

# Nature 誌ヒト全ゲノムリシーケンス論文の筆頭著者に特別インタビュー

## 次世代シーケンサーの世界展開と日本の進む道

Illumina Inc. シーケンス部門最高統括責任者 David Bentley  
聞き手：イルミナ株式会社

医学、生物学の研究を大きく変えるツールとして高い関心が寄せられている次世代シーケンサー技術。日々開発が進むこの技術の今後の発展と、それを日本でどのように使いこなしていくべきか、英国ウェルカムトラストサンガー研究所でヒトゲノム部門を率い、数多くの国際プロジェクトに参加し、現在イルミナ社にてシーケンス部門最高統括責任者を務める David Bentley 博士にお話を聞きました。

博士は、2008年11月7日に開催されたイルミナ Genome Analyzer セミナーのために来日されました。その前日（11月6日）の Nature 誌にヒトゲノムリシーケンスの論文が掲載された博士に、絶好のタイミングでインタビューできました。



David Bentley 博士

英国ウェルカムトラストサンガー研究所に13年間所属し、Human Genetics部門代表として国際ヒトゲノムプロジェクト、SNPコンソーシアム、HapMapプロジェクトに貢献する。2005年にGenome Analyzerコア技術を開発したSolexa社に移り、'07年からIllumina Inc.にてシーケンス部門最高統括責任者となる

### ゲノム研究における日本の位置づけ

——2008年11月6日付けのNature誌で、米国、英国、中国からヒト全ゲノムリシーケンスが3報続けて発表されました。ご自身もそのうちの1報の筆頭執筆者ですが、日本の次世代シーケンサーを用いたゲノム研究が世界と比べて出遅れているといわれている風潮に対して、日本の強みと、それを今後どう生かしていくべきかを、ヒトゲノムとHapMapという2つの国際プロジェクトのリーダーだった観点から、お聞かせください。

まず、どの国が進んでいる、遅れている、という考えは適当ではありません。特に日本のように革新的な研究開発をこれまでに行った国に対しては、Nature誌で全ゲノムリシーケンスを発表した英国ウェルカムトラストサンガー研究所、中国北京ゲノム研究所、米国ワシントン大学は以前から新しい技術を評価し、ヒトゲノムに特化したプロジェクトにフォーカスを定めて研究を続けたグループです。今回のNature誌での発表は、その姿勢を反映したものといえます。ただし、世界には違う分野に特化して研究が続いているグループが多数あり、日本にも、他のゲノム解析分野でリーダーとなる研究グループがいくつかあります。

例えばDNA解析でも、さまざまな国際プロジェクトに日本は貢献しました。ヒトゲノムプロジェクトに協力した機

佳之先生、国際HapMapで1研究所として最大の貢献を果たした中村祐輔先生率いる理化学研究所などが世界的にも有名です。中村先生は特にこの分野で何年も優れた研究をしており、世界でもリーダーとして認められている研究者です。遺伝多型と疾患にフォーカスした中村先生の研究およびDNAサンプルの収集は多くの情報に富み、優れた資産でもあります。これらは日本だけでなく海外でも貴重な価値をもつサンプルです。こうした日本の研究における先見的視野、リーダーシップ、資産などは将来を見据えたものであり、今後のゲノム解析の革新的な研究成果につながるものです。あるグループが1つのことに専念し、成果を発表する、その間に他のグループは他のことに専念し、異なる成果を発表する、というだけなのです。

RNA解析においても日本のグループが世界でユニークかつリーダーとなる存在となっています。理化学研究所の林崎良英先生のグループが取り組んでいるゲノムアノテーションは、まさにDNAとRNAのシーケンス情報が融合した成果といえるでしょう。転写開始サイトの探索は非常に重要な研究対象です。というのも、遺伝子調節や制御といったことがここで行われており、ゲノムがどのように機能を持定し、それがうまくいかない時に何が起ころのか、ということ深く観察できるからです。林崎先生も、やはり前を見据えた研究を行い、遺伝子構造を理解する分野で世界をリードしています。

東京大学の菅野純夫先生も、日本を代表するRNA解析のリーダーです。菅野先生はわれわれの技術が世に出る1年も前から、学会などで次世代シーケンサーが研究にもたらす重要性を提唱していました。RNAにおけるシーケンス情報は生命科学研究の中核をなす情報になると、数年前に予測し、細胞が環境にどう適応し、どう遺伝子を使うのか、という点において現在精力的に研究を進めていらっしゃいます。

## シーケンサー技術に最も重要なこと

——現在のイルミナ Genome Analyzer のコア技術を開発したのは Solexa 社ですが、何が一番の技術的なブレイクスルーだったのですか？

すべての技術が重要ですが、1つ挙げるとしたら可逆的ターミネーター法でしょう。ターミネーター法はすでにサンガー法でも使われていましたが、可逆的にこれを使う（効率よく意図的に反応の停止、開始を行うことができる）という技術はこれまでになく、きわめて革新的なものでした。基盤にDNAを貼り付け、ゲル泳動することなく1分子でサイ



イルミナ Genome Analyzer セミナーにて講演中の David Bentley 博士

2008年11月7日に行われた「イルミナ Genome Analyzer セミナー」にて、11月6日付Nature誌でも発表されたアフリカ人の全ゲノムリシーケンスの結果を発表する Bentley 博士

クルごとに正確なシグナルを読み取り、塩基配列を決定する。これを大量のDNAに対して同時並行に反応を行い、既存技術と比べて、100倍の化学反応を100分の1のコストで行う。これを可能にしたのが可逆的ターミネーター技術でした。

——第3世代シーケンサーでは、1分子シーケンスが話題になっていますが、ご意見をお聞かせください。

過去30年の間、われわれ研究者はDNAシーケンスにたった1つの方法を使ってきました。それがサンガー法です。それと比べて、われわれは現在、多様な技術、手法を目の前にしています。ただ、どの手法も目的は1つで、「シーケンスをいかに低コストで、精度よく行うか」にあります。多くの人が新しいアイデアへの挑戦と失敗を繰り返しています。決して、平坦な道のりではありません。

イルミナ社は1分子シーケンス技術から開発を進めましたが、読み取り精度を追求するために、増幅ステップを行



### イルミナ Genome Analyzer システム

Solexa社の技術を用い、高精度、高スループット、低コストを実現。どの規模の研究室でも実験できるようにデザインされた次世代シーケンサーシステム

う決断をしました。正確なヒトゲノム、トランスクリプトーム、エピゲノム解析は、正確な配列決定から成ると考えたからです。第2世代、第3世代、第4世代と、いろいろな技術が登場してくるでしょう。しかし、重要なことは、「いかに難しい技術を開発するのか」ではなく、「いかにして正確な配列決定を、どの規模の研究室でも安定して行えるか」にあるのです。次世代シーケンサーの到来により、実験手法の中核がシーケンスとなりつつある今、これはとても重要な意味をもちます。

1分子シーケンスというのは非常に魅力的ですが、根本的な欠点をもっています。それは、1分子しかシーケンスできないことです。これは裏を返せば、もし1分子シーケンスが失敗すれば、結果を得ることができない、ということです。1つのミスも許されないのです。イルミナ社は、このために1分子シーケンスを行う技術をもっていながら、1分子シーケンスを行わない選択をしました。

とても重要なことなので繰り返しますが、DNAシーケンスは、「正確なシーケンスをいかに低コストで行うか」が重要です。これは今後も変わることはありません。サンガー法は、過去30年にわたり使用され、検証された方法で、いろいろなことがわかっています。30年という経験は、たやすく捨てられるものではありません。イルミナ社はこの経験から学び、この経験から技術をさらに改良する道を選びました。それが可逆的ターミネーター法です。発表後わずか18カ月でこの技術は大きな成長を遂げ、世界中で幅広く使われています。それを実証したのが前述のNature誌の論文です。「より正確な情報を低コストで得る」という観念は、第2世代も、第3世代も、その後の世代でも変わるものではない、変わるべきものではないのです。そのため、研究者

は第2世代、第3世代という言葉に捕らわれるのではなく、その技術で何ができるのか、どれだけ塩基を読めるのか、どれだけ精度よく解析できるのか、どれだけ低コストに実験が進められるのかを、じっくりと考えるべきです。われわれ研究者にとって大事なことは、難しい技術を開発することではなく、開発された技術をどのように使って研究を進めるのかにあります。そしてそのためには、優れた性能をもつ技術を見極めることが重要なのです。

## イルミナ Genome Analyzer の今後

——イルミナ Genome Analyzer の今後について、お聞かせください。

イルミナ Genome Analyzer は拡張性が高いシステムです。現在の装置で、第2世代、第3世代、そしてその先につながる改良の余地を十分に含むシステムです。現在は、シーケンス精度を高め、スループットを増やし、低コスト化することを目標に、日々開発が進められています。

精度と読み取り長でいうと、2008年6月の時点で、イルミナ社内では50塩基ペアエンドを安定して行っていました。現在は75塩基ペアエンドを日常的にこなしています。イルミナ社は1,000人ゲノムプロジェクトに参加していますが、そのデータでは精度をさらに高めることにも成功しています。以前は配列クオリティ値（Q）がQ20以上の塩基が全体の90%以上を、Q30以上の塩基は50%以上を占めていましたが、現在はQ30以上の塩基が80%以上のデータを得られています。

フローセルのイメージ解析、ソフトウェアの改善も進めています。これにより、クラスター密度を上げ、ランあたりのデータ量を上げることが可能になります。フィルター



後に1億リード(ペア両端で計算すると2億リード)が読み取れるとすると、50塩基ペアエンド法では10 Gb、75塩基ペアエンド法では15 Gb、100塩基ペアエンド法で20 Gbのデータを得ることができます。

また、コスト面でも改良を進めています。ヒトゲノムリシーケンスを行ううえで、一番重要な点は、正確に、精度よく遺伝多型や構造変化を検出することにあります。低コスト化が進んでも、この点が損なわれるのでは意味がありません。精度よくこうした遺伝情報を引き出す十分なカバレッジ数を考慮したコスト計算が重要になります。現在、市場で低コストを積極的にアナウンスしているところもありますが、正確性のよいデータを出すのに十分なカバレッジ数を得られているか、きちんと確認する必要があります。イルミナ社では現在約25万ドルでヒトゲノムのリシーケンスを行うことができます。これが今後5万ドル、2.5万ドル、と下がり、かなり近い将来に1万ドルにできると考えています。もちろん当面の目標である1,000ドルゲノムのことを考えると、製造コストなど挑戦すべき点は多数ありますが、思った以上に早く1,000ドルゲノムが可能になる日が来るのではないかと感じています。

——どうもありがとうございました。

David Bentley 博士のシーケンスにかかる情熱を感じました。何度も何度も、「シーケンス技術自身が大事なのではない、大切なのは、正確性、低コスト実現のためのスループット、そしてどの研究室でも使えるようにすることだ」と繰り返していたことが印象的でした。その目標のために、イルミナ社が1分子シーケンス技術をもちながら、それを使わない選択を行ったことも初めて聞きました。こういった高い目標と情熱に研究開発が行われていることが現在そして今後のイルミナ Genome Analyzer を支えていくのでしょう。イルミナ社の今後の活躍をご期待ください。



イルミナ株式会社 (Illumina K. K.)

〒100-0005 東京都千代田区丸の内2-2-2  
丸の内三井ビル2F

TEL : 03-5252-7771 FAX : 03-3218-0080

URL : <http://www.illumina.co.jp>

E-mail : [contact@illumina.co.jp](mailto:contact@illumina.co.jp)





# i can

3つの理由から、Genome Analyzerの導入を決めました。  
データポイントあたりのコスト、実験のワークフロー、  
そしてデータ情報の使いやすさ。  
Genome Analyzerのおかげで、小さな研究室を  
ゲノムセンターに変えることができたのです。

Greg May, Ph.D.  
Director, Genome Sequencing Center  
National Center for Genome Resources

次世代シーケンサーは、Genome Analyzer.

Next-gen sequencing  
*now*
[www.illumina.co.jp/](http://www.illumina.co.jp/)

# SOLiD™3 システムで、遺伝子解析の新たな研究戦略を立ててみませんか？

アプライドバイオシステムズジャパン株式会社  
SOLiD プロダクトマネージャー 稲垣美香

一般的に物事に対する概念を変えることはたやすいことではありませんが、あるきっかけで容易に変わることもあると思います。今、遺伝子解析の世界で多くの研究者にそのきっかけを与えているもの、それがアプライドバイオシステムズ社の次世代シーケンサ「SOLiD™システム」であると感じています。

## 新しい時代の幕開け

私が次世代シーケンサに携わるようになったのは今から2年半前の2006年の春のことでした。その時、次世代シーケンサの登場で「なにか大きな変化が起こる」ということを想像するのは難しくありませんでした。次世代シーケンサは米国国立衛生研究所（NIH）が提唱した1,000ドルゲノムプロジェクトを達成するための1つのツールとして市場に登場しました。この次世代シーケンサは、シーケンサに対する概念を変える非常に大きな技術革新でした。

アプライドバイオシステムズ社の次世代シーケンサ、SOLiD™システムは、1リードあたり35塩基～50塩基という短い配列情報を、1ランで数千万～数億本を産出することができます。この能力は、従来の「塩基配列を読む」というシーケンサの機能だけではなく、リード（タグとよびます）を数えるといったタグシーケンシングを可能にします。つまり次世代シーケンサはアレイ技術の代わりになるツールとしても位置づけることができます。このタグシーケンシングが多くの研究者の希望の扉を開け、さらに研究の戦略さえも変えるきっかけとなったと私は感じています。



SOLiD™3 システム

## 次世代シーケンサで重要なこと

SOLiD™システムの大きな特徴の1つに、超ハイスループットであることがあげられます。2009年初春に発売予定の「SOLiD™3 システム」は、200億塩基（20Gbp）/ 4億タ

表 SOLiD™3 システムデータアウトプット

ライブラリタイプ	解読塩基長	スライド枚数／ラン	マッピング可能塩基／スライド	マッピング可能塩基／ラン（2スライド）	タグ数／ラン	日数／ラン
フラグメントライブラリ	50塩基まで 75塩基*	1～2	50億塩基以上 (> 5 Gbp)	100億塩基以上 (> 10 Gbp)	2億以上	約3～5日（35塩基の場合） 約6～7日（50塩基の場合）
メイトペアライブラリ	2×50塩基まで	1～2	100億塩基以上 (> 10 Gbp)	200億塩基以上 (> 20 Gbp)	4億以上	約8～9日（2×35塩基の場合） 約12～14日（2×50塩基の場合）

\* R&Dによる実証値 ※最新の処理能力などの仕様に関しては、アプライドバイオシステムズ社（inagakmn@appliedbiosystems.com（稲垣））までお問い合わせください

グ以上を1ランで産出する能力をもつ、業界でも類を見ない次世代シーケンサです(表)。すでにR&Dグループにおいては、400億塩基(40Gbp)/ランも実証されています。SOLiD™システムではシステムの大規模なアップグレードなしに、スライドガラス上のビーズ密度を上げることで処理能力の向上を実現することができます。この処理能力は、ゲノムシーケンシングにおいて、数十カバレッジというデータを1ランで、しかもより低価格に解析できることを意味し、ヒトゲノムを含むさまざまな生物種のゲノム解析について、従来ではあまり現実的でなかった「ゲノム全体をシーケンスする」ことを可能にします。さらに、ChIP解析、メチル化解析やSAGE、CAGE、トランスクリプトーム解析などの遺伝子発現解析においても同様のことがいえます。特にマイクロアレイ解析では解析困難だったゲノム全体を解析することが可能になり、さらにより多くのタグ数を産出する能力から、従来ではみえなかった部分や低発現遺伝子を検出することができるようになりました(図)。これらの結果は、従来みていたものは氷山の一角であったことを気づかせるものであり、新しい発見との出会いでもあったと、SOLiD™システムのユーザは語ります。水面下に隠れた氷山全体の大きさを、誰がわかっているのでしょうか？処理能力が高いということは本当の全体像をみるために必要不可欠なことであると私は思います。

さらに産出されるデータ量というものは、データそのものの精度に密接にかかわっています。より高い精度のデータを大量に産出すること、それがデータ解析の時間を削減し、より短時間で研究成果をあげる近道なのです。SOLiD™3システムがもち合わせる高い精度と無比の処理能力は、従来法ではできなかった新しい研究戦略の確立へと導かせるのです。

## 次世代シーケンサを提供する企業として

膨大なデータを産出する次世代シーケンサを利用して研究成果を出していくのは容易なものではないと思います。アプライドバイオシステムズ社では、End to End Solutionをモットーとして、できるだけ研究者の負担を軽減し、意味のあるデータを短期間で取得していただけるように、アプリケーションごとにサンプル調製からデータ解析までの一連のソリューションを提供しています。現在では、アンビオンブランドをベースにしたRNA解析をはじめ、複数のアプリケーションをEnd to Endでサポートしています。

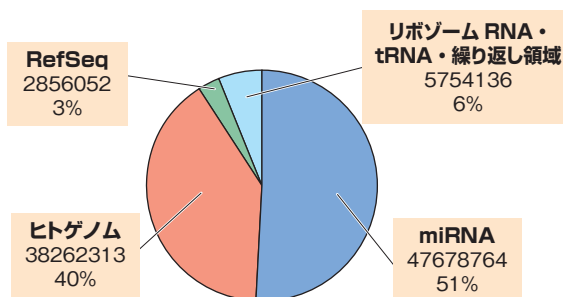


図 Small RNA Expression キットを用いたmiRNAの探索とプロファイリング

SOLiD™ Small RNA Expression キットを用いて調製したサンプルをSOLiD™システムで解析し、個々のデータベースで1カ所(ユニーク)にマップされた9,400万個のタグを分布しました。40%以上のタグが、既知のRNA種にマッピングされず、ゲノム領域にマッピングされました。また、miRNAにマッピングされず、ヒトゲノムにマッピングされたタグのサブセットは、未だ特性が決定されていないmiRNAに相当する可能性があります。

また、現地法人のアプライドバイオシステムズジャパン社では、システム導入にあたっての、設備関係のコンサルティング・サンプル調製やデータ解析などを全面的にバックアップしています。質のよいサポートができるのも、アプライドバイオシステムズジャパン社のラボにSOLiD™システムを導入し、実際にデータを出している成果だと考えています。

## 最後に

SOLiD™システムで研究をされている日本国内のユーザから「驚いた」「すごいね」という言葉を頂戴します。これがSOLiD™システムのパワーであると思います。みなさんもSOLiD™3システムで新しい研究戦略を立ててみませんか？より詳しい情報は、下記枠内のWebサイトより入手可能です。ユーザの生の声、原理がよくわかるムービーなどもございます。また、SOLiD専任技術者をご訪問してご説明いたしますので、お気軽に弊社担当までお声をおかけください。



アプライドバイオシステムズジャパン株式会社

〒104-0032 東京都中央区八丁堀 4-5-4

TEL : 03-5566-6100 (代表)

URL : <http://www.appliedbiosystems.co.jp/website/SOLiD.html>

E-mail : [inagakmn@appliedbiosystems.com](mailto:inagakmn@appliedbiosystems.com) (稲垣)