

# 序

この増刊号を手にする読者は、生命医科学の急激な変革、とりわけ大量データ処理の重要性を肌で感じているはずだ。「次世代」という言葉に象徴される技術革新は、従来の生命医科学におけるパラダイムをひっくり返した。これまではデータこそが最重要であった。大多数の人はデータを囲い込み（クローズド）、データの収集に比較すれば解析に要するコストは僅かであった。バイオインフォマティクスも、データの見栄えにかかわる程度の「おまけ」分野にみえた。しかし今は違う。ヒトゲノム1000人分のデータを誰でも無料でダウンロードでき<sup>1)</sup>、解析ソフトウェアもソースコードを公開して開発するオープン時代である。データは欲しいだけ入手でき、解析こそがボトルネックなのだ。加えて、これまでの解析結果が必ずしも信用できないこともわかっている。2004年に出版されたがんに関するマイクロアレイ研究のうち（有名雑誌も多く含まれる）、半数の統計結果が誤っているという報告がある<sup>2)</sup>。また、計測者や手法によるデータの偏りも正しく考慮されていない場合が多い<sup>3)</sup>。いま最も重要なのは、データ処理の過程を正しく把握し、利用可能な最新のリソースやソフトウェアを知ることに尽きる。

本書は、臨床家・研究者・教育者が座右のリファレンスとして使えるように企画・編集した。1章を読むと日本の主だったデータベースプロジェクトおよびそのビジョンを概観できる。2章と3章は主要なバイオリソース、4章には最新のソフトウェアが集めてある。執筆陣には生命情報処理の現場で活躍する気鋭の若手を中心に選び、各項目にリソースやソフトウェアの客観的な比較表をつけてもらった。2008年に出版された実験医学増刊号「バイオデータベースとソフトウェア最前線」（森下真一、阿久津達也／編）とあわせ、『使える一冊』を心がけたつもりである。読者の期待に応える内容になっていることを願うとともに、编者からの細かい注文に対応してくださった執筆者の方々に感謝したい。

2011年8月

有田正規

1) 1000人ゲノムプロジェクト：<http://www.1000genomes.org/>

2) Dupuy, A. & Simon, R. M. : Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. : J. Natl. Cancer Instit., 99 : 147-157, 2006

3) Leek, J. T. et al. : Tackling the widespread and critical impact of batch effects in high-throughput data. : Nat. Rev. Genet., 11 : 733-739, 2011