

特集「ゲノム情報」の編集にあたって

高 木 利 久[†]

ある生物の遺伝情報の総体をゲノムという。ゲノムは各細胞の中にある染色体に DNA の塩基配列の形で書かれている。分子生物学は、複雑にみえる生命現象もその基本的な仕組みはゲノムの構造 (= DNA 塩基配列) に記述されていることを明らかにした。ヒトゲノム計画の目的は、ヒトにおけるさまざまな生命現象をこのゲノムの構造情報から解き明かすことにある。ゲノムの構造情報は、転写・翻訳という過程を経てたんぱく質の構造情報に変換され、それらが生体内で空間的・時間的に相互作用することにより各種の生物学的な機能が実現される。ゲノムの構造情報から生命現象を読み解くには、これらの情報の変換や操作を規定する法則を明らかにしなければならない。

しかしながら、現状では構造から機能に至る道筋で成り立つ法則はほとんどわかっていない。そこで、ヒトゲノムを解読するために必要な生物学医学情報を網羅的に収集・整理・統合してデータベース化をはかるとともに、その上でさまざまな解析やシミュレーションを行い、ゲノムの構造情報から生物学的な意味を探ることが必要となる。これには最先端の情報処理技術が不可欠である。

平成3年度に始まった文部省ヒトゲノム計画 (<http://www.genome.ad.jp>) では、この計画において情報科学/工学が本質的な役割を担うとの認識から、その発足に際して、分子生物学者から構成されるゲノム解析研究グループとは別に、ゲノム解析のための情報処理技術を研究開発するためのグループを科学研究費補助金重点領域研究「ゲノム解析に伴う大量知識情報処理の研究」(略称:ゲノム情報)として設けた。

この重点領域研究は本年3月に5年間にわたる研究を終えたが、この間情報科学/工学の分野から多くの研究者がこのプロジェクトに参加し、有用なツールを提供するという面においても新しい方法論を開拓するという面においてもヒトゲノム計画の進展に寄与してきた。また、この重点領域研究で得られた成果や問題設定は情報処理の他の

研究分野へも少なからぬ影響を与えつつある。

重点領域研究「ゲノム情報」の終了を機に、情報科学と生物学とにまたがる新しい研究分野「ゲノム情報学」を情報処理技術研究の立場から総括するとともに、今後取り組むべき課題を明らかにすることを目的としてこの特集を企画した。

本特集は10編の解説記事よりなる。第1編ではゲノム情報学とはどのような研究分野であるか今後の目標や課題は何かを概観する。

第2編から4編までは、タンパク質のアミノ酸配列や立体構造の解析にそれぞれ並列最適化、幾何アルゴリズム、隠れマルコフモデルの観点から取り組んだ研究事例について紹介する。ゲノムの機能はおもにタンパク質として実現されるため、タンパク質における解析や予測はゲノム情報における重要な課題である。

第5編と第8編とは、RNAの二次構造の予測問題にそれぞれ形式言語理論、超並列計算機の応用の立場からアプローチした研究について解説したものである。第5編ではDNAの組み換え技術を利用した新しい計算方式いわゆるDNAコンピューティングについても紹介する。

第6編ではゲノムの配列データベースからその配列データの特徴づける仮説をいかに発見するかについて解説する。これには計算論的学習理論が用いられる。第7編では生命系の理解にはどのようなモデル化が必要か、また、それを用いていかにシミュレーションするかについて考察する。

昨年度で終了した重点領域研究「ゲノム情報」は、新たに発足した重点領域研究「ゲノムサイエンス」の中の「ゲノムの生物知識情報」班として本年4月に再スタートした。第9編と第10編は、今年度から新たに加わったメンバによる研究の概観である。ゲノム情報に対してこれまでとは異なる観点から取り組んでいる。

最後に、大変お忙しい中にもかかわらず、本特集の執筆、閲読、編集にご尽力いただいた多くの方々に深く感謝いたします。

(平成8年9月17日)

[†] 東京大学医科学研究所ヒトゲノム解析センター