

次世代シーケンサーの動向と結核研究

結核研究所

生体防御部長 土方 美奈子

遺伝情報とシーケンサー

ヒトや結核菌のゲノムDNAの塩基配列（A, T, G, Cの並び）は、遺伝情報を保持している。結核菌ゲノムは、約440万個の塩基の並びから成る環状の2本鎖DNAで、細菌が分裂する時にゲノムDNAも複製されて受け継がれていく。DNA複製は写し間違えのないように工夫されているが、稀にミスが起きて塩基配列が変化することがあり、これが変異である。ゲノムDNAの違いを検出することで、結核菌か非結核性抗酸菌かの見分けや、北京型などの結核菌の遺伝系統の判断ができ、結核菌ゲノム配列の変異パターンを解析することで、分子疫学調査や薬剤耐性検査も行われる。

この塩基配列そのものを直接読み取る装置がシーケンサーで、最初の自動DNAシーケンサーが市販されたのは1986年である。第一世代とも呼ばれるサンガー法によるキャピラリーシーケンサーを用いた場合、30億塩基もあるヒトゲノムの塩基配列を決定するには年単位の時間と億単位の費用が必要だったが、2007年に塩基配列大量並列読み取り装置である「次世代シーケンサー（next-generation sequencer, NGS）」が登場し、大量の塩基配列データを高速・安価に得られるようになった^{1) 2)}。2021年にはヒトゲノムは、2日間のシーケンス、1人あたり十数万円近くの費用で解析できるようになっている³⁾。本稿では、世界の結核研究で非常に重要なツールとなっているNGSとさらに新しいシーケンサーの仕組みと使われ方について特徴をまとめる。

ショートリード・シーケンサー

DNAの複製時、元のDNA鎖を鋳型にして、Aに対してはT、同様にT→A, G→C, C→Gのルールで新しいDNA鎖が合成され、相補鎖と呼ばれる。イルミナ社のNGSでは、蛍光発光する塩基を用いて相補鎖を合成させ、1塩基ずつ色を読み取って塩基配列を決めるが、その過程には様々な工夫がある。まず数百塩基程度の長さにしたDNAをガラス板（フローセル）に貼り付け、DNAをフローセル上で増幅させて同じ配列を持つDNAの集合体（クラスター）を形成させる（図A）。クラスターの数は、NGSの機種によって400万から200億までの幅があ

り、これが各機種のデータ出力の違いとなる。クラスターの数だけ、同時並列で塩基配列（リード）を得ることができるが、最長でも600塩基までという長さの制限がある。大量の配列データの解析を行うにはコンピューター技術が不可欠で、コンピューターによって検体と標準ゲノムの塩基配列の違いを検出したり、リードをつなぎ合わせてゲノム全体を明らかにするなどの情報解析分野をバイオインフォマティクスと呼び、この解析技術の発展がNGSを支えている。配列解析の場合、短いリード配列で同じ場所が何回もシーケンスされていることで、正確な配列を決定することができる。

NGSのもう一つの強みは、大量の短いDNA断片の塩基配列を決定するので、ある配列が元のサンプルに何本あったか、という数の情報を得ることができることにある。例えば、結核患者の血液中のRNAをDNAに変換してNGSでシーケンスすれば、どのような種類のRNAの量が患者検体で多いのか検討することができ、バイオマーカー探索が行える。別の例としては、様々な細菌を含むサンプルから抽出したDNAを断片化してシーケンスすることで、どのような種類の細菌がどのような割合で含まれていたかを推定することもできる。NGSから非常に多彩なアプリケーションが生まれ、様々な研究分野で使われている。

一方で、イルミナ社のNGSの弱点はリードが短いことであり、例えば、結核菌ゲノム解析の場合、IS6110やVNTR、あるいはPE/PPE遺伝子内などの繰り返し配列は、数百塩基程度の断片では正確な解析が困難な場合が多く、また、ショートリードのデータから1本に環状化させた完全長ゲノム配列の決定をすることもできない。

ロングリード・シーケンサー

イルミナ型のNGSがショートリード・シーケンサーと呼ばれるのに対し、長い塩基配列を決定できるロングリード・シーケンサーが次に登場してきた⁴⁾。第3世代シーケンサーと呼ばれることもある。

Oxford Nanopore Technologies社が提供するロングリード・シーケンサーは、ナノポアというタンパク質でできた小さい穴が人工膜に埋め込まれており、穴にDNA

分子を通す時に塩基の種類の違いにより生じるイオン電流の変化によって、DNAの配列を解析する(図B)。電流値から塩基配列への変換には機械学習の手法が使われる。市販された当初は、手のひらに乗るサイズのMinIONという機種をUSB接続でパソコンにつなぎ、どこでもシーケンスできる新規性が大きな注目を集めたが、現在は高データ出力の大型機種もある。純度の高い、切断されていないDNAが抽出できれば、DNA分子の長さに応じて長い配列データを得ることができる。増幅などの過程を全く含まない、抽出されたままのDNAをナノポアに通せば、メチル化などの修飾状態も解析でき、さらにRNAを直接シーケンスすることも可能である。弱点は、シーケンスの精度がやや劣ることで、特にホモポリマーと呼ばれる同じ塩基の並びが苦手な、結核菌ゲノム研究では、多くの場合、イルミナの短い正確なリード配列とあわせて完全長ゲノム配列の決定に使われている。しかし、新製品の開発で精度は年々改善されてきている。

もう一つのロングリード・シーケンサーは、Pacific Biosciences社が提供するPacBioで、一分子リアルタイムシーケンスという技術を用いている。イルミナのシーケンサーが、クラスター形成過程でDNAを増幅して蛍光を検出するのに対し、PacBioの技術では、1分子のDNAの相補鎖合成時、取り込まれる瞬間の塩基を蛍光発光させて検出する(図C)。数十キロ塩基長のDNA断片のシーケンスができ、また、DNA修飾の検出もできる。さらに近年開発されたHiFiリードは、2本鎖DNAを環状のシーケンス鑄型にすることで、1分子のDNAを回転

させて何度も読み続け、そのコンセンサス配列を得ることで、各リード配列のエラーが修正され、このリードをさらに多数解析することで、長さの高い正確性を兼ね備えた方法となっている。

ヒトゲノムには、まだ配列解析ができていないリピート配列領域が数多く残っており、ロングリードによる解析で次々にリピート配列と疾患との関連が明らかになってきている⁵⁾。結核菌研究においても、ロングリードを用いて得られた完全長ゲノム配列が持つ情報は、参照配列へのショートリードのマッピングで解析されるゲノム情報より多く、リピート配列も含めて解析できるようになるため、さらに結核菌ゲノムの知見が深まることが期待されている。

シーケンサーと塩基配列解析の今後

シーケンサーの技術革新により、ヒト・病原体のゲノム解析、RNA発現解析、エピゲノム解析など、塩基配列から得られる情報は多岐にわたり、様々な疾病に関する新しい知見が得られてきた。本稿では2022年執筆時点のシーケンサーの基礎知識をまとめたが、これがあつという間に古くて役立たない知識になるくらい、シーケンサーの技術革新のスピードは速いものであり、おそらく今の予想を超える次のシーケンサーが登場する日も、そう遠くないのではないかと考えられる。🍵

参考文献:

- 1) Shendure J, et al. Nature 550:345-353, 2017.
- 2) Metzker ML. Nat Rev Genet 11:31-46, 2010.
- 3) www.genome.gov/sequencingcostsdata
- 4) Logsdon GA, et al. Nat Rev Genet 21:597-614, 2020.
- 5) Depienne C, et al. Am J Hum Genet 108:764-785, 2021.

