

生物系

次世代シーケンサーによる日本人の全ゲノム配列の初めての包括的解析

独立行政法人理化学研究所 ゲノム医科学研究センター
情報解析研究チーム チームリーダー
角田 達彦



【研究の背景】

一人一人がもつDNAの塩基配列の違いにより発症リスクが異なる病気の解明に向けて、ゲノムワイド関連解析が爆発的に行われています。それは例えばある病気になった人達の多くがそうでない人達に比べ共通に持つ塩基の違いに着目し解析する効率の良い手法ですが、さらにより細かな多様性も解析する次世代の方法として、「全ゲノムシーケンス解析」が期待されています（図1）。しかし、未だに精度良く解析する方法が確立していない上、日本人のゲノム解析についてはまだ報告がありませんでした。

【研究の成果】

我々は今回、次世代シーケンサーを使い、日本人男性1人の全ゲノム配列を取得、高精度な解析を達成し、日本人のゲノム配列の多様性を初めて包括的に検出することに成功しました（図2）。まず、全ゲノムシーケンスデータにバイズ決定法という数学的手法を適用し、約313万個の一塩基多様性（DNA上の塩基ひとつの違い）を約99.9%の高精度で検出しました。そして、海外の別々の研究グループから報告されている欧米人、アフリカ人、中国人、韓国人の6人の全ゲノム配列と日本人の全ゲノム配列を比較し、これまで集団での解析では見失われていた一塩基多様性が個人個人には多いことを発見しました。また、高精

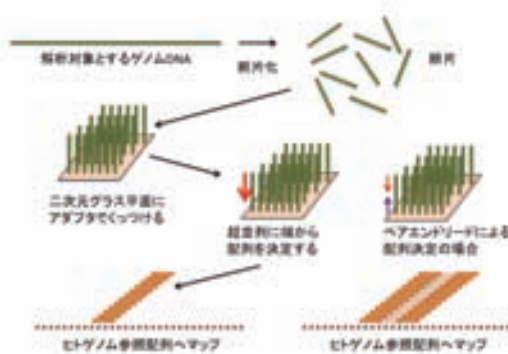
度な方法で1万塩基対より小さい欠失を約5,300個検出し、コピー数の多様性や構造的多様性も網羅的に見いだすことができました。さらに、ヒトゲノム計画で配列決定した参照配列にない約300万塩基対の新規配列を発見し、これらの配列がヒトゲノムの多様性を反映する可能性を見いだしました。

【今後の展望】

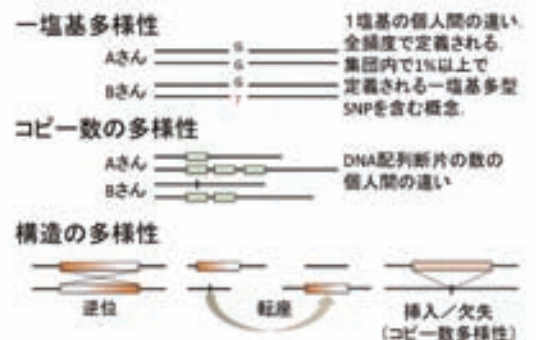
今回、個人のゲノム配列上には、病気に関わる希な多様性や、参照配列にはない新規配列が、未発掘のままである可能性があることがわかりました。全ゲノムシーケンス解析は、それらを完全に理解するために極めて重要な手法です。今後さらに日本人固有の多様性を検出することで、日本人のための病気の研究、そしてオーダーメイド医療が益々進展することが期待できます。また、この解析技術を駆使してがんに関わるゲノム上の包括的情報を解明し、日本も参加している国際プロジェクト「国際がんゲノムコンソーシアム（ICGC）」を進めていきたいと思ひます。

【関連する科研費】

平成22-26年度 新学術領域研究（『システムがん』）「がんのバイオインフォマティクスと遺伝統計学的解析」



▲図1 次世代シーケンサーによるシーケンス



▲図2 ゲノムDNA配列の多様性