



1人あたり約14万個所のDNA塩基多型情報を用いて日本人の集団構造を解明

— 病気と遺伝子の関連を調べるケース・コントロール解析のよりよい研究デザインが可能に —

病気のかかりやすさ、薬の副作用の有無は、私たちそれぞれで違います。この個人の体質の違いの遺伝的な要因を明らかにして、それぞれの人の特質に見合った治療を行うオーダーメイド医療がスタートしています。心筋梗塞、糖尿病、関節リュウマチなど、さまざまな疾患の原因遺伝子を同定するために、ゲノム全体をカバーする1塩基多型 (SNP) を使った関連解析が、わが国ばかりか世界規模で展開されています。わが国では、この研究の中心となっているバイオバンクジャパンが、約30万人の日本人のDNA、血清を集め、情報基盤としての役割を果たしています。

この研究を推進してきた理研ゲノム医科学研究センターは、これらの情報を活用して、大部分の日本人が2つのクラスター、本土クラスターと琉球クラスターに大別できることを明らかにしました。この結果は、従来から提唱されていた日本人集団の「二重構造説」と矛盾しない内容です。2つのクラスター間のSNPの頻度の違いは小さいのですが、耳垢のタイプを決めるSNPや髪の毛の太さに関わるSNPの頻度には明らかな違いがありました。また、本土の中でも遺伝的な地域差があることが明確になりました。

病気と遺伝子の関連を調べるケース・コントロール解析において、今回の研究結果を活かすことによって遺伝的背景のバランスを調整できますので、よりよい研究デザインが可能になります。

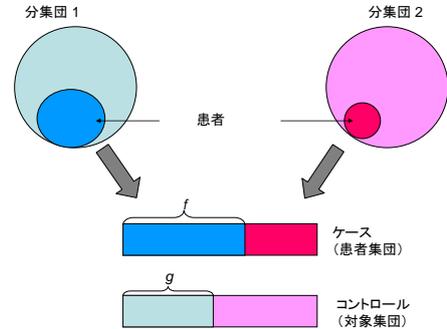


図1 ケース・コントロール解析で偽陽性率が上がる条件

分集団1、分集団2では、多数のSNPの座位(形質に関係のないゲノム領域のSNPの場所)について、SNPの頻度が少しずつ異なっている。この例では、ケース(患者集団)のサンプルにおける分集団1からの人の割合1(f)が、コントロール(対照集団)における分集団1からの人の割合(g)よりも大きい。このようなサンプルで関連解析を行うと疾患に関連していないSNPであっても、統計的に有意な関連(偽陽性)を検出する率が上がってしまう。

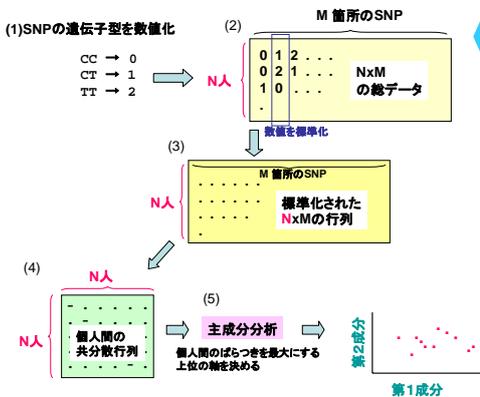


図2 たくさんの個所のSNPの遺伝子型を使った個人間の近縁関係の解析

個人からなるべくたくさんのSNPの遺伝子型を数値化(1)して、人数×SNPサイト数の総データを作る(2)。次に、統計解析のために数値を標準化する(3)。そして、個人対個人の総当りの、遺伝子型の相関係数などのデータを得る(4)。主成分分析により、個人間のばらつきを最大にする軸をきめ、上位の軸(例えば、第1成分と第2成分)でプロットを作成する(5)。第1成分の軸は、個人間のばらつき最大になるように選ばれている。第2成分の軸は、個人間のばらつきが次に大きくなるように選ばれている。

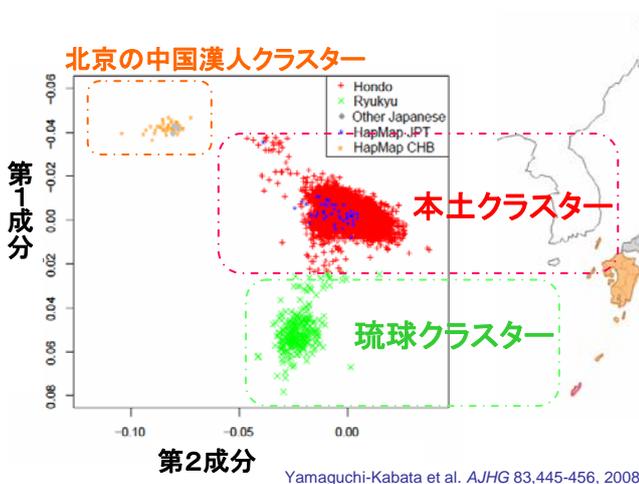


図3 東アジア人の近縁関係と3つのクラスター

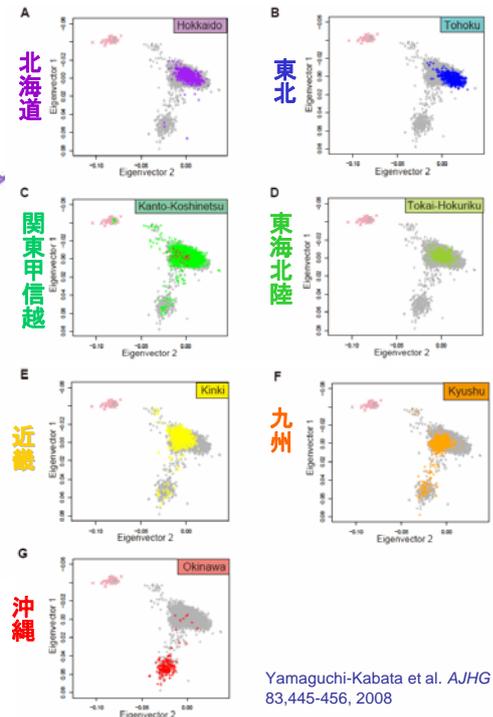


図4 それぞれの地域の人をハイライト

Yamaguchi-Kabata et al. AJHG 83,445-456, 2008

Yamaguchi-Kabata et al. AJHG 83,445-456, 2008