



ゲノム科学の研究に携わって

足立 淳（予測発見戦略研究センター ゲノム解析研究グループ）

生物のゲノム情報が解読されるようになって、生命科学の研究方法は急速に変わりつつある。多くの研究はゲノム情報を活用するのが当然となり、その大量のデータから必要な情報を抽出するために統計処理や情報処理の技術が必要不可欠となった。この流れは、我々統計科学の研究者にとっても応用分野が広がり、非常にありがたいことである。しかし、ゲノム情報を利用した研究は、生命科学の研究者との共同作業になることが多く、そこには幾つかの関門がある。私は、統計数理研究所に着任するまでの3年間、日本のゲノム科学の中心的な研究機関に在籍していた。そのおかげで、外にはわからない研究体制の現状にも触れることができた。その当時に実感した事のうち、巨大プロジェクト体制と共同研究の問題について感じたことを書いてみたい。

最近のプロジェクト研究体制の一番の問題は、研究の継続性を軽視しがちな点にある。研究資金の獲得競争の過熱から、流行を追うように新しいプロジェクトが立ち上げられ、以前のプロジェクトはより速く結果を出すことが求められてしまう。私が研究に携わったころは、ゲノムプロジェクトの全盛期であった。しかし、ゲノムのドラフト配列が公開されるとすぐに、ポストゲノムという言葉が一人歩きを始め、ゲノムプロジェクトはもう終わったかのような雰囲気を作られてしまった。ゲノム配列の公開はDNAの塩基配列を電子情報に起こしただけに過ぎず、遺伝情報の解読は始まったばかりだった。解読することが目的であったにも関わらず、ポストゲノムという言葉の登場と共に、次のプロジェクトの立ち上げが議論され、より精度の高い配列の決定とその解読への研究体制は資金の削減と共に縮小されてしまった。ゲノム配列は、今後の生命科学の研究の礎となる大切なものなのに、精度が十分でなく不完全な部分を残したままになっている。

ゲノムの次に脚光を浴びたのはcDNAプロジェクトであった。しかし、新たにタンパク立体構造プロジェクトが立ち上がると、cDNAプロジェクトは資金面の継続性の不安から質よりも量が優先されてしまう。同じことは繰り返される。SNP、遺伝子ネットワーク、RNAi等と流行が変わるかのごとくプロジェクトは入れ替わっていく。その結果として玉石混合な大量のデータが公共データベースに登録され公開されてしまった。プロジェクトに関わった者やその側にいた者は、それらのデータの質に様々な問題があることを知っているが、外部の者がそれを知ることは難しい。将来の生命科学の礎となるデータの現状に、我々のようなデータ解析者は細心の注意を払わなければならないし、また、その現状を変えていくように働きかける必要性を強く感じている。

ゲノム情報を前提とした研究では、生命科学の研究者と統計・情報科学の研究者との共同作業が必要不可欠となったが、上記のようなプロジェクト体制の下では障害も存在する。事実、プロジェクトでは多くの統計・情報系の研究者が採用されたが、そのような研究者の転職率は極めて高い。これには転職先のポストが多いこと以外に、幾つかの理由が考えられる。

一般に、生命科学者はある特別な現象を追い求め、一方で統計・情報科学者はより普遍的な法則を追求する傾向がある。この固有性と汎用性の志向の違い（または、実験系と理論系の違いとも考えられる）を、互いに良く理解していないと共同研究はうまくいかない。別な問題として、プロジェクト体制と競争の激化によってもたらされた研究スピードの変化もある。より早く結果を求めるあまりに、データの質を疎かにしがちな前者と、より美しい結果を求めるあまりにデータに質も量も要求しがちな後者の違いである。

プロジェクトの下では生命系の上司と統計・情報系の部下という組み合わせになることが多く、両者のバランスをとることが難しく、時には偏ってしまう。そういう意味では、組織内部にいるよりも、外部にいる我々のような研究者の方が、共同研究をうまくやれる可能性が高いのかもしれない。共同研究においては、両者の立場を理解した仲介役がいることが望ましいが、それは研究現場では理想論に過ぎない。現実には、どちらか一方がその役割を兼ねなくてはならないのである。そして、統計・情報科学の研究者が、その役割を果たすことが暗黙の内に期待されている。

その点では、我々のような共同利用機関の研究所にいる者に有利な点があるのではないだろうか。これまで様々な分野のデータの統計的解析を行い、共同研究の実績を上げてきた経験がある。相手の立場を理解した上で両者のバランスをとり、共存共栄の共同研究の実現することに馴れているからだ。

我々に期待されている役割が非常に大きなものであると痛感しているこのごろである。