

遺伝的多型の関連解析における構造化と異質性について

中村好宏

防医大誌 (2022) 47 (1) : 11-18

要旨：塩基配列決定の技術の発展に伴い、2000年頃から遺伝的多型と疾患の関連解析の報告が増加した。疾患に関連する遺伝子（疾患感受性遺伝子）の報告が増加するにつれて、1つの現象が問題となった。それは追試の失敗である。特定の集団で疾患に関連があると報告された遺伝的多型について、他の集団で追試を行うと関連が確認されなかったという結果の齟齬が多くみられるようになった。その原因の1つは最初の報告が偽の発見であった、すなわち関連が無い遺伝子が関連ありと判断されてしまった場合である。偽の発見は統計学的仮説検定を用いる際は確率的に起こり得るが、データの偏り、交絡因子の影響などによっても引き起こされる。遺伝的関連解析については、交絡因子となる変量は少ないと予想されるが、構造化が交絡因子になることが知られている。そのため、構造化の影響を排除する方法が数多く提案されている。

また、追試で関連が確認されないもう1つの原因として、最初の研究で用いた集団と追試で用いた集団の間に多型の効果の異質性があることが挙げられる。この場合、追試で関連が確認されないことがむしろ正しい結果であることも有り得る。

本稿では、遺伝的関連解析における構造化と多型の効果の異質性について概観し、それらを解決するための手法を紹介する。

索引用語： 遺伝的多型 / 関連解析 / 構造化 / メタアナリシス / 効果の異質性

緒 言

疾患の発症に関連する遺伝子の研究はこの20年で急速に数を増やし、多くの疾患について、その発症に関連する遺伝子（疾患感受性遺伝子）が報告されている。これらの業績は、1865年、メンデルにより発見された遺伝継承法則（メンデルの法則）によって遺伝についての定量的な研究が可能となったこと、そして1953年、ワトソンとクリック¹⁾によりDNAの二重らせん構造が発見され、遺伝情報はDNAを介して継承されることが判明したことによる。これらの成果より、遺伝子の多様性に関する研究が発展し、DNA配列の研究が急速に進む中、2003年、ヒトゲノムの全配列が解明されることとなった。この前後、PCR法の発展に伴い、DNAの

塩基配列を決定するタイピングの技術が急速に発展した。タイピングにより塩基配列が決定されると、次はDNA配列に関する様々な研究が行われるようになり、DNA配列と表現型（疾患・薬剤の副作用など）の関連を調べる遺伝的関連解析が行われるようになった。そこで必要とされたものが統計学であり、遺伝子の関連解析に特化した統計学である遺伝統計学が発展することとなった。関連解析において、遺伝統計学の手法をもって解決すべき問題は今も多い。本稿では、関連解析における2つの問題、構造化、多型の効果の異質性についての解説を行う。

遺伝的関連解析

人間の全染色体は約30億の塩基対で構成され

向検定を適用する場合もあり、表現型が連続変数の場合には回帰分析を用いることもある。また、GWASでは欠測データを埋めるインプテーション⁴⁾という方法が用いられ、インプテーション後のデータに対し、ロジスティック回帰分析で関連の有無を確認する方法も存在する。

関連解析の問題点

関連解析の報告が増加するにつれて、1つの現象が問題となった。それは追試の失敗である。特定の疾患との関連が報告されたSNPについて、他の集団でも追試を行ったところ、関連が確認されないという現象が多くみられた。例えば、Tokuhiraら⁵⁾は日本人集団において関節リウマチの発症と *SLC22A4* のSNPに関連があることを報告した。しかし、Kuwaharaら⁶⁾による別の日本人集団を用いた関連解析では関節リウマチの発症とこのSNPの関連は確認されなかった。また、Suzuki⁷⁾らは関節リウマチと *PADI4* のSNPに関連があることを報告し、Iwamotoらが行った他の日本人集団での追試でも関連が確認されたが、欧米人における追試では関連が確認されなかった⁸⁾。

結果が食い違う原因の1つは、最初の報告が第1種過誤、すなわち偽陽性の場合である。偽

陽性の原因はデータ収集の不備、多重比較の問題など複数あるが、構造化の影響も考えられる。遺伝的多型の関連解析は、一般的な医学データの解析に比べて交絡因子の影響は少ないと予想されるが、構造化が交絡因子となり偽陽性率を増加させることが知られている。もう1つの考えられる原因は人種による多型の効果の相違である。追試で用いられた集団と最初の報告で用いられた集団の遺伝的背景が異なる場合、それぞれの集団において多型の効果が異なることが有り得る。例えば最初の報告で用いられた集団では真の関連があり、追試で用いられた集団では本当に関連が無い場合、このような齟齬が起きることはむしろ正しい結果である。本稿では、上記の構造化、人種間の効果の異質性について概観し、その対処法として提案された方法を紹介する。

構造化の問題

遺伝的多型と疾患の関連解析においては、構造化についての検討を行う必要がある。構造化とは、関連解析の対象となる母集団に複数の異なる遺伝的背景をもつ副母集団が含まれることを指す。研究対象の母集団に構造化があると、偽陽性が生じる確率が増加することが知られている。

遺伝型	A/A	A/T	T/T	計
患者	n_{10}	n_{11}	n_{12}	$n_{1\cdot}$
健常者	n_{20}	n_{21}	n_{22}	$n_{2\cdot}$
計	$n_{\cdot 0}$	$n_{\cdot 1}$	$n_{\cdot 2}$	n



アレル	A	T	計
患者	$2n_{10} + n_{11}(N_{11})$	$2n_{12} + n_{11}(N_{12})$	$2n_{1\cdot}(N_{1+})$
健常者	$2n_{20} + n_{21}(N_{21})$	$2n_{22} + n_{21}(N_{22})$	$2n_{2\cdot}(N_{2+})$
計	$2n_{\cdot 0} + n_{\cdot 1}(N_{+1})$	$2n_{\cdot 2} + n_{\cdot 1}(N_{+2})$	$2n(N)$

遺伝子型の分割表
(2 × 3)をアレルの
分割表(2 × 2)に
変換

検定統計量

$$X_0^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{\left(N_{ij} - \frac{N_{i+}N_{+j}}{N} \right)^2}{\frac{N_{i+}N_{+j}}{N}}$$

有意水準を α としたときの棄却域

$$X_0^2 > \chi_{\alpha}^2(1)$$

図2. 分割表の独立性の検定を用いた関連解析

患者、健常者それぞれにおけるA/A、A/T、T/Tの遺伝型をもつ人数が上の表で表される場合、患者と健常者におけるアレルAとTの頻度は下の表になる。この表に分割表の独立性の検定を用いることで、患者と健常者のアレル頻度の違いを検出する。

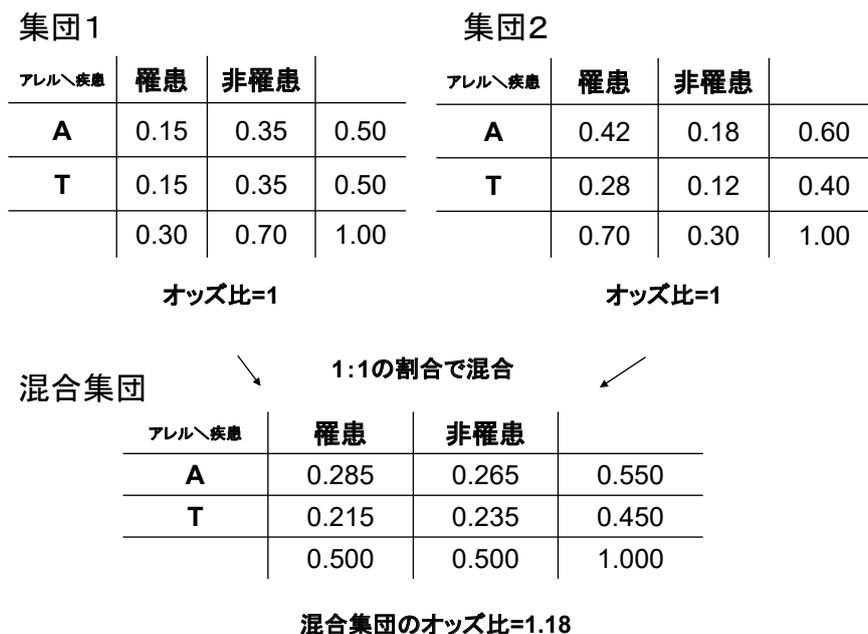


図3. 構造化の例

集団1と集団2では、どちらもオッズ比が1であり、このSNPは疾患の発症に無関係である。しかし、集団1と2ではアレル頻度と疾患の頻度が異なっており、この2集団を1:1の割合で混合させた集団では、患者と健常者の間にアレル頻度の偏りが発生し、分割表の独立性の検定を行うと、疾患とSNPに関連があると結論付けてしまうことになる。

医学データの解析において、交絡因子の特定、除去は重要な課題である。遺伝的多型データにおける構造化とは、各副母集団において、遺伝的多型のアレル頻度、疾患の頻度がそれぞれ異なる場合を指す。関連解析において、性別、喫煙歴、年齢、BMIなどの項目は遺伝的多型の頻度には無関係と考えられるので、交絡因子ではなく独立した共変量と考えられる。しかし、構造化はアレル頻度と疾患頻度の両方に関係するため交絡因子となる。図3は、疾患に無関係なSNPについて、2つの副母集団におけるアレルの相対頻度を記したものである。2つの副母集団は、どちらもオッズ比が1であるが、アレル頻度と疾患の頻度が異なる。この2集団を1:1の割合で合わせた表ではオッズ比が1ではなく、偽の関連が見られる。この問題は、データ数の増加では解決しない。解決するには、異なる遺伝的背景をもつ副母集団を特定が必要である。(図3)

構造化を検出するためには、複数の座位における遺伝型データの収集が必要である。多座位の遺伝型データを用いた方法としては、古くはWrightにより F_{st} が提案されている⁹⁾。これは、副母集団が分かっている場合、その副母集団間

の構造化の指標となる値である。 G 個の集団で、あるSNPのアレル頻度を計測したデータがあるとき、 F_{st} は

$$F_{st} = \frac{\sum_{k=1}^G (\hat{p}_k - \hat{p})^2 / (G - 1)}{\hat{p}(1 - \hat{p})}$$

のように計算される。ここで、 $\hat{p}_k = \sum_{i=1}^{n_k} X_i^{(k)} / n_k$ 、 $\hat{p} = \sum_{k=1}^G \hat{p}_k / G$ 、 n_k は k 番目の副母集団の個体数の2倍、 $X_i^{(k)}$ は k 番目の副母集団の i 番目のアレルを表す2値変数である。複数のSNPにおいて F_{st} を計算し、その平均値が大きければ、 G 個の副母集団には遺伝的背景に相違があることが示唆される。

副母集団が判明している場合は、上記の方法で構造化の確認が可能だが、実際には研究対象としている集団が複数の遺伝的背景が異なる副母集団の混合母集団であることが分からない場合が多く、その対応を考える必要がある。Nakamuraら¹⁰⁾は、遺伝的多型データを元に、階層型クラスタリングや分割最適化型クラスタ分析を用い、遺伝的背景が異なる副母集団を推定するアルゴリズムを提案した。また、Pritchardら¹¹⁾は、遺伝形式をモデル化したSTRUCTUREを提案した。STRUCTUREは各個体の各副母集団への帰属度を求める方法で

ある。構造化があるデータの解析法は、副母集団毎に解析を行い、それをメタアナリシスで統合することが一般的である。Nakamura¹⁰⁾は、クラスタリングにより推定した副母集団毎の解析結果をMantel-Haenszel検定で統合するアルゴリズムを提案している。

また、副母集団の推定とは異なるアプローチで構造化による偽陽性の問題を解決する方法が提案されている。構造化が偽陽性の確率を上昇させる理由は、分割表の独立性の検定における検定統計量の値を増加させるためである。検定統計量を増加させる量を推定し、その増加分を補正する方法が、Devlinらにより提案されたGenomic controlである¹²⁾。また、Priceらにより提案されたEIGENSTRATは、主成分分析を行い、その主成分得点を共変量として関連解析を行い、構造化の影響を補正する方法である¹³⁾。

このように構造化の影響を排除する様々な方法が提案されたが、現在はEIGENSTRATのように主成分分析や多次元尺度法により個体のデータを低次元空間へ射影し、その散布図から副母集団の存在をクラスター分析で読み取る方法が一般的になっている。これは副母集団の推定のみならず、遺伝的背景が異なる外れ値を検出する方法として有効である。精度が高いデータの収集が可能になった現在は、何かしらのアルゴリズムで副母集団を推定するのではなく、人種などあらかじめわかっている情報を元に、均質な集団で解析を行い、それを統合することが進められており、副母集団の推定を行うアルゴリズムは均質な集団における外れ値の検出の役割を負うことが多い。Yamaguchiらは、日本人の大規模データを収集し、日本人の遺伝的背景を調べ、日本人集団には構造化は見られないことを示した¹⁴⁾。この研究は、日本人集団における関連解析では構造化の影響が少ないことを示唆している。

多型の効果の異質性

関連解析において、多型の効果には人種間の異質性があることが知られている。Nakamuraら¹⁵⁾はASPNの変形性関節症への効果に異質性があることを報告している。ASPNにおける

D14アレルの膝変形性関節症への効果には異質性が見られ、東アジアの集団では強い関連が見られるものの、ヨーロッパの集団では関連がみられなかった。人種間の異質性の原因は環境要因と遺伝的要因に分けられる。環境要因とは食生活、文化、気候などの外的要因であり、遺伝的要因とは遺伝子により決まる個人差を指す。遺伝的要因により異質性が発生するメカニズムを単純な例で説明すると次の通りである。AというSNPとBというSNPに交互作用があり、どちらのSNPにも変異をもつと、ある疾患の発症リスクが高くなるとする。研究者がBの存在を知らずにAのみで関連解析を行った場合、Bの頻度が低い人種ではAの効果は低く、Bの頻度が高い人種ではAの効果が高くなる。このように遺伝的背景が異なる集団においては多型の効果が異なる可能性があるが、そのようなデータを解析する場合、異質な集団それぞれで効果の推定を行い、メタアナリシスにより結果の統合可能性を検証する。もし統合可能であれば統合した結果で多型の効果の検証を行うことが推奨される。

関連解析のメタアナリシス

1つの遺伝的多型について複数の集団で関連解析を行うと、結果にはばらつきが見られる。メタアナリシスは、結果がばらついた複数の研究の統合可能性を検討し、可能な場合は結果を統合して研究の精度を高める手法の総称である。

今、あるSNPと疾患の関連について、 K 個の施設で別々にデータ収集が行われているものとする。エビデンスの質を高めるため、この K 個の施設のデータを統合することを考えるが、その際、データを単純に合計するだけでは、構造化の問題と同様に多型の効果の推定に悪影響を及ぼし、誤った結果を導き出す可能性があり、それを避けるためにメタアナリシスを行う必要がある。メタアナリシスには多くの手法があるが、多型の効果をオッズ比で測る場合の最も基本的な手法は漸近分散法であろう。漸近分散法は以下の通りである¹⁶⁾。

施設 i における 2×2 分割表の数値を A_i, B_i, C_i, D_i ($i = 1, 2, \dots, K$)、真のオッズ比を θ とする。すると施設 i におけるオッズ比の推定式 $\hat{\theta}_i$

は $(A_i D_i)/(B_i C_i)$ であり、 $\hat{\theta}_i$ の対数 $\log \hat{\theta}_i$ は漸近的に平均 $\log \hat{\theta}$ 、分散 $\frac{1}{A_i} + \frac{1}{B_i} + \frac{1}{C_i} + \frac{1}{D_i}$ の正規分布にしたがう。この性質を利用し、全施設のデータを統合した $\log \hat{\theta}$ の推定式 $\log \hat{\theta}_{all}$ は $\log \hat{\theta}_i$ の重み付き平均として

$$\log \hat{\theta}_{all} = \frac{\sum_{i=1}^K w_i \log \hat{\theta}_i}{\sum_{i=1}^K w_i}$$

とする。ただし、 $w_i = \left(\frac{1}{A_i} + \frac{1}{B_i} + \frac{1}{C_i} + \frac{1}{D_i}\right)^{-1}$ である。

すると $\log \hat{\theta}_{all}$ は $\theta=1$ のとき、平均0、分散 $1/\sum_{i=1}^K w_i$ の正規分布にしたがうので、

$$Q_1 = \frac{(\sum_{i=1}^K w_i \log \hat{\theta}_i)^2}{\sum_{i=1}^K w_i}$$

は自由度1のカイ二乗分布にしたがう。これにより、オッズ比が1であるか（すなわちSNPと疾患に関連があるか）の仮説検定が可能である。

オッズ比のメタアナリシスとして、漸近分散法とDerSimonian-Laird法が有名である。漸近分散法は全ての研究において真のオッズ比が共

通であることを前提とした母数モデル（Fixed effect model）であり、DerSimonian-Laird法は各研究における真のオッズ比に多少のばらつきがあることを許容する変量モデル（Random effect model）である。一般的にDerSimonian-Laird検定の方が保守的な結果を導出する傾向を有するため、こちらを採用することが多い。

遺伝的効果の異質性の検討

複数の研究において真の効果が共通で、結果のばらつきは偶然の範囲で起こり得る程度である場合、メタアナリシスで結果を統合し、効果の推定精度を高めることが出来る。変量モデルは、真の効果が研究によりばらつくことを許容しているが、効果が全く異なる研究を統合すると、誤った結論を導き出す可能性がある。そこで、効果の異質性を評価し、それにより結果の統合可能性を検討する必要がある。異質性の評価は仮説検定を用いて行われることが多かった

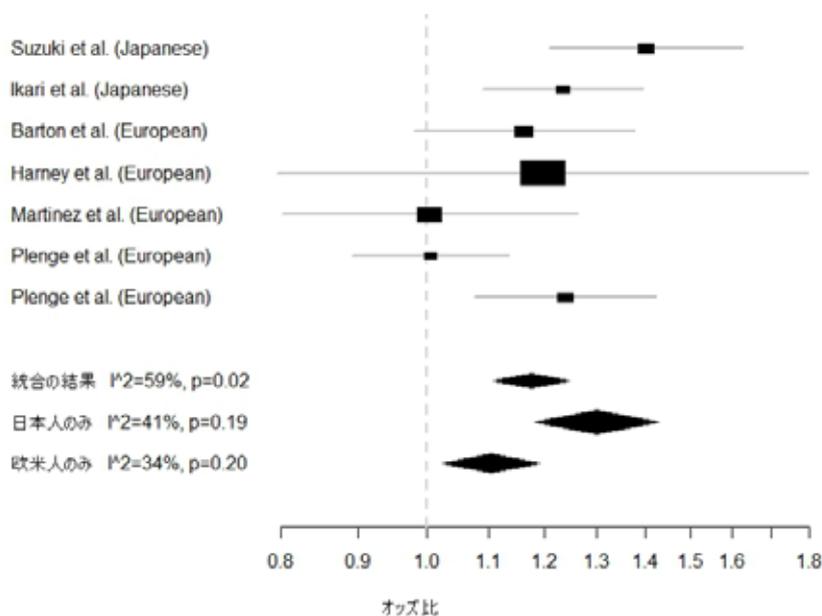


図4. 関節リウマチとPADI4のSNP94のメタアナリシス

図はそれぞれの研究および統合した結果のオッズ比とその95%信頼区間を表す。Iwamotoら⁹⁾による関節リウマチとPADI4のSNP94に関する遺伝型データについてメタアナリシスを行った。日本人集団に関する2つの研究と欧米人集団に関する5つの研究を漸近分散法で統合した結果、オッズ比は1.17（95%信頼区間が1.11-1.25）となり、このSNPと関節リウマチには関連が見られる。ただし、7つの研究間の I^2 は59%となり、強い異質性が見られるため、日本人集団と欧米人集団を分けて、それぞれについて I^2 を計算すると、41%、34%と強い異質性が確認されなかった。したがって、集団ごとに解析を行うと、日本人集団におけるオッズ比は1.30（95%信頼区間が1.18-1.43）、欧米人集団におけるオッズ比が1.10（95%信頼区間が1.02-1.19）となった。このことからPADI4のSNP94はどちらの集団でも有意に関節リウマチに関連があるものの、どちらもオッズ比が低く、特に欧米人における効果はかなり弱いものであることが示唆される。この解析は統計解析ソフトRのmetaパッケージで行った。

が、現在は I^2 をもって行われることが多い。どちらも以下の式、Cochraneの統計量

$$Q = \sum_{i=1}^K w_i (\log \hat{\theta}_i - \log \hat{\theta}_{all})^2$$

を利用する。 Q は各施設の真のオッズ比が共通であるとき、自由度 $K-1$ のカイ二乗分布にしたがい、この性質により仮説検定が可能である。ただし、検定が棄却されれば施設間に異質性があることが示されるが、棄却されなかったからといって、等質であるとは主張できない。したがって現在は仮説検定より、異質性の指標である次式、

$$I^2 = \frac{Q - (K - 1)}{Q} \times 100 (\%)$$

を用いる方が多い。 I^2 が0~25%のときは異質性無し、25~50%のときは異質性中程度、50~75%のときは異質性が強い、75~100%のときは異質性が非常に強いと評価する¹⁷⁾。 I^2 が50%より大きいときは、メタアナリシスは行わず各施設での解析に留める、または等質な施設のみでメタアナリシスを行うべきである。

Iwamotoら⁸⁾は関節リウマチと*PADI4*のSNP94 (rs2240340)に関するメタアナリシスを行っている。ここでは日本人集団に関する2つの研究と欧米人集団に関する5つの研究を統合した結果、このSNPと関節リウマチには関連があることを報告している。この報告にあるデータを用いてメタアナリシスを行うと図4のようになった。7つの研究間の I^2 は54%となり、強い異質性が見られる。そして、日本人集団と欧米人集団を分けて、それぞれについて I^2 を計算すると、41%、34%と強い異質性は確認されなかった。したがって、それぞれの集団ごとに解析を行うと、日本人集団におけるオッズ比は1.30 (95%信頼区間が1.18-1.43)、欧米人集団におけるオッズ比が1.10 (95%信頼区間が1.02-1.19)となった。このことから*PADI4*のSNP94はどちらの集団でも有意に関節リウマチに関連があるが、欧米人における遺伝的効果は日本人のものよりも弱いことが示唆される。なお、その後、別の日本人集団における追試¹⁸⁾とマレーシアにおける追試¹⁹⁾でもこの関連が確認されたが、イギリスにおける追試では関連

がないことが示された²⁰⁾。このことからアジア人における関連は確かなものであると考えられるが、欧米人ではほとんど関連が見られないといえるであろう。(図4)

また、Nakamuraら²¹⁾はEMアルゴリズムによる混合正規分布の推定法を用いて、各サブグループにおける多型の効果の推定と、異質性の有無を情報量基準による最適なサブグループ数の推定する方法を提案している。

結 語

これまでの研究で、多くの疾患が多因子疾患であり、疾患感受性遺伝子が多数存在することが明らかになっている。それらの中には効果が弱いため、未だに発見されていないものも多いと考えられる。したがって関連解析による疾患の原因遺伝子探索は、今後も進められていくであろう。そして発見された疾患感受性遺伝子間の交互作用の研究が進み、効果の異質性の原因の解明も進むと思われる。今回解説した構造化に限らず、偽陽性を避けるための解析手法もまだ改良の余地があると思われる。今後もデータからバイアスを取り除く、精度が高い解析法の発展が望まれる。

利益相反

この研究に関する利益相反はない。

文 献

- 1) Watson JD, Crick FH.: Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*. 171: 737-738, 1953.
- 2) 鎌谷直之: 遺伝統計学入門岩波書店. 東京, 2007.
- 3) 杉山高一, 藤越康祝: R・Pythonによる統計データ科学. 勉誠出版, 東京, 2020.
- 4) Marchini J, Howie B.: Genotype imputation for genome-wide association studies. *Nat Rev Genet*. 11: 499-511, 2010.
- 5) Tokuhiro S, Yamada R, Chang X, et al.: An intronic SNP in a *RUNX1* binding site of *SLC22A4*, encoding an organic cation transporter, is associated with rheumatoid arthritis. *Nat Genet*. 35: 341-348, 2003.
- 6) Kuwahara M, Ikari K, Momohara S, et al.: Failure to confirm association between *SLC22A4* polymorphism and rheumatoid arthritis in a Japanese population. *Arthritis Rheum*. 52: 2947-2948, 2005.
- 7) Suzuki A, Yamada R, Chang X, et al.: Functional haplotypes of *PADI4*, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. *Nat Genet*. 34: 395-402, 2003.

- 8) Iwamoto T, Ikari K, Nakamura T, et al.: Association between *PADI4* and rheumatoid arthritis: a meta-analysis. *Rheumatology*. **45**: 804-807, 2006.
- 9) Wright, S: The Genetical Structure of Populations. *Ann Eugen.* **15**: 323-354, 1951.
- 10) Nakamura T, Shoji A, Fujisawa H, et al.: Cluster analysis and association study of structured multilocus genotype data. *J Hum Genet.* **50**: 53-61, 2005.
- 11) Pritchard JK, Stephens M, Donnelly P.: Inference of Population Structure Using Multilocus Genotype Data. *Genetics*. **155**: 945-959, 2000.
- 12) Devlin B, Roeder K, Wasserman L.: Genomic control, a new approach to genetic-based association studies. *Theor Popul Biol.* **60**: 155-166, 2001.
- 13) Price AL, Patterson NJ, Plenge RM, et al.: Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* **38**: 904-909, 2006.
- 14) Yamaguchi-Kabata Y, Nakazono K, Takahashi A, et al.: Japanese population structure, based on SNP genotypes from 7003 individuals compared to other ethnic groups: effects on population-based association studies. *Am J Hum Genet.* **83**: 445-456, 2008.
- 15) Nakamura T, Shi D, Tzetzis M, et al.: Meta-analysis of association between the *ASPN* D-repeat and osteoarthritis. *Hum Mol Genet.* **16**: 1676-1681, 2007.
- 16) 丹後俊郎：新版メタ・アナリシス入門。朝倉書店，東京，2016。
- 17) 野口善令，福原俊一：はじめてのメタアナリシス。NPO 法人健康医療評価研究機構，東京，2009。
- 18) Takata Y, Inoue H, Sato A, et al.: Replication of reported genetic associations of *PADI4*, *FCRL3*, *SLC22A4* and *RUNX1* genes with rheumatoid arthritis: results of an independent Japanese population and evidence from meta-analysis of East Asian studies. *J Hum Genet.* **53**: 163-173, 2008.
- 19) Too CL, Murad S, Dhaliwal JS, et al.: Polymorphisms in peptidylarginine deiminase associate with rheumatoid arthritis in diverse Asian populations: evidence from MyEIRA study and meta-analysis. *Arthritis Res Ther.* **14**: R250, 2012
- 20) Burr ML, Naseem H, Hinks A, et al.: *PADI4* genotype is not associated with rheumatoid arthritis in a large UK Caucasian population. *Ann Rheum Dis.* **69**: 666-670, 2010.
- 21) 中村好宏，高橋篤，鎌谷直之：遺伝的効果の異質性に基づくクラスター分析。統計関連学会連合大会講演報告集，P97，2006。

Population stratification and heterogeneity in genetic association studies

Takahiro NAKAMURA

J. Natl. Def. Med. Coll. (2022) **47** (1) : 11 – 18

Abstract: The number of genetic association studies has increased with the rapid development of DNA sequencing technology. Many disease-susceptibility genes have been detected by genetic association studies that have identified positive associations between polymorphisms and diseases; however, replication studies often fail to confirm the associations.

One cause of such discrepancies may be that the initial positive report was a false discovery. Although false discoveries can happen stochastically, they can also be caused by confounding factors, such as population stratification in genetic association analyses. Therefore, several methods for excluding the influence of stratification have been proposed.

Another possible cause of the discrepancies is the heterogeneity of the effects of polymorphisms. The effects may vary according to the genetic background of each population. For example, two conflicting reports may both be correct if the examined polymorphism is associated with a disease in one population, but not in the other population.

In this review, population stratification and the heterogeneity of the effects of polymorphisms are overviewed, and methods for dealing with the problems caused by the stratification and heterogeneity are introduced.

Key words: Genetic polymorphism / Association study /
Stratification of population / Meta-analysis / Heterogeneity of effects