

2008年9月26日

独立行政法人 理化学研究所

## 1人あたり約14万個所のDNA塩基多型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っています。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。

今後は、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。

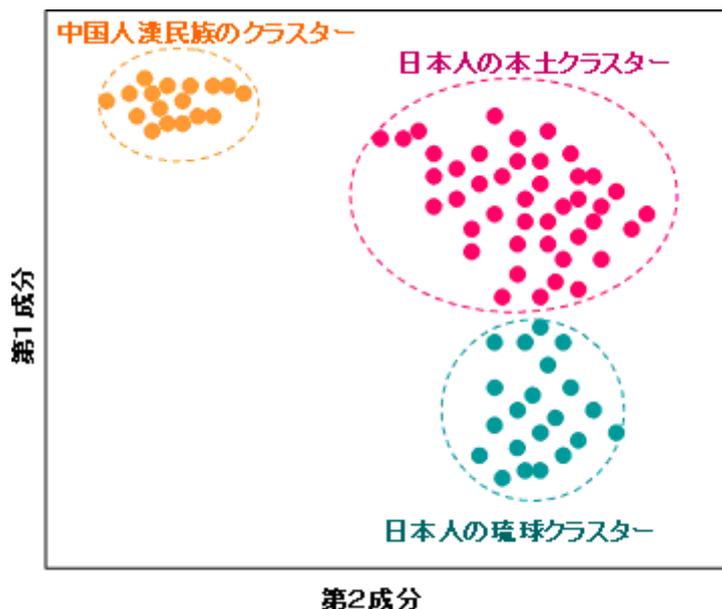


図 SNP を活用した日本人の集団構造の解析  
ほとんどの日本人は、本土クラスター（赤）  
と琉球クラスター（緑）に大別

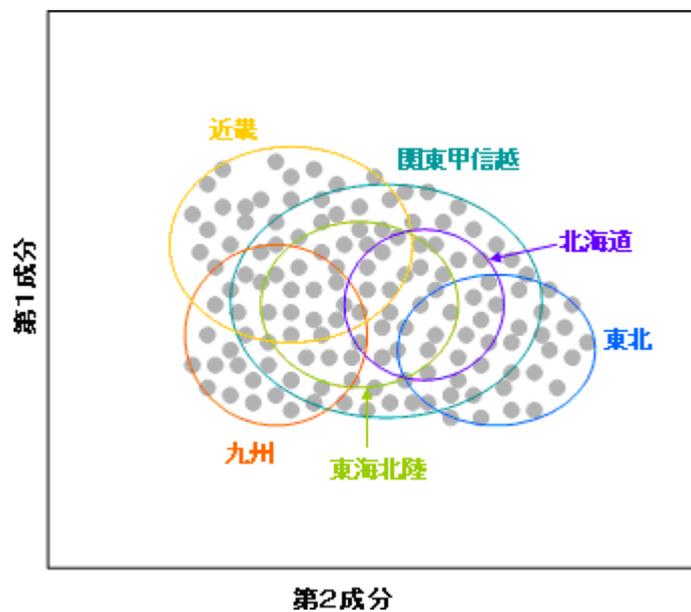


図 SNP を活用した日本人の集団構造の解析  
図 本土の遺伝的な地域差

2008年9月26日  
独立行政法人 理化学研究所

## 1人あたり約14万個所のDNA塩基多型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 -

### ◇ポイント◇

- 7,000人以上の日本人の常染色体上のDNA塩基多型情報を解析
- ほとんどの日本人は、本土クラスター、琉球クラスターの2つに大別
- 本土でも遺伝的な地域差があることが判明

独立行政法人理化学研究所（野依良治理事長）は、1人あたり約14万個所のDNA塩基多型の解析により、日本人の集団構造を大きく2つのクラスター（集団）に分けることができることを明らかにしました。これは、理研ゲノム医科学研究センター（中村祐輔センター長）統計解析・技術開発グループの鎌谷直之グループディレクター、加畑（山口）由美研究員らによる研究成果です。

私たちそれぞれが持っている遺伝子と個人の病気のかかりやすさ、薬の副作用の有無などの関係を調べる場合、病気や薬の副作用を持つ集団（患者集団：ケース）と持たない集団（対照集団：コントロール）について、遺伝子上のSNP（1塩基多型）<sup>\*1</sup>を測定し、その頻度を比較する手法が一般的に行われています。この手法はケース・コントロール解析と呼ばれ、この解析からわかるSNPの頻度情報は、個人に合った投薬や治療を行う、オーダーメイド医療を目指した研究の基盤情報として非常に有用です。しかし、対象にした集団に分集団構造があり、ケースとコントロールのサンプルの仕方に偏りがある場合、偽陽性の結果を検出する率が上昇してしまいます。今後のゲノムワイドなケース・コントロール解析のよりよいデザインのためにも、ゲノム全体を網羅するSNPを用いて、日本人の集団構造を調べることが大変重要です。

研究グループは、日本人7,001人と中国人45人のサンプルについて、1人あたり約14万個所のSNPの遺伝子型データを用いた主成分分析<sup>\*2</sup>を行いました。その結果、日本人の大部分が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。さらに、本土の中でも遺伝的な地域差があることが明確となりました。

今回明らかにした、日本人の集団構造に関する知識により、ゲノムワイドなケース・コントロール解析の研究デザインの向上が期待されます。また、今後さらに多くのアジアの近隣諸国の人のSNPと共に解析することによって、集団間の違い、民族の歴史、人の移動の程度を含めた詳細な集団構造の理解などが可能になると期待できます。

本研究は、文部科学省「個人の遺伝情報に応じた医療の実現プロジェクト（オーダーメイド医療実現化プロジェクト）」の一環として行われたもので、研究成果は、米国の科学雑誌『*The American Journal of Human Genetics*』（10月号）に掲載されるに先立ち、オンライン版（9月25日付け：日本時間9月26日）に掲載されます。

## 1. 背景

病気のかかりやすさ、薬の副作用の有無などと遺伝子の関係を調べる場合、病気や薬の副作用を持つ集団（患者集団：ケース）と持たない集団（対照集団：コントロール）について、ゲノム上のDNA塩基多型のタイプを測定し、その頻度を比較する手法「ケース・コントロール解析」が一般的に行われています。このケース・コントロール解析で用いる1塩基多型（SNP）の頻度情報は、病気や副作用と遺伝子の関係を解明し、個人に適合した投薬や治療を行うオーダーメイド医療を目指した研究の基盤情報として非常に有用です。

理研ゲノム医科学研究センターは、ゲノム全体を網羅するSNP解析の手法を確立し、これまでに、心筋梗塞、関節リウマチ、変形性関節症、糖尿病性腎症などをはじめとする数多くの疾患に関連する遺伝子を特定してきました。また近年、欧米の研究機関においても、同様の解析による疾患関連遺伝子研究が急速に進んできています。一方、SNPの頻度は人種、国、地域によって異なっており、民族の歴史の影響を受けています。このような状況を踏まえ、日本人の病気や薬の副作用に関連する遺伝子の探索をさらに深めていく上で、日本人の集団の遺伝的多様性と集団構造の理解の必要性が高まっています。

現在のケース・コントロール解析は、解析対象の集団の中では、疾患に関係のないゲノム領域のSNPの頻度は個人の疾患の有無と関係がない、という仮定を立てています。そのため、もし対象の集団に、さらに細分化した分集団構造があって、ケースとコントロールのサンプルの取り方に偏りがある場合（図1）、あるいは、コントロール集団に比べてケース集団の方がより近縁である場合は、疾患に関係のないゲノム領域でも、SNPの頻度に統計的な有意差が出やすくなり、偽陽性の結果を得る率が上昇します。

これまで理研ゲノム医科学研究センターは、日本人の集団を対象として、多くの疾患関連遺伝子を特定してきました。しかし、今後のさらなる研究を進める場合、より多数のサンプルを用いて解析を行うと、検出力が上がる反面、集団の構造化による偽陽性の結果を得る率が上がってしまうことが懸念されます。

日本人の遺伝的多様性と起源についての研究は、これまでミトコンドリアDNA<sup>\*3</sup>やY染色体<sup>\*4</sup>のDNA塩基多型を用いた研究が主流でした。それらの先行研究では、興味深いことに、日本人集団の「二重構造」説<sup>\*5</sup>を支持するものや、ミトコンドリアDNAのハプロタイプ<sup>\*6</sup>の多様性を示すものがあります。そのため、今後のゲノムワイドなケース・コントロール解析が、偽陽性の結果を得ることがないようにするためにも、ゲノム全体を網羅するSNPを用いて、日本人の集団構造を調べる必要があります。

## 2. 研究手法と成果

今回の研究では、国際ハップマッププロジェクト<sup>\*6</sup>の4つの集団（西・北欧系ユタ州住民60人、ナイジェリアのヨルバ族60人、東京在住の日本人45人、北京在住の中国人漢民族45人の合計210人）のSNPのデータに加えて、バイオバンクジャパン<sup>\*7</sup>の日本人7,003人の、常染色体<sup>\*8</sup>上にある1人あたり140,387個所のSNPを解析に用いました。この日本人7,003人は、心筋梗塞、糖尿病、関節リウマチなど35種類の疾患のいずれかの患者であり（バイオバンクジャパンでは47種類の対

象疾患があります)、病院の所在地により、7つの地域(北海道、東北、関東甲信越、東海北陸、近畿、九州、沖縄)にグループ分けされています。

研究チームは、これらの人々からの常染色体上のSNPの遺伝子型データを用いて、主成分分析を基礎にした解析手法により、個人間の遺伝子型の相関を基に個人間の近縁関係を解析しました(図2)。まず、欧米人、アフリカ人を含んだ解析により、日本人7,003人のほぼ全員が東アジア人のグループに属することを確かめました。次に、この内の7,001人を中国人45人のサンプルと共に解析した結果、7,001人は本土クラスターと琉球クラスターの2つの主なクラスターに大別されることがわかりました(図3)。つまり、前者には本土の6つの地域で採血された大部分の人が含まれ、後者には沖縄で採血された人の大部分が含まれていました。本土クラスターと琉球クラスターの遺伝的分化の程度は非常に小さく、そのためSNPの頻度の違いは大部分についてはわずかでしたが、約14万個所という数多くのSNPを用いたために、2つのクラスターを観察できたと考えられます。さらに、本土の中でも遺伝的な地域差があることが明確にわかりました(図4)。今回の結果は、従来から提唱されている日本人集団の「二重構造」説と矛盾しないものです。

本土クラスターと琉球クラスターの違いが、どのSNPでもっとも顕著であるかを調べたところ、6番染色体のHLA領域<sup>\*9</sup>に見つかりました。また、アミノ酸を変化させるSNPの頻度の違いを比較したところ、髪の毛の太さと関連のある*EDAR*遺伝子<sup>\*10</sup>のSNP、耳垢のタイプと関連のある*ABCC11*遺伝子<sup>\*11</sup>のSNPの頻度がもっとも大きい違いを示しました。

さらに研究グループは、日本人の集団構造が疾患関連遺伝子探索のケース・コントロール解析にどの程度影響するかを調べるために、2つのクラスターや地域を分集団として、個人をランダムにサンプル抽出し、シミュレーションを行いました。その一つとして、本土クラスターのケース集団(200人)とコントロール集団(200人)を基に、ケース集団における琉球クラスターからの人の割合を増やしていき、ゲノム全体のSNPの遺伝子型頻度の違いの統計量がどのように増大するかを調べました。その結果、ケース集団における琉球クラスターからの人の割合が23%になると、偽陽性の結果を得る率の増大が無視できなくなることを明らかにしました。この結果は、ケース・コントロール解析では、患者の住む地域や遺伝的背景を考慮した解析デザインが必要であることを示しています。

### 3.今後の期待

日本人における遺伝的な集団構造と地域差が、ゲノム全体を網羅するSNPで解明されたことにより、次のような発展が期待されます。まず、今回明らかになった日本人の集団構造についての知識は、今後のゲノムワイドなケース・コントロール解析の研究デザインの向上に生かすことができます。今後の解析では、検出力を上げるために、より多数のサンプル数で解析すると予想でき、今回の研究の知見を生かせば、ケース集団とコントロール集団のバランスがとれるように前もって調整することで、偽陽性をなるべく抑える研究デザインが可能になります。また、進行中あるいは検証中の研究であっても、今回の研究で示した通り、サンプルの抽出の偏りが及ぼす偽陽性への影響の程度を考慮することによって、構造化による偽陽性の問題を解決することが可能になります。さらに、本土と琉球の2つのクラスターの

間で頻度の違いが大きい SNP があることを示しているため、どの SNP やどのゲノム領域で偽陽性が起こりやすいか、ということについても役立つ情報を提供します。このように、集団構造の理解や、分集団間の違いを理解することからケース・コントロール解析の精度が上がるため、疾患関連遺伝子同定の精度の向上が期待され、より確かなオーダーメイド医療の実現化につながると考えられます。

また、日本人の集団構造が、今後、アジアの近隣諸国の人のデータと共に解析されることによって、集団間の違い、民族の歴史、人の移動の程度を含めた詳細な集団構造の理解が進むこととなります。近縁な集団間における遺伝的な関係や違いの程度を理解することで、SNP と疾患の関係を解明する研究が、アジアの近隣諸国でも加速していくと注目されます。

今回の研究により、アミノ酸を変化させている SNP の中で、本土クラスターと琉球クラスター間のもっとも大きな違いが、髪の毛の太さや耳垢のタイプのような、表現型の違いにかかわるものであったことは、非常に興味深い結果です。今後、近縁な分集団の間で頻度の違いの大きな SNP を調べることによって、人類の歴史において過去に働いた自然選択が、表現型に違いを生じるような SNP の頻度の変化に影響を与えてきた例をさらに探し出すことができる可能性があるかと期待できます。

(問い合わせ先)

独立行政法人理化学研究所

ゲノム医科学研究センター統計解析・技術開発グループ

グループディレクター 鎌谷直之 (かまたになおゆき)

Tel : 03-5449-5708 / Fax : 03-5449-5564

横浜研究推進部 企画課

Tel : 045-503-9117 / Fax : 045-503-9113

(報道担当)

独立行政法人理化学研究所 広報室 報道担当

Tel : 048-467-9272 / Fax : 048-462-4715

Mail : koho@riken.jp

## <補足説明>

### ※1 SNP(スニップと発音されることが多い)

DNA の 1 塩基多型 (Single Nucleotide Polymorphism) で、過去に起きた DNA の点突然変異による。ヒト集団では染色体をランダムに 2 本選んで比較すると、1,200~1,500 塩基に 1 つの割合で、SNP があると推定されている。ヒトゲノム上では、約 12,000,000 の SNP が報告されている。

## ※2 主成分分析

複数の変数間の共分散（相関）を少数の合成変数で説明する手法。共分散行列の固有値問題の解として得ることができる。この研究では、各個体点を座標のばらつきがなるべく大きくなるように直交軸を取り直して眺めている。

## ※3 ミトコンドリア DNA

細胞内小器官のミトコンドリアが独自に持っている環状の DNA。過去に細胞内で共生したほかの細胞が由来である、と考えられている。ミトコンドリア DNA は、受精卵の細胞質により母から子に伝えられる。DNA 複製のエラー修復機構を欠き、分子進化速度が核の DNA よりも速い。ヒトのミトコンドリア DNA は約 16,600 塩基対と短いが、塩基多型の割合が核の DNA よりも高い。ヒトの遺伝的多様性と移動の歴史を推測する研究によく使われている。

## ※4 Y 染色体

性染色体の 1 つで、ヒトでは Y 染色体があると男性になる。通常の男性の性染色体の核型は X 染色体と Y 染色体が 1 本ずつで、Y 染色体は父親から息子に遺伝する。ヒトの遺伝的多様性と移動の歴史を推測する研究によく使われている。

## ※5 日本人集団の「二重構造」説

埴原和郎（はにはらかずろう）氏の「二重構造モデル」。日本人の祖先集団は、縄文人、そして北東アジアから渡来した弥生人に由来しており、この 2 集団は日本列島内で徐々に混血したが、アイヌ人と南西諸島の人においては、北東アジアからの渡来人の影響は少なかった、とする説。

## ※6 国際ハップマッププロジェクト/ハプロタイプ

ハップマップとは、ハプロタイプ地図を略したもの。ハプロタイプは 1 本のゲノム（染色体）上の DNA 塩基多型の組み合わせである。このプロジェクトは、疾患関連遺伝子同定のためのゲノムワイドなケース・コントロール解析を加速させることを第一の目的に発足した。日本人、中国人、ヨーロッパ系アメリカ人、アフリカ人からの 4 集団 270 人について、ゲノム全体を網羅する SNP の遺伝子型が決定され、データベースで公開されている。

## ※7 バイオバンクジャパン

文部科学省「個人の遺伝情報に応じた医療の実現プロジェクト（オーダーメイド医療実現化プロジェクト）」の基盤となる DNA サンプルおよび血清サンプルを収集し臨床情報とともに保管している世界でも有数の資源バンクの名称。バイオバンクへ約 30 万人の DNA および血清試料を集め、それらを利用して SNP と病気との関係、薬剤の効果などの関係を明らかにする研究を行っている。心筋梗塞、糖尿病、関節リウマチなど 47 種類の対象疾患がある。情報は個人情報管理に配慮し幾重にも厳重に管理されており、東京大学医科学研究所内に設置されている。

## ※8 常染色体

ヒトでは1~22番染色体のこと。細胞内の染色体の全セットのうち、性染色体（X染色体とY染色体）を除いたもの。

## ※9 HLA(human leukocyte antigen)領域

ヒトの6番染色体短腕（6p21）上に存在し、その領域にある多重遺伝子族（遺伝子重複によって生じた遺伝子の中で相同性が高いもの）は、自己と非自己の認識、免疫応答の誘導に関与する白血球抗原をコードしている。MHC（major histocompatibility complex;主要組織適合性複合体）領域とも呼ばれる。

## ※10 EDAR 遺伝子(ectodysplasin A receptor)

シグナル伝達に関与する遺伝子の1種で、2番染色体上に存在する。370番目のアミノ酸変異（アラニンまたはバリン）の頻度は、東アジア人と他の人種で大きく異なっており、髪の毛の太さに関与することがわかっている。東アジア人は他の人種と比べて髪の毛が太く、アラニン型の頻度が高い。

## ※11 ABCC11 遺伝子(ATP-binding cassette, subfamily C, member 11)

ATP結合機能を持ち、輸送にかかわるタンパク質のファミリーをコードする多重遺伝子族のメンバーの1つで、16番染色体上に存在する。薬剤排出の機能を持つ。180番目のアミノ酸変異（グリシンまたはアルギニン）は、耳垢の乾型、湿型を決めることがわかっている。そのSNPの頻度は東アジアと他の地域で大きく異なっており、東アジア人では乾型のアルギニンのタイプの頻度が高い。

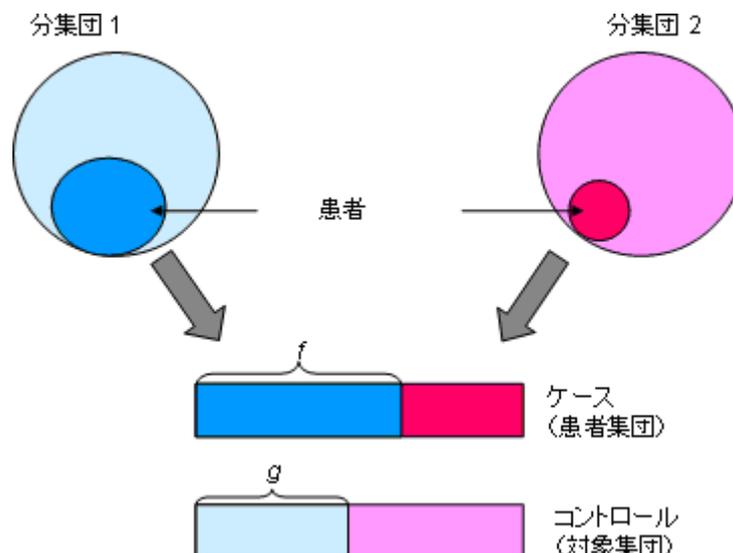


図1 ケース・コントロール解析で偽陽性率が上がる条件

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  $f$  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。

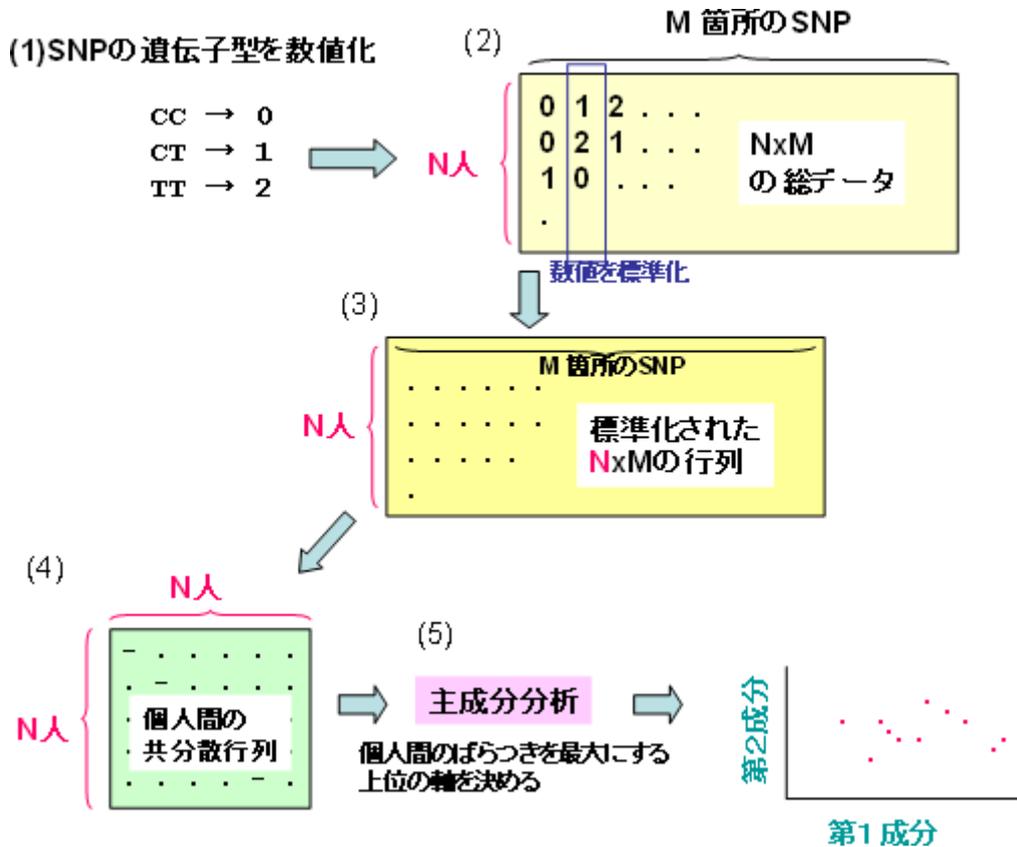


図2 たくさんの個所のSNPの遺伝子型を使った個人間の近縁関係の解析

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、遺伝子型の相関の度合いのデータを得る (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。

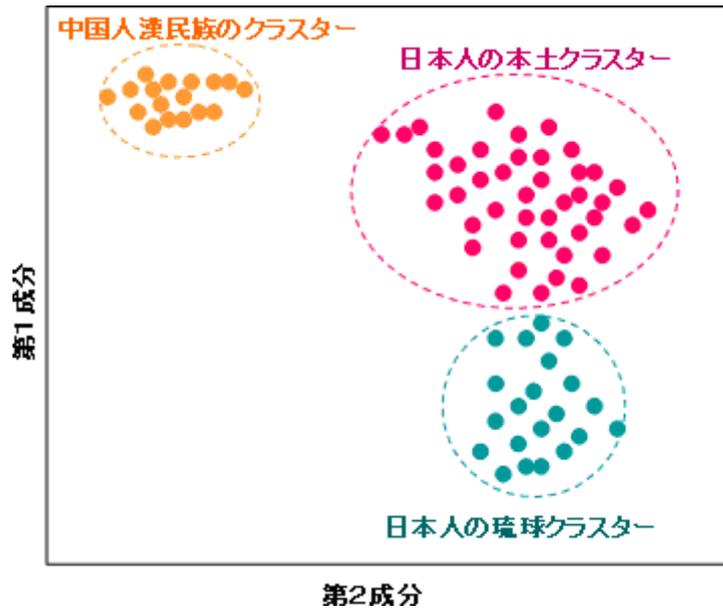


図3 ほとんどの日本人は、本土クラスターと琉球クラスターに大別される

日本人と中国人漢民族で個人間の近縁関係を解析したところ、主な3つのクラスターに分かれた。1つは中国人のクラスターである（橙）。日本人は、本土クラスター（赤）と琉球クラスター（緑）の2つに大別される。

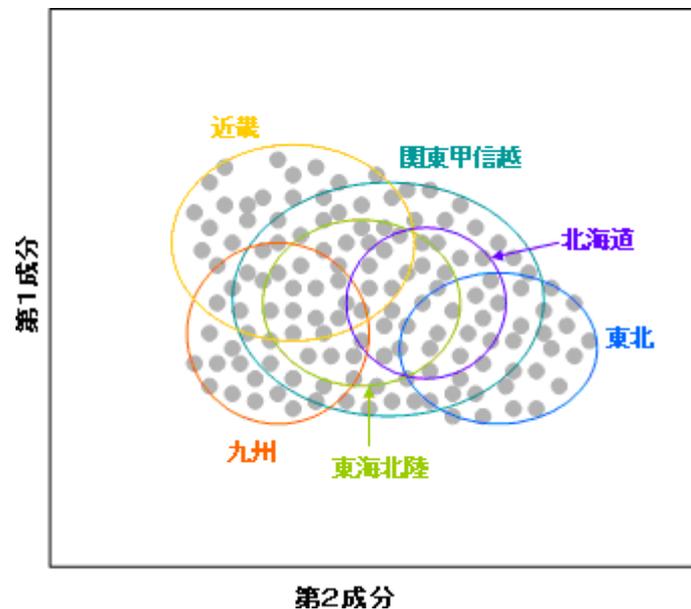


図4 日本の各地域の遺伝的な分化と地理的な傾向

本土クラスターの中の人分布は、地域によって偏りがあることがわかった。近畿の人は、上の図では本土クラスターの左上に偏りがあり、中国人のクラスターに比較的近い。対照的に、東北の人は、本土クラスターの右下に偏っていて、中国人のクラスターから見て離れている。