はじめに

生物のもつ遺伝情報総体であるゲノム情報から，さまざまな生命現象を解明しようとする研究分野がゲノム科学（Genomics）です．次世代型DNAシークエンサー（NGS）の登場以降は，シークエンシングした塩基配列情報を生物学的な意味が付随する遺伝情報としてではなく，単なるシグナル情報として利用するRNA-Seq解析やChIP-Seq解析などの新しい研究手法も登場し，ゲノム解読だけでなく多様な目的にゲノム科学が応用されるようになりました．技術革新のペースはさらに加速していて，メタゲノム解析，Hi-C解析やシングルセル解析など，高度解析技術による新たな解析手法が猛烈な勢いで進展しています．これは科学や社会にとってはよいことなのですが，裏を返せば，これまで培ってきた解析手法が2～3年で陳腐化するということを意味しているわけで，研究者にとっては大きな負担となっています．この負担を軽減するために，さまざまな解析ツールが開発・公開されており，それらツールを支える理論やアルゴリズムを理解していなくても，コマンドを入力すれば，もしくはソフトウェアのボタンをクリックすれば，最新の解析結果を簡単に得ることができるようになっています．

さて，ゲノム科学における代表的な推論方法はアブダクション（仮説形成）なので，仮説を裏付けるために多様なデータ群が必要となります．特にNGS登場後は，マルチオミクスと称する膨大かつ多様なデータ群の解析が求められるため，解析アルゴリズムも多岐にわたり，また複雑さも増しています．しかし研究を進めるうえでは，使用するアルゴリズムの中身を理解しておく必要があり，単なるツールとして使い解析結果を出しただけでは誤った結論に至る可能性もあるし，そもそも科学的推論とは胸を張って言えなくなります．

かねてより私たちはこのような状況を背景に，生命科学系研究者の情報解析をプラットフォームとして支援しており（先進ゲノム支援，奥付プロフィール参照），その事業の一端として，プログラミング言語Pythonを用いた初級者向けおよび中級者向けの講習会を行っています．本書で扱うトピックの多くは，この中級者向け講習会をもとにしています．講習会で得られた学習のノウハウを詰め込みつつ，単にツールの使い方を紹介する便利な本ではなく，解析の本質やアルゴリズムを理解できる教科書にしたいと思いました．したがって，手っ取り早くツールの使い方を知りたい方には，冗長でほしい知識に簡単に辿り着けない面倒臭い本，として分類されてしまうかもしれません．自らプログラミングを実践し，解析を進め，結果を得る．その途上で，仮説形成のためになぜそのような解析が必要になるのか，目的のためにはどのようなアルゴリズムが有効なのか，などを理解しつつ，D. E. Knuth先生のThe Art of Computer Programmingのように，常に戻って参照できる，そこを基礎として発展できる，というような教科書をめざして編集しました．各章の文章は原則的に筆者の皆さんの原文ママとしています．その結果，章ごとに文体が異なる，少々「荒削り」な書籍となっていますが，著者の人物像を垣間見つつ，より臨場感をもって楽しく学習していただければと思っています．

最後に，講習会の際にいつも手伝ってくださっている国立遺伝学研究所 生命情報・DDBJセンターの皆様，ライフサイエンス統合データベースセンター（DBCLS）の皆様，先進ゲノム支援の皆様，さらには私どもの編集方針を理解し出版にまでこぎ着けていただいた羊土社の皆様に心より感謝申し上げます．

2021年2月

編集を代表して
黒川　顕