

次世代シーケンサーへようこそ!

最新版はこちらからダウンロード



illumina®

次世代シーケンサーの誕生

すべての生物学研究において、ゲノム配列を理解することは重要です。キャピラリー電気泳動を用いたサングーシーケンス法の開発により、様々な生物システムの遺伝情報の解読にアクセスすることが可能となりました。この手法は世界の多くの研究室で幅広く使われていますが、データのスループット、拡張性、スピード、解像度、コストといった面で制限があり、研究者のプロジェクトを遂行する上ですべての必要な遺伝情報を引き出すことは難しいものでした。これらの制限を打開するために、次世代シーケンサーが誕生しました。次世代シーケンサーは以前の手法とは大きく異なるアプローチで、非常に数多くの画期的な発見をもたらし、ゲノムサイエンスに大きな革命をもたらしたのです。

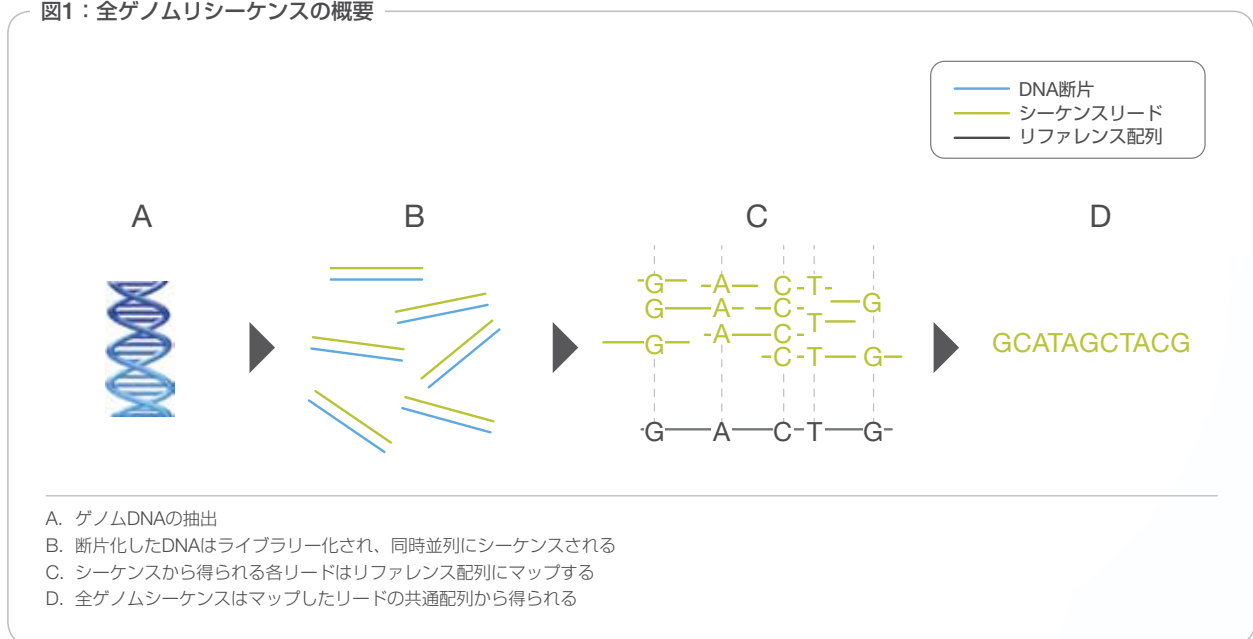


次世代シーケンサーへようこそ

次世代シーケンサーが誕生して5年が経過し、我々は大きな研究手法の変化を目の当たりにしました。次世代シーケンサーが登場したことで研究者は生物システムから遺伝情報を抽出し、ゲノムだけではなくエピゲノム、トランスクリプトームなどの分野において無限の見識を明らかにすることができるようになったのです。次世代シーケンサーの原理は、基本的にはキャピラリー電気泳動を用いたサンガー法に似ています。DNA断片をテンプレートとし、1塩基ずつ再合成する時の蛍光強度を検出し、塩基配列を決定します。従来のサンガー法との違いは、サンガー法では1~96のDNA断片を同時処理するのに対し、次世代シーケンサーでは数千万から数億のDNA断片に対して大量並列に処理します。これによりシーケンス解析のスピードは飛躍的に向上し、大きなゲノム領域を対象とする研究ができるようになりました。現在、次世代シーケンサーの上位機種では1回のランで実に数千億塩基の情報を得られるまでとなりました。

実際のステップをわかりやすく考えるために、ひとつのサンプルをイメージします。サンプルからゲノムDNAを抽出し、さらに大量並列シーケンスが精度よく均一に行われるように断片化し、ライブラリを作成します。このライブラリのDNA断片は次世代シーケンサーによって解読されますが、各断片の塩基配列の単位はリードと呼ばれます。ゲノム配列がすでにわかっている生物種の場合はこのリードをリファレンス配列にマップして比較するリシーケンス解析を行い、ゲノム配列のない生物種の場合は新たにアセンブルを行い配列決定をする *de novo* シーケンス解析を行います (図1)。

図1：全ゲノムリシーケンスの概要



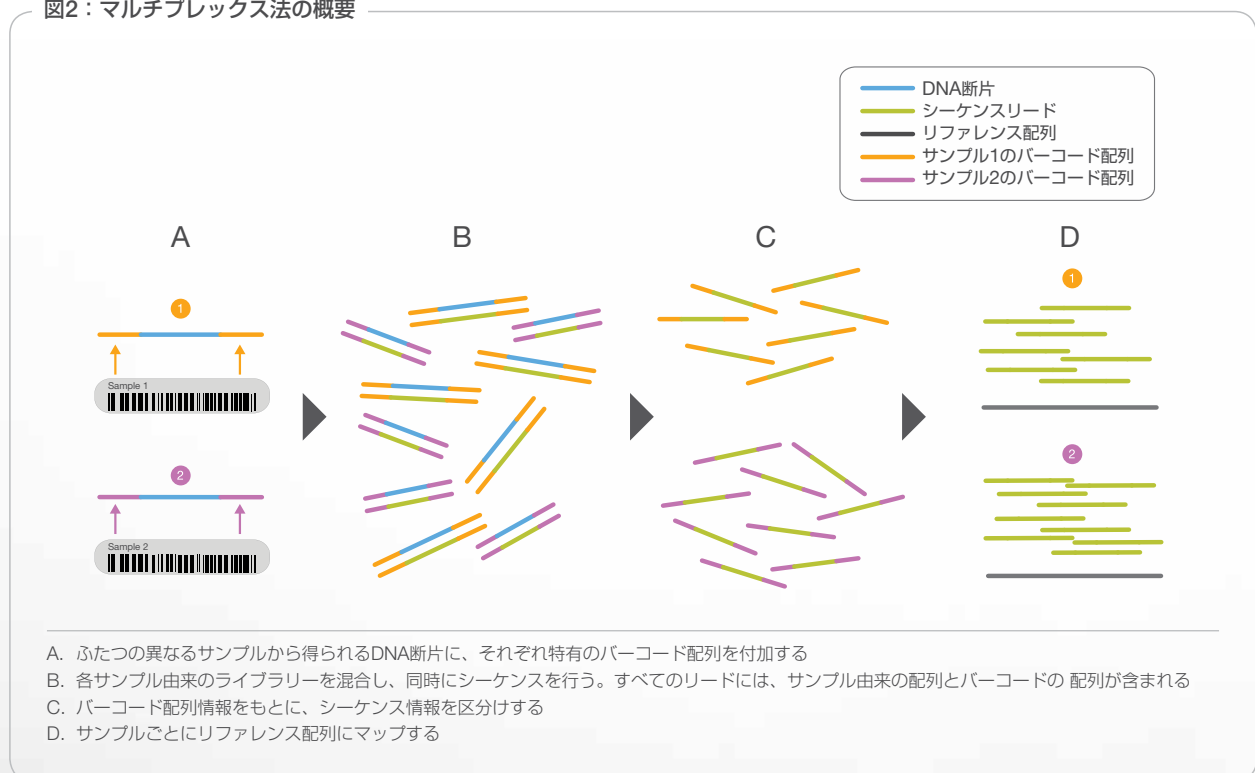
- A. ゲノムDNAの抽出
- B. 断片化したDNAはライブラリ化され、同時並列にシーケンスされる
- C. シーケンスから得られる各リードはリファレンス配列にマップする
- D. 全ゲノムシーケンスはマップしたリードの共通配列から得られる

高スループットで行うサイエンス

次世代シーケンサーの出力データ量はムーアの法則を上回るスピードで上昇しており、テクノロジーが誕生してから毎年倍以上のペースで増加しています。次世代シーケンサーが誕生した2007年は1回のランで10億塩基（1 Gb）のデータを得ることができました。それが、2011年には1000億塩基（1 Tb）に届くまでのデータを産出し、わずか4年で1000倍もの増加を達成したのです。迅速に大規模な量の配列解析能力をもつ次世代シーケンサーは、わずか数時間から数日で、研究者のアイデアを実際のデータセットにすることを可能にしました。現在、次世代シーケンサーを使えば1回のランで5名のヒト全ゲノム解析を約10日間、70万円で行えますが、2003年に終了したヒトゲノムプロジェクトで初めてヒト全ゲノムの解読を行った時は、以前のキャピラリー電気泳動のサンガーシーケンスを用い、配列情報の産出に10年、さらに解析に3年かかりました。このプロジェクトには3000億円（3兆ドル）近くもの予算が投じられました。

大型の次世代シーケンサーは大規模なデータを産出する能力がありますが、一方で次世代シーケンサーは拡張性のあるテクノロジーでもあります。同じケミストリーを使って、小さなゲノム生物種の解析や、ターゲット領域にフォーカスした用途に使うデータ量を押さえた機種も誕生しています。こうしたシステムの拡張により、研究者のニーズに応じて柔軟に機種を選ぶことが可能となりました。小さいバクテリアやウィルスゲノムの解析、あるいはエクソンにターゲットを絞ったプロジェクトを行う研究者は、小型の次世代シーケンサーを用いてランあたり複数サンプルの解析を行うことができますし、また大型の次世代シーケンサーとマルチプレックス法を使えば大規模なサンプル数を一度に処理することもできます。マルチプレックス法とは、1回の実験で多数のサンプルを同時にシーケンス解析を行う手法です（図2）。この手法にはバーコードの配列が各サンプルに付加され、データ解析時にそのバーコードに応じてサンプルごとの解析を行います。

図2：マルチプレックス法の概要



マルチプレックス法を使うことで、次世代シーケンサーは多数サンプルを対象とするプロジェクトでも、データ取得までの時間を大幅に削減することに成功しました。キャピラリー電気泳動シーケンサーでは、数百のPCR産物をシーケンスするには数週間から数ヶ月かかりましたが、次世代シーケンサーを使えば解析を含めて2日間で終了します。自動化されたシンプルなプロトコールとの併用で、研究者は実験から成果の発表まで、より速く、より簡単に行えるようになりました（表1）。

表1：イルミナ次世代シーケンサーとキャピラリー電気泳動を使ったサンガーシーケンス法の比較

テクノロジー	必要 DNA 量	ランあたりのサンプル数	ランあたりの時間	リード数	ランあたりのデータ量	ランあたりのデータ量	アプリケーション
キャピラリー電気泳動を使ったサンガー法	1-3 µg	1-96	0.5 時間	550 塩基 [†]	1-96	0.55-52.5 kb	DNA シーケンス、リシーケンス、マイクロサテライト解析、SNP ジェノタイピング
			3 時間	900 塩基 [†]	1-96	0.9-86.4 kb	
イルミナ MiSeq システム	50 ng (Nextera)	フローセルあたり 1 レーン	4 時間	36 塩基 x1 [§]	~700 万 (シングルリード)	~2 Gb	DNA シーケンス、遺伝子調整解析、トランスクリプトームの定量解析、SNP 探索、構造解析、細胞遺伝解析、DNA タンパク相互作用解析 (ChIP-Seq)、メチル化解析、Small RNA 探索と解析、 <i>de novo</i> 、メタゲノム、メタトランスクリプトーム
	0.1-1 µg (TruSeq)		27 時間	150 塩基 x2 ^{**}			
イルミナ HiSeq システム	50 ng (Nextera)	フローセルあたり 8 レーン、2 フローセル使用で 16 レーン	1.5-11 日	150 塩基 x2 [‡]	30 億 (シングルリード)	~600 Gb	
	0.1-1 µg (TruSeq)						

* キャピラリー電気泳動ベースではロード前のサーモサイクラーを使った4時間のシーケンス反応は含まず。NGSではシーケンスおよび塩基検出はランで同時に行われる。

† クオリティ値がQ20の塩基対

§ > 90% の塩基対がQ30以上のクオリティをもつ

** > 75% の塩基対がQ30以上のクオリティをもつ

‡ > 80% の塩基対がQ30以上のクオリティをもつ

拡張性のある解像度

次世代シーケンサーは実験に応じて解像度を柔軟に設定することができます。データ量を決めるのは研究者で、ゲノムのある特定領域にフォーカスをしてより高解像度な解析を行ったり、少ないデータ量でざっくりと全体像を把握する解析を行ったりすることができます。解像度の調整は、ある実験におけるシーケンスのカバレッジをどれぐらいにするのかを決定することから始まります。ここでいうカバレッジとは、ある生物種のDNAサンプルをシーケンスする際、その生物種のゲノムサイズの何倍量のデータを得ることができるかを意味します。例えば、ゲノムに対して平均30xカバレッジで解析したいとすると、ゲノム上の各塩基は平均30の冗長性でシーケンスされ、カバーされていることとなります。

このようにカバレッジ、あるいは解像度を研究者が自由に決定することができるので、様々な実験デザインを描くことができます。例えば癌研究において、ある組織サンプルにおける数～数十パーセントの細胞にのみ存在する体細胞変異を検出したいとします。癌サンプルには癌や正常細胞など異なる種類の細胞が混在するため、ある特定領域におこる低頻度の変異を検出するには、高いカバレッジ（例 1000x）のシーケンスを行わなくてはなりません。次世代シーケンサーではこれが可能となりましたが、キャピラリー電気泳動シーケンサーの場合、このような解析は可能ではあるものの高額なコストがかかり、多サンプルを解析する拡張性のあるプロジェクトを行うのは現実的ではありません。

研究者が全ゲノムを対象とした多型の探索を行う場合は、低めのカバレッジで解析を行うことができます。低い解像度で探索を行うかわりに多くサンプルを解析することで、ある特定集団における多型検出を高い統計力をもって実施することができます。

制限のないダイナミックレンジ

次世代シーケンサーはリード数をデジタルに数値でカウントします。これによりダイナミックレンジが広がり、遺伝子発現解析では非常に高感度の定量解析が可能となります。従来のマイクロアレイ手法と比べ、次世代シーケンサーはより高い解像度で転写産物の活動を定量することができ、生物学的なプロセスに応じてわずかに変化する遺伝子発現を捉えることができます。マイクロアレイ手法は連続的なシグナル強度を測定するので、低いレンジではノイズにより、また高いレンジでは飽和により検出が制限されます。一方で次世代シーケンサーはリードを個別のデジタルカウントとして捉えて定量を行います。実験ごとにシーケンスのリード数を増やしたり減らしたりすることで、研究者は研究の目的に応じて実験の感度を調整することができます。

あらゆる目的に使える生物学ツール

強力で柔軟性の高い次世代シーケンサーは、多くの研究に広がりをもたせ、生物学研究において必要不可欠な万能ツールとして定着しつつあります。どの生物種においても遺伝上の基盤を解析能力をもつ次世代シーケンサーは、研究デザインを大きく変える多様なアプリケーションの開発の原動力となっています。これまでの限界を超え、以前では考えられなかった情報を得ることができるようになったのです。

次世代シーケンサーを使って、あなたの研究では何ができるようになるのでしょうか。

