

次世代シーケンサーの医療への応用と課題

Next generation DNA sequencer

— Applications and the prospects for the clinical medicine

みずしま すが の じゅん こ すが の すみ お
水島一菅 野純子¹⁾ : 菅野純夫²⁾
Junko MIZUSHIMA - SUGANO Sumio SUGANO

要 旨

現在、医療現場で働く多くの方々には、次世代シーケンサーはあまりなじみのないものであり、次世代シーケンサーについての知識を持ってすぐには役立つかもしれない。しかし、次世代シーケンサーの登場は、病気の診断、治療、創薬に確実に変革をもたらし始めている。いまだに実現しておらず、現実的ではないと思われがちな個人化医療も、この高速シーケンサーの登場で夢物語ではなくてきている。本稿では次世代シーケンサーとは何であるのかという解説と医学へのかかわりについて述べる。

I. はじめに

次世代シーケンサーが登場したのは数年前のことである。その当時は、機器が高額であったこともあり、それほど普及せず、限られた研究室でしか用いられていなかった。しかし、現在（2011年8月）までに爆発的に使用が広まり、機器の性能のさらなる向上とあいまって、ゲノム科学および疾患研究に質的な転換をもたらしつつある。

次世代シーケンサーはアメリカのベンチャー企業が開発したもので、アメリカを中心にイギリスなどのヨーロッパ諸国や中国などがこぞって機器を購入し、これを利用した研究を積極的に進めている。中国は、特に大量の次世代シーケンサーを買い込み、早速、中国人の個人ゲノム（全遺伝情報）を決定した。米国に次ぐ業績である。中国のゲノム研究は主にBGI社

（旧名：北京ゲノム研究所、Beijing Genomics Institute）が先導して行っており、BGIは2010年には128台もの次世代シーケンサー（HiSeq™2000（イルミナ社の最新型シーケンサー））を購入している。機器のほとんどは香港のBGIゲノムセンターに設置され、世界で最も多くの次世代シーケンサーが設置されているセンターとなっている。しかし、日本で次世代シーケンサーを設置している研究室は数が少なく、現在でも日本全国合わせて100台程度である。また、次世代シーケンサーを用いて成果を上げている研究室となるとさらに少ない。世界の潮流に乗り遅れてしまった感のある日本であるが、最近になって、癌ゲノムの解析で成果をあげるなど、徐々に挽回してきてはいる。

II. 次世代シーケンサーとは どういうものか？

それでは、次世代シーケンサーとはどういうものなのであろうか。従来型のシーケンサーと比べて何が違うのだろうか。次世代シーケンサーがあまり取り入れられていない日本は、ゲノム研究のどのような点で世界に遅れをとる恐れがあるのだろうか。

まず、塩基配列決定法の原理であるが、従来型のシーケンサーではサンガー法が用いられているのに対し、次世代型ではこの方法を用いていない。サンガー法とは、ジデオキシヌクレオチドを用いてDNAポリメラーゼの伸長を止めるチェインターミネーター法である。一方、次世代シーケンサーで

1) 工学院大学工学部応用化学科特任教授
東京大学大学院新領域創成科学研究科
メディカルゲノム専攻、ゲノム制御医科学分野非常勤講師
2) 東京大学大学院新領域創成科学研究科
メディカルゲノム専攻、ゲノム制御医科学分野教授
〒108-8639 東京都港区白金台4-6-1

Laboratory of Functional Genomics, Department of Medical
Genome Sciences, Graduate school of Frontier Sciences,
The University of Tokyo
(4-6-1, Shirokanedai, Minato-ku, Tokyo, JAPAN)

は、機種ごとに個別のシーケンシング原理が用いられている。Illumina社のSolexaゲノムアナライザーおよびHiSeq™2000 (Solexaゲノムアナライザーの改良型機種)では合成シーケンシング法¹⁾、454 ライフサイエンシズ社 (後にロシュ・ダイアグノスティクス社が買収)のシーケンサー GSLXおよびFLXシステムではパイロシーケンシング法²⁾、ABI社 (後にInvitrogen社と合併後lifetechnologiesに名称変更)のSoLiD™および5500xlではリガーゼ反応シーケンシング法³⁾が用いられている(図1)。

従来型とは全く異なるシーケンシング原理を採用したことにより、塩基配列決定量が飛躍的に高まった。詳細は省くが、ヒト一人のゲノムを読むのに要する時間と費用で比較してみると次のようになる。人類が初めてヒトゲノムを解読したのは2003年のことで、13年間かかり、3千億円(30億ドル)の



図1 次世代シーケンサー

(ロシュ・ダイアグノスティクス社、ライフテクノロジーズ社、イllumina社の好意による。)

費用を費やしている。2007年5月にはアメリカでジェームズ・ワトソン博士(DNAの二重らせん構造の解明によりノーベル賞を受賞)個人のゲノムが解読された。2003年の人類初のヒトゲノム解読では、数人のゲノムが混じった状態で解読されたが、ワトソン博士のゲノム解読は個人のゲノム情報が解読されたものとしては世界初の出来事である。この時は次世代シーケンサーで解読を行い、たった2カ月、1億2千万円(100万ドル)足らずで読み切っている。この時に使われた機種は、米国のバイオ企業で当時の454 ライフサイエンシズ社の次世代シーケンサーであった。現在(2011年8月)では、それがわずか1カ月の期間と約百万円の費用で個人ゲノムが解読可能となっている。次世代シーケンサーの登場で、1塩基あたりの配列決定費用は数千分の1と大幅に低下している。その性能は、今現在もすさまじい勢いで向上を続けており、今後5年ほどで、個人のゲノム解読は1時間で終わり、数万円でできるようになると言われている。

個人ゲノムとしては、アメリカのワトソン博士のゲノム解読に続き、中国人、アフリカ人、韓国人などのゲノム解読が行われた。日本人の個人ゲノムはこれに遅れ、2010年になってはじめて解読されている。大ざっぱな言い方をすると、次世代シーケンサーは、1台で従来型シーケンサーの200~1000台分の能力があるといえる(図2)。従来型はシーケンシングに電気泳動を必要としたが、次世代型は必要としないため1回に解析できるサンプル数が非常に多いことがその要因であろう。

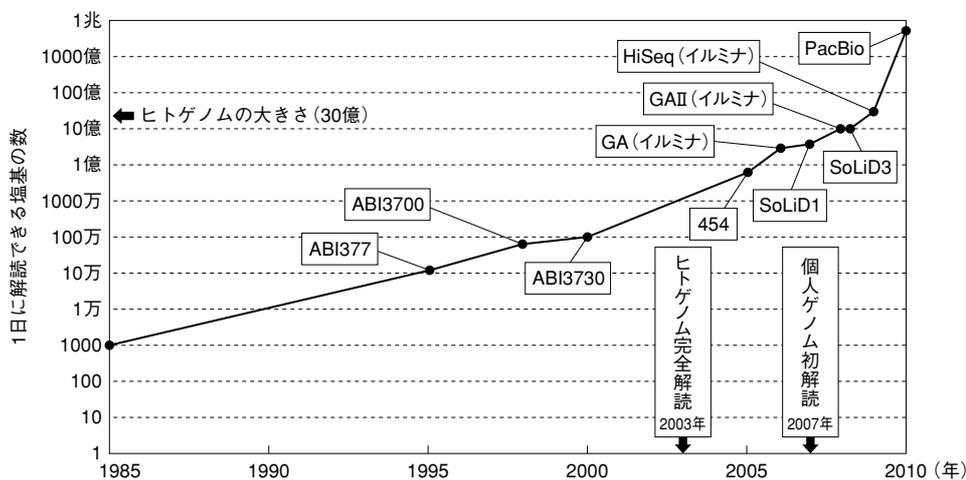


図2 DNAシーケンサーの性能の向上

Ⅲ. 癌ゲノム解析

次世代シーケンサーの登場により塩基配列決定能力が飛躍的に伸び、大きく前進しつつあるのが癌研究である。癌は遺伝子の疾患である。厳密には、ゲノムからの遺伝情報発現が異常になることが癌の原因で、体細胞ゲノムの塩基配列に変化(変異)が起こることによる。遺伝性乳がんなど、全身の細胞が変異している遺伝性の癌もあるが、癌全体の中に占める割合はごくわずかで、ほとんどの癌が遺伝せず、体細胞の遺伝子異常がその原因である。

最近になって、癌の遺伝子異常は非常にさまざまであることがわかってきた。大腸癌を例にとると、同じ大腸癌でも塩基配列変化が起きている箇所が個人によって異なることが明らかにされた。これが、大腸癌にさまざまな性質のものがあることの原因であると考えられている。また同じ個人でも、大腸癌の癌部分の中で、場所によって起きている変異が異なっているという報告もある。したがって、一概に大腸癌といっても、実際にはさまざまな原因によって癌が生じており、おのずと治療効果のある抗がん剤も異なることになる。ここでは大腸癌を例にあげたが、これは肺癌や乳がんを始めすべての種類の癌に共通していえることである。そこで、さまざまな癌の組織からDNAを抽出して癌細胞と同じ個人の正常細胞の全ゲノム配列を網羅的に解析することが行われている。癌細胞と正常細胞のゲノム配列の中でどこがどのように違うのかを解析すれば、癌化のメカニズムの解明につながるというわけである。これには次世代シーケンサーの性能がいかに発揮されている。この研究により、癌の悪性度、治療反応性、予後等の指標となる変異が明らかになると予想される。この指標が得られれば、癌の治療方法をより有効なものとするができるだろう。

Ⅳ. 疾患関連遺伝子を見つける

もう1つ、次世代シーケンサーの登場で大きく前進した研究が遺伝性疾患の原因となる遺伝子の探索である。疾患の原因となる遺伝子を見つけるには1塩基多型(Single Nucleotide Polymorphisms (SNPs))などを使った遺伝学的な解析が必要であ

る。それにより、原因遺伝子の存在する遺伝子領域が明らかになるが、そこから遺伝子の変異を見つけ原因遺伝子を同定するには、大量のシーケンシングが必要で、次世代シーケンサーの登場以前は、時間がかかった。それが大幅に短縮されたのである。

さらに、SNPなどを使わずに、直接ゲノム配列決定を行い、その情報を使って遺伝学的解析を行うことで、原因遺伝子を見出す試みが始まっている。この場合、ゲノム全体ではなく、エクソン(exon:ゲノム上で最終的に成熟 mRNA となる部分)部分のみを取り出してシーケンシングすることで、コストを10分の1にできる。このため、タンパク質の構造を変化するような変異による遺伝病は、急速にその原因遺伝子の同定が進みそうだ。エクソン部分はすべてを合わせると約5千万塩基対でヒトゲノム(30億塩基対)の60分の1だが取り出すにはコストがかかる。それでも全ゲノム配列を決定するよりは、はるかに低コストでできる。

エクソン部分のみを取り出してシーケンシングすることを exome シーケンシングというが、これは多因子疾患の疾患感受性遺伝子の探査にも使用されつつある(図3)。従来、多因子疾患を対象にした全ゲノム解析(GWAS (Genome Wide Association Study)(ジーバスと発音する)またはWGA study (WGASとも略される)(Whole Genome Association Study)と呼ばれている)はSNPを用いて行われてきた。これを exome シーケンシングで行うことで、それまで解析できなかった稀な変異も見出すことが期待できるのである。もっとも、数百人、数千人レベルの exome 解析が必要なので、簡単ではない⁴⁾。

Ⅴ. エピゲノム解析

最近になって、癌や精神疾患などの病気にエピジェネティックな変化が大きくかかわっていることが明らかになってきた。エピジェネティックな変化とはDNAの塩基配列変化を伴わないが、遺伝情報発現が変化することである。この変化は細胞が分裂した後も、娘細胞に引き継がれるのが特徴である。分子レベルでどのような変化が起きているかという、DNAのメチル化とヒストンタンパク質の化学修飾である。DNAのメチル化ではシトシン残基(C)にメチル基が導入される。ヒストンタンパク質とは、

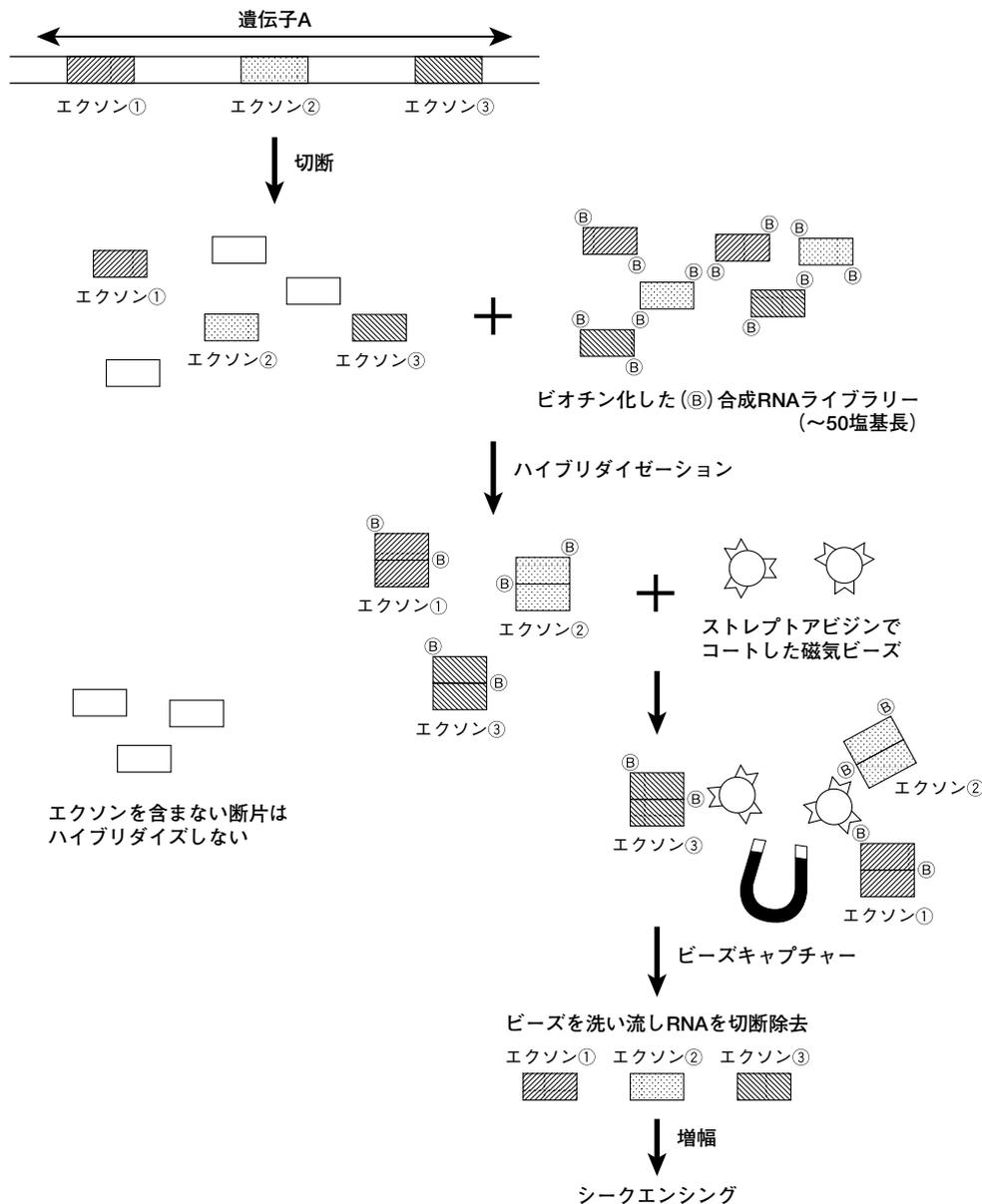


図3 エクソーム配列解析の方法

その周りにDNAを撒きつけてクロマチンすなわち染色体を構成している塩基性のタンパク質である。このヒストンタンパク質にアセチル化、メチル化、リン酸化、ユビキチン化などの修飾が起きることによりクロマチン構造が変化し、遺伝情報発現が促進されたり、抑制されたりする。DNAがメチル化されると、遺伝情報発現は抑制される。例えば、癌細胞ではメチル化異常が起きており、脱メチル化剤やメチル化阻害剤の中には、血液のがんの治療に有効なものもある (HDAC 阻害剤やDNAメチル化阻害剤)。

エピジェネティックな変化の解析はゲノム科学の進展とともに加速度的に進んだ。次世代シーケン

サーが登場してからは、ゲノムの一部分ではなく、全ゲノム領域にわたってエピジェネティックな変化を調べることができるようになった。これをエピゲノム解析と呼ぶ。DNAのメチル化を調べるには主にバイサルファイト・シーケンス (BSP) 法が、ヒストン修飾状態を調べるためにはChIP-seq法が主に用いられる。どちらも最後にシークエンシングをする必要があるが、次世代シーケンサーの登場によって、解析スピードが著しく上がった。今後、この分野は病気との関連性、創薬において、ますます重要性を増してくるであろう。

VI. 個人化医療（テーラーメイド医療）の実現に向けて

ゲノム科学が発展し、個人個人の体質や病気の質に即した、いわゆる個人化医療の実現が夢物語ではなくなっている。地球上には約67億人もの人々が暮らしているが、各個人が独自のゲノム配列をもつ。すなわち、約67億通りのゲノム配列が存在することになる（一卵性双生児は除く）。これらを個別に読むことができれば、癌や糖尿病などへの罹りやすさがわかり、生活習慣を改善するなどして病気を未然に防ぐことも可能となるだろう。また個々人の薬物代謝酵素遺伝子（P450など）の配列から薬の効き方の予測ができ、抗がん剤の副作用を最小限にしたり、個々人に適正な量の投薬をすることも可能となる。

ここで、注意しなければならないことは、倫理的、法的、社会的な問題（Ethical, Legal and Social Issues (ELSI)）に備える必要があるということである。例えば、病気の原因となる遺伝子を見つけるためには上述した通り、患者さんと健常者のゲノム配列の比

較が必須であるが、究極の個人情報であるゲノム情報の取り扱いには細心の注意が必要である。コンピュータからの情報流出などはあってはならないから、ゲノムの塩基配列情報を入力するコンピュータはインターネットに接続しない、などの措置がとられる。また、血液サンプルや組織サンプルの採取の際にはどのような研究に用いられるのか、研究に使われることによって個人が不利益を被る危険性はないのか、など十分な説明がなされる必要がある（インフォームドコンセント）。

文 献

- 1) Ronaghi et al. "Improved performance of Pyrosequencing using single-stranded DNA-binding protein". *Analytical Biochemistry* **286** (2): 282, 2000.
- 2) Bentley, D.R. et al. "Accurate whole human genome sequencing using reversible terminator chemistry" *Nature* **456**: 53-59, 2008.
- 3) <http://www.appliedbiosystems.co.jp/website/jp/multi/page.jsp?CD=116437>
- 4) Li Y et al. Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants". *Nat. Genet.* **42**: 969-972, 2010.