



第34回

医学生物学における 次世代シーケンサー利用に関する解説

ペンシルバニア州立大学医学部ゲノム科学
バイオインフォマティクスコアディレクター 川沢(今村)百可 Yuka Imamura Kawasaki

はじめに

初の次世代シーケンサー(next-generation sequencer; NGS)が発売されたのは2000年初期ヒトゲノム解読とほぼ同時期である。サンガーシーケンス(first-generation sequencer)によってヒトゲノムが解読されるまで、所要期間13年、費用は3億ドルに上ったが、NGSの進化によって、今やヒトゲノムは1日1,000ドルで解読できる時代となった。

一方で、NGSはその用途の幅も広がり、数層の高かった解析手法も様々な改良がなされ、医学生物学において必要不可欠なツールとなってきた。筆者自身、開発途上のNGSマシンに触れ、2009年以降は一ユーザー兼プロバイダーとしてその進化を目の当たりにしてきた「生き証人」の1人として、NGSの実際と、現在の医学生物学にどう役立っているか、さらにどのような今後の発展が望めるか、全2回に分けて解説してみたい。

NGSの原理と実際

イルミナヤ454、サーモフィッシャーの次世代シーケンスは、サンガーシーケンスと同じSBS(Sequencing By Synthesis)の方法を採用している。一

本鎖DNAをテンプレートとし、ポリメラーゼを使って相補鎖を合成する際に生じる電気化学的シグナルから、該当するDNA配列を1塩基ずつ決定する。サンガーシーケンスはこれをSmall Scaleで行っていたが、次世代シーケンスでは短く断片化したDNAを大量並列に処理する。これにより高速にシーケンスすることが可能になり、NovaSeqやHiSeq X Tenなどの最新機種では1回のランで数兆塩基の情報を得られるまでになった。これは30億塩基のヒトゲノムを30倍の深度(depth)で読んだとして10人以上一度に解析できる量である。イコール、このレベルのランには1回当たり1万ドル以上はかかるので、オペレーターとしては緊張でピペットを握る手に冷や汗をかきながら実験することになる。

では何をどうやってシーケンスするかということだが、基本的に「短く切って(ショートリード)大量並列処理」という戦略はどのNGSメーカーにおいても同じである。この、短く切って後処理をしたDNA断片のことを「ライブラリー」と呼ぶ。使用するDNAやRNAが少ない場合はPCR反応を挟んで量を増やすが、大量であればPCRはしなくても構わない。ライブラリープレップの最後の段階で、大体6塩基~16塩基からなる