

バイオインフォマティクスとは何か？

大量データ時代の到来

近年の分子生命科学の大きな特徴の一つとして、計測技術の進展に伴う大量かつ多様なデータの生成が挙げられます。このような傾向は、70～80年代の塩基配列決定技術の進展に伴う配列データベースの構築に端を発し、その後のゲノムプロジェクトやポストゲノム研究が進む中で加速化しています。当面は次世代DNAシーケンサーがこのような傾向を牽引することになると思われます。分子生命科学に携わる研究者は、自身が網羅的（ゲノムワイド）な研究に関与していなくても、これらのデータと向き合っていかなざるをえない状況にあり、このデータの海から必要な情報を効率的に抽出できるかが、研究を成功させる重要なファクターとなってきました。しかし、そのような大量かつ多様なデータを人力で処理することは不可能であり、情報科学の助けなしに研究を進めることは困難になってきています。バイオインフォマティクスの生命科学における重要な役割の一つとして、情報の大海の中で研究者をナビゲートし、医薬、農学を含む生命科学の新たな知識獲得へ至る過程を手助けすることが挙げられます。

仮説構築による研究の加速化

バイオインフォマティクスが生命科学の重要な局面で大きな役割を果たした例は多数あります。近年、京都大学の山中 伸弥氏らは、ES細胞の多能性維

持に関する転写因子の「狩り」にあたり、実験による成果がなかなか得られなかったところを、データベースを利用した*in silico* 解析でブレイクスルーを得ています^[1]。また、創薬ターゲットとして新規サイトカインやその新規受容体をゲノムデータベースやESTデータベースから相同性探索によって抽出することに成功した例もあります^[2]。このような事例は多数見つけることができますが、これらに特徴的なことは、バイオインフォマティクスが仮説構築に利用されていることです。これまでの仮説検定型と異なり、バイオインフォマティクスは大量のデータから新たな仮説を導きだし、研究を新たなステージに引き上げることができます。仮説構築を言い換えると「予測」ということになります。ゲノム創薬におけるターゲット探索も仮説構築の一種といえるでしょう。仮説検定型研究では、研究者の経験や知識から立てられた仮説を検証する形で研究が進められます。しかし、自然はしばしば人間の憶測から外れた挙動を示すことがあり、その場合は仮説構築型の研究が有効で、バイオインフォマティクスはセレンディピティーの加速装置としてはたります。

「量」から「質」への転換

生命科学の情報爆発は始まったばかりであり、今後ますます大量のデータが生成されると予想されます。そのような状況では、いかに“量”を“質”に転

換できるかを考える必要があります。バイオインフォマティクスは、先に述べた仮説構築という形で“量”から“質”への転換を行っています。このような処理では、構築される仮説の品質が問われます。精度の悪い予測（仮説）は、かえって研究や産業化を遅滞させるだけです。また、“量”という観点からは、それを“質”に変換するのに要する時間も大きなファクターとなるでしょう。いかに精度の高い予測が得られるとしても、その計算に100年かかるようでは意味がありません。さらに、“量”を“質”に転換するという意味で重要なこととして、情報の統合が挙げられます。大量データには、ゲノム配列だけではなく、トランスクリプトームやプロテオームのようなさまざまな種類のものがあります。これらのデータはそれぞれ生命活動の1側面を切り出したものであり、これらを統合することで初めてシステムとしての生命の全体像がみえてきます。そのためには、これらのデータを統合的に扱える柔軟なプラットフォームの構築が必要になるでしょう。よりよい仮説をより迅速に構築する技術と情報統合の基盤の開発が、バイオインフォマティクスにこれまで以上に求められることになると思われます。

生命情報工学研究センター 副研究センター長
藤 博幸

参考文献

- [1] 田中 幹人：iPS細胞 ヒトはどこまで再生できるか？, 日本実業出版社(2008).
- [2] 野村 仁：ゲノム創薬 個別化医療とゲノムデータマイニング, サイエンス社(2005).