胞はタンパク質で支えられており、体内に取り入れたものを身体のあちこちに運ぶタンパク質も存在する。さらに、消化など、生体内で起こる化学反応の触媒である酵素もまたタンパク質である。

タンパク質は、すべての生き物が自身の設計図として細胞内にもっているDNAから作られる。DNAからタンパク質が作られるまでの一連の過程を発現という。

発現の過程を見ていこう。まず、DNAの塩基配列が、DNAによく似た物質であるRNAに写し取られ、mRNAという鎖が作られる。これを転写という(図-左)。次に、リボソームという細胞小器官がこのmRNAを雛形として読み取り、配列に対応したアミノ酸がtRNAという生体内分子によって運ばれてくる。これを翻訳といい、この

アミノ酸が一系列につながって、タンパク質ができるのだ(図-右)。

発現量とは、発現によって得られたタンパク質の量のことである。ある生き物を構成する細胞はすべて同じDNAをもっているが、すべての細胞で常に同じタンパク質が同じ量だけ作られているわけではない。細胞のある場所や、その生き物の年齢などによって、どのタンパク質がどれくらい作られるかが変化してくる。

例えば、ヒトの体内において、牛乳に含まれるラクトースという糖を分解する酵素であるラクターゼは、年をとるにしたがって作られなくなる。これは、母乳を飲まなくなることによってラクトースを分解する必要がなくなるためである。このため、子供の頃は牛乳をよく飲んでいたので、大人になるとお腹を下してしまう人が出てくる。生き物は、体の状態に応じて遺伝子の発現量を変え、生命活動を調節しているのだ。

