

## 特集

# 生活習慣病の遺伝因子群と ポリジェニックリスクスコア

尾崎 浩一\*

## 内容紹介

虚血性心疾患や糖尿病、認知症などの生活習慣病は、その名のとおり生活習慣が疾患の発症に左右するが、実際は遺伝因子が発症に大きく関与していることが知られており、これを遺伝率として表す(たとえば虚血性心疾患、糖尿病で40~50%程度、アルツハイマー病で70%程度)。遺伝因子の同定を通じた疾患の全容解明から、エビデンスに基づいた精密医療が可能になると考えられる。疾患に関連する遺伝因子を正確に効率よく同定するには、多数の解析試料および可能な限り正確な臨床情報を収集する必要がある。

欧米諸国では、近年UKバイオバンクをはじめとして、10~100万人規模のバンクが稼働しているが、本邦においては2000年代初頭よりすでにバイオバンクジャパン(BBJ)が始動し、その後、ナショナルセンターバイオバンクネットワーク(NCBN)、東北メガバンク(ToMMo)が加わり、本邦の三大バイオバンクとして、生体試料、情報、ゲノムデータ等の収集が進められている。

近年、このようなバイオバンクの基盤整備が進むにつれて、数万人~数十万人を対象として、全

ゲノムにわたる一塩基多型などの多様性(バリエーション)を用いた大規模ゲノムワイド関連解析や、そのメタ解析などの体系的解析が進められるようになり、真の疾患感受性遺伝因子群の同定が加速してきた。遺伝因子データは生涯不変であり、これにより構築したポリジェニックリスクスコアなどによる疾患発症の早期予知、予防が注目を浴びつつある。

## はじめに

2000年に人体の設計図とも言われる全ゲノムのドラフト配列が解読されて以来、国際ハップマッププロジェクト、1000人ゲノムプロジェクトなどによるゲノムバリエーションの整備や、さらにゲノム解析技術、機器、大規模データを処理する情報技術が著しく躍進し、当時では考えられないスピードで疾患や表現型、形質に関する遺伝因子群が明らかとなってきた。2015年にはプレジジョンメディシンという言葉が米国オバマ大統領(当時)の一般教書演説において言及され、疾患や形質に関連したゲノム解析を基盤としたオミクス解析が、医療研究開発の中で大規模に進められるようになってきた。

## I. 生活習慣病と遺伝因子

図1に示したように、生活習慣病は多数の遺伝因子の異常が集合することにより発症することから、ポリジェニック疾患と呼ぶことができる。ただ、その一つひとつの影響は非常に小さく(オッ

—Key words—  
生活習慣病, アルツハイマー病, ゲノム解析,  
ポリジェニックリスクスコア

\* Kouichi Ozaki: 国立長寿医療研究センター/メディカルゲノムセンター/理化学研究所生命医学研究センター

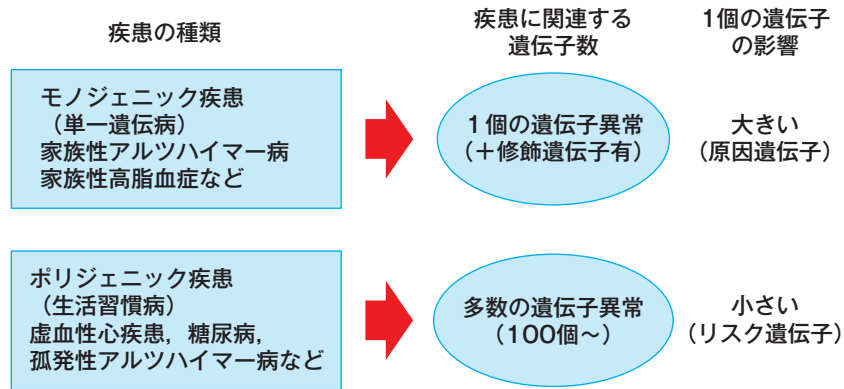


図1 モノジェニック、ポリジェニック疾患の分類

(筆者作成)

ズ比 1.0 を超えるものは疾患のリスクになると考えられる), 検出力の問題点からすべての遺伝因子が同定できるわけではなく, たとえば日本人の糖尿病では 300 個<sup>1)</sup>, 虚血性心疾患では 170 個<sup>2)</sup> を超えるリスク遺伝因子が見つかるが, これらすべてを合わせても, それぞれの疾患での遺伝率をカバーできていないのが現状である。一方で, 家族性の単一遺伝病はモノジェニック疾患とも呼ぶことができ, ひとつの遺伝因子異常がある特定の疾患を発症する(図1)。

近年, このような遺伝因子群を網羅的に同定するために, 大規模なゲノム解析が行われるようになってきた。大規模ゲノム解析の特徴として以下の3点があげられる。①ヒトのサンプル(DNA)が出発点であり, 成果をダイレクトに臨床現場に戻すことが可能になる。②研究仮説を作らないために主観が入らないことから, 未同定かつ思いもよらぬ疾患関連遺伝子やパスウェイを発見し, まったく新しい切り口で疾患研究を進めることにより, 革新的な診断法, 治療法の開発につながる可能性が高い。③ゲノム情報は生涯不変であることから, 一度, 遺伝因子の情報を得るだけで, 将来の疾患発症も含めた, さまざまな形質・表現系の予知・予測が可能となる。

近年進められている大規模ゲノム解析として, 全ゲノムにわたる一塩基多型(single nucleotide polymorphism : SNP)を用いたゲノムワイド関連

解析 (genome-wide association study : GWAS), 最近急速に発展してきた次世代シーケンサーを用いた, 全エクソン, 全ゲノム配列解析をもとにしたGWASなどがあり, 生活習慣病をはじめとしたさまざまな疾患の解明に大きく貢献している。特にGWASは仮説を作ることなく, 遺伝因子群を網羅的に同定する手法として, 近年多くの生活習慣病感受性遺伝子の同定に利用されている。

GWASは2000年代初頭に我々(理化学研究所)が, 全ゲノムにわたる10万個の遺伝子領域SNPを用いて心筋梗塞を対象として行い, 世界に先駆けて報告した手法である<sup>3)</sup>。この大規模ゲノム解析から, 遺伝統計学, 分子生物学的解析を組み合わせることにより, 日本人における心筋梗塞の遺伝因子, 病因解明から創薬開発のヒントを得ていることも事実である<sup>4)~7)</sup>。

2000年代初頭から始まった国際ハップマップ計画(アフリカ人, アジア人[日本人約100人], 欧米人のジェノタイプング)<sup>8)</sup>により, 全ヒト染色体のSNPが地図化され, SNPハプロタイプ, 連鎖不平衡(linkage disequilibrium : LD)関係にあるSNP群の全ゲノム上の分布がデータベース化された。LDとは, 物理的に近い距離にあるゲノム領域は組み換えが起こらず, 集団内で特定のSNPと他のSNP(群)間で組み換えがなく, 次世代に受け継がれていく現象であり, ひとつのSNPのジェノタイプがわかれば, LD関係にある

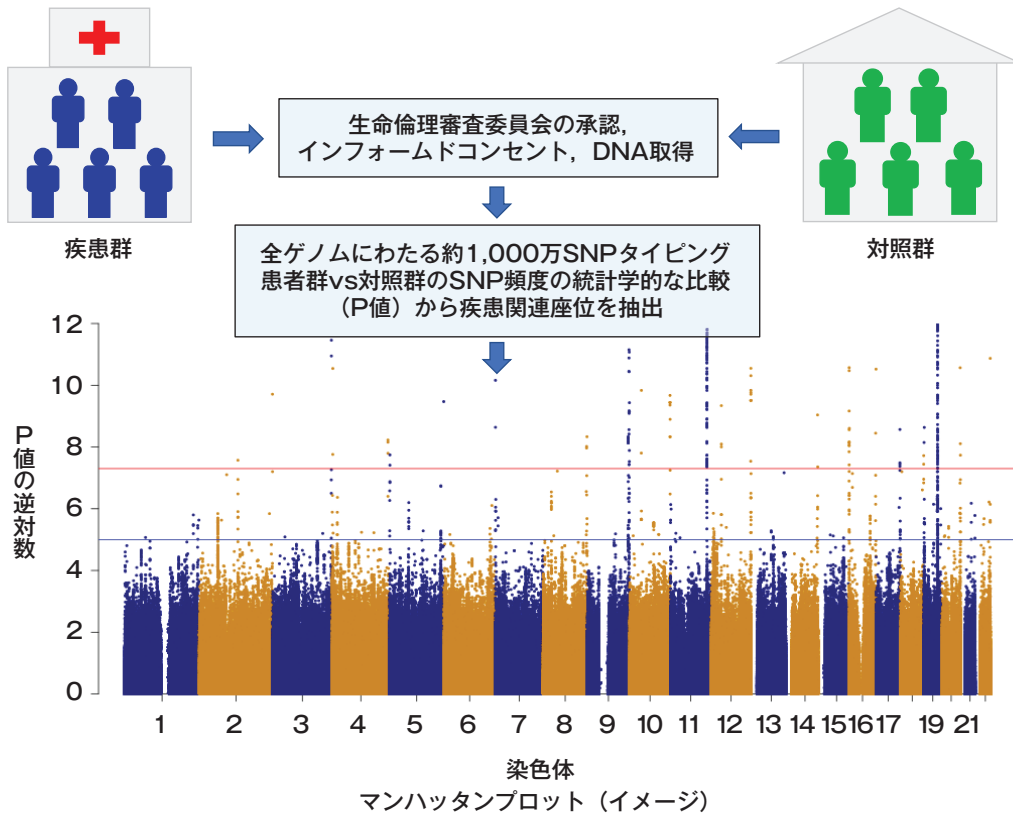


図2 ゲノムワイド関連解析(GWAS)の概要

1個のプロット=1個のSNP, 赤線: GWAS有意性, 青線: 示唆的有意性P値( $P = 1 \times 10^{-5}$ 未満の関連を示したバリエント)

(筆者作成)

他のSNP(群)のジェノタイプもほぼ推測できる。したがって、このハップマップ計画の成果としては、すべてのSNPマーカーを計測しなくても、代表SNPマーカーを計測すれば効率よく多くのSNP情報が得られることを実現したところにある。

この代表SNPをマイクロチップ上に貼り付けてジェノタイピングを行うのがSNPチップマイクロアレイであり、一挙に全ゲノムジェノタイピングを加速させた。さらに1000人ゲノムプロジェクト<sup>9)</sup>による、各民族における深度のあるゲノム配列多型の検出とLD関係情報の整理もあり、現在のGWASプラットフォームとして民族特異的なジェノタイピングアレイも使用できる。また、この全ゲノムリファレンスパネルを利用したインピュテーション解析(実際にはジェノタイピングしていないSNPジェノタイプを予測すること)に

より、ゲノム全体にわたる約1,000万個以上のSNPデータを復元し、迅速かつ体系的に表現型や質的量的座位における形質の関連解析を行うことができるようになった(図2)。

関連解析を行う統計的手法としては、メンデル遺伝学を仮定した顕性(優性)・潜性(劣性)、加算モデルや従来の疫学的手法と同様に回帰モデルが使われ、帰無仮説( $\beta = 0$ )を否定するSNPについて有意水準を設定( $\alpha$ 値)することで、確率(P値)により評価する。GWASにおける有意水準はマルチプルテスト(インピュテーション前のSNP数約100万が使われている)を考慮し、 $P < 5 \times 10^{-8}$  (0.05/100万)と通常設定する。このような基盤を背景として、GWASデータは、これまでに200以上の形質について約5,000論文の報告が成されている(<https://www.ebi.ac.uk/gwas/>)。

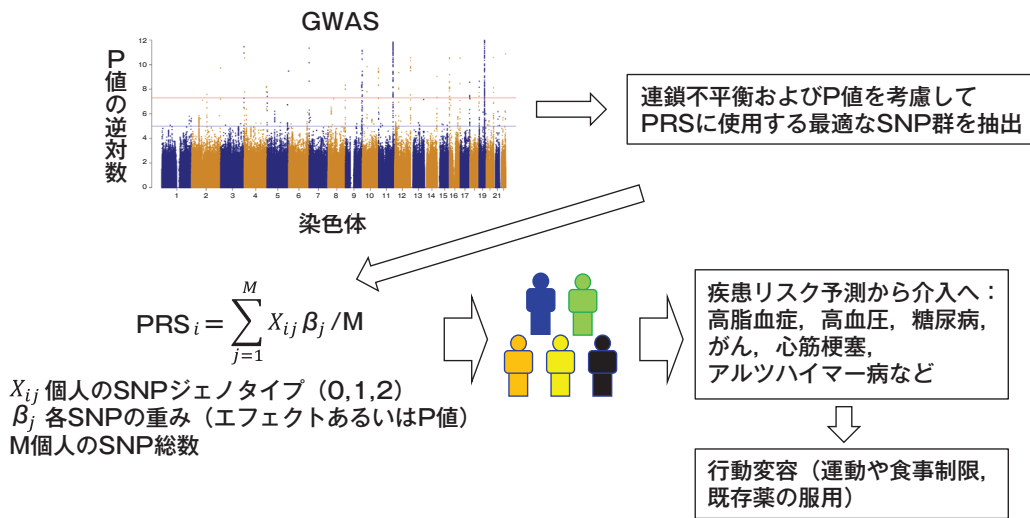


図3 GWAS データによるポリジェニックリスクスコア(PRS)の構築から疾患の発症を早期に予測 (筆者作成)

さらに、全ゲノム解析において正確な遺伝因子情報、大規模データを得るためにもっとも重要なことは、解析する試料数および正確な臨床情報であることは言うまでもない。このような大規模解析に耐え得るサンプル数、情報を収集、提供できるのがバイオバンクである。冒頭に述べたとおり、本邦においては三大バイオバンクが遺伝子解明に大きく貢献してきている。

我々の所属する国立長寿医療研究センター(NCGG)メディカルゲノムセンターのバイオバンク (<https://www.ncgg.go.jp/mgc/biobank/index.html>)はNCBNに属しており、認知症などの老年病を中心として被験者をリクルートし、2020年12月時点で、疾患サンプル約1万例、対照群として約1万例を超える血液、DNAサンプルおよび臨床情報を有すると共に、これらの外部提供も積極的に行っている。また、全ゲノムジェノタイプングデータ 約2万例、全ゲノムシーケンス(WGS)データ 2,500例、全エクソームデータ 750例、オミクスデータ 6,000例(全RNA配列解析1,000例、マイクロRNA発現情報5,000例)についても蓄積済みであり、データベース化すると共に、GWAS等のゲノム解析、遺伝子発現情報統合解析およびデータ共有(共同研究や公的データベースへの登録)も進めている。

## II. ポリジェニックリスクスコアの構築と臨床応用

遺伝因子を使った疾患リスクのスコア化については、GWASが進められた当初から考えられてきており、GWAS上位の10個程度を使ったジェネティックリスクスコア(GRS)として報告されてきている。GRSを用いた予測評価の正確性はそれほど高いものではないと考えられるが、指標として、GRSを用いたDTC遺伝学的検査(direct-to-consumer genetic testing)がいくつかの企業などから提供されている。一方、近年の数万人~数十万人に及ぶ大規模GWASが進むにつれて、もっと多くの遺伝因子を利用したポリジェニック(多遺伝因子)リスクスコア(polygenic risk score: PRS)という概念が浮かび上がってきた<sup>10,11)</sup>。

再度になるが、個々人のゲノム情報は生涯不変であり、一度算出した個人のPRSは生涯変化せず(計算方法が同じであれば)、予知診断や疾患の階層化あるいは薬剤の効力、副作用予測に有用であると考えられるため、多くの遺伝因子(ポリジェニック)が関係する生活習慣病や、形質を予測する方法として、正確なPRSアルゴリズムの構築が精力的に進められるようになった。もともとの方法論はシンプルで、基本的にはGWASより関



連 SNP 群を抽出し、個人における各 SNP ジェノタイプ 0, 1, 2 にエフェクトサイズ(オッズ比等)で重みづけした値の各 SNP の総和を算出する(図 3)。しかし、PRS を臨床応用するためには、異なる集団(同じ日本人でも地域が違うなど)においても高い正確率をもって疾患発症予測を再検証できるかが課題となってくる。

最近では正確性を増すために、年齢、性別に加え、各 SNP における LD 関係の情報を詳細に加味した PRS 構築アルゴリズムが開発されてきており、海外における虚血性心疾患、心房細動などの発症を予測する PRS 研究では、ある程度高精度で発症を予測できることが示されている<sup>12)</sup>。しかし、結局のところ正確性を上げるためには、より多くのサンプルを用いた検出力の強い GWAS、あるいはそのメタ解析を用いた SNP 群抽出が鍵となる。

実際に、我々が最近報告した虚血性心疾患の日本人における大規模 GWAS (BBJ サンプル 17 万人)データと、欧米人 48 万人の GWAS データを統合して構築した民族横断型 PRS が、日本人データあるいは欧米人データそれぞれより算出した PRS と比較して、もっとも正確に日本人の虚血性心疾患発症を予測できることが判明している<sup>2)</sup>。

今後、PRS の臨床応用として期待されることは、リスクを知った上で、生活習慣、すなわち食生活の改善、飲酒、喫煙の抑制、適度な運動の実施などといった行動変容、意識改革を助長することである。現時点では、PRS を知らせた方々において、行動変容の有無を統計的に精度が確保された方法で調べた報告は見当たらないが、欧米の研究では、異なる 3 つの前向きコホートからなる数万人のデータから、遺伝的リスクの高低に関わらず、動脈硬化や虚血性心疾患の発症が、行動変容により抑制できることが示されている<sup>13)</sup>。言い換えれば、若年で PRS を測定し、リスクが高い方々は、早期の段階から積極的に行動変容すれば疾患の発症を阻止、遅延できることを暗示しており、PRS の有用性を示している。

## おわりに

日本人における生活習慣病の正確な PRS を構築するためには、さらなる大規模 GWAS (おそらく 100 万人規模)を行う必要性など、未だ課題は残っている。我々は最近、NCGG バイオバンクサンプルを用いた血液細胞における RNA 発現データ、一般の健康診断で行われる特定の生化学検査、および遺伝因子データを組み合わせることにより、高精度のアルツハイマー病発症予測モデルを構築できることも確認しており<sup>14)</sup>、オミクス、臨床情報を統合することにより、さらに正確な疾患予知法が確立できるものと考えている。また、PRS を臨床現場で使用するためには、医師をはじめとした医療関係者が、遺伝因子データと疾患の関係、疾患遺伝学を理解する必要があり、医療関係者間での知識や情報の共有も課題になる。

## 利益相反

本論文に関して、筆者が開示すべき利益相反はない。

## 文献

- 1) Spracklen CN, et al : Identification of type 2 diabetes loci in 433,540 East Asian individuals. *Nature* 2020 ; **582** : 240-245.
- 2) Koyama S, et al : Population-specific and trans-ancestry genome-wide analyses identify distinct and shared genetic risk loci for coronary artery disease. *Nature Genetics* 2020 ; **52** : 1169-1177.
- 3) Ozaki K, et al : Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. *Nature Genetics* 2002 ; **32** : 650-654.
- 4) Ozaki K, et al : Functional variation in LGALS2 confers risk of myocardial infarction and regulates lymphotoxin-a secretion in vitro. *Nature* 2004 ; **429** : 72-75.
- 5) Ozaki K, et al : A functional SNP in PSMA6 confers risk of myocardial infarction in the Japanese population. *Nature Genetics* 2006 ; **38** : 921-925.
- 6) Ozaki K, et al : SNPs in BRAP associated with risk of myocardial infarction in Asian populations. *Nature Genetics* 2009 ; **41** : 329-333.
- 7) Ozaki K, et al : Molecular genetics of coronary artery diseases. *J. Hum. Genet.* 2016 ; **61** : 71-77.
- 8) The International HapMap Consortium : The International HapMap Project. *Nature* 2003 ; **426** : 789-

- 796.
- 9) The 1000 Genomes Project Consortium : A global reference for human genetic variation. *Nature* 2015 ; **526** : 68-74.
  - 10) Dudbridge F: Power and predictive accuracy of polygenic risk scores. *PLOS Genetics* 2013 ; **9** (3) : e1003348.
  - 11) Lambert SA, et al : Towards clinical utility of polygenic risk scores. *Human Molecular Genetics* 2019 ; **28** (R2) : R133-R142.
  - 12) Khera AV, et al : Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature Genetics* 2018 ; **50** : 1219-1224.
  - 13) Khera AV, et al : Genetic Risk, Adherence to a Healthy Lifestyle, and Coronary Disease. *N Engl J Med.* 2016 ; **375** : 2349-2358.
  - 14) Shigemizu D, et al : Prognosis prediction model for conversion from mild cognitive impairment to Alzheimer' s disease created by integrative analysis of multi-omics data. *Alzheimer's Research & Therapy* 2020 ; **12** : 145.