

## 精神科ゲノム学

## ——一塩基多型 (SNP) とコピー数変異 (CNV) を用いた遺伝子関連解析——

池田 匡志

近年のゲノム解析の進展は、精神疾患を含めた複雑疾患の疾患感受性遺伝子同定に革命的な結果をもたらし続けている。本総説では、方法論で主体となっている全ゲノム関連解析 (GWAS) の結果を理解するため、一塩基多型 (SNP) とコピー数変異 (CNV) とは何か、そして関連解析の基本についてを概説する。加えて、最新の精神科ゲノム学の研究結果を紹介する。

<索引用語：統合失調症，全ゲノム関連解析，精神科遺伝学>

## はじめに

精神科遺伝学 (psychiatric Genetics) における分子遺伝学的研究は、1990 年代の連鎖解析から関連解析がメインとなって以降、急速に進展している。現在までの 20 数年間に、30 億塩基対解読をとりあえずめざした「ヒトゲノムプロジェクト」、一塩基多型 (Single Nucleotide Polymorphism : SNP) をはじめとした遺伝子多型のカタログ化をめざす「HapMap 計画」、1,000 人の全ゲノム配列解読をめざす「1,000 ゲノム計画」など、ホールマークとされるビッグプロジェクトも完遂・継続している。

そうした恩恵もあり、50~100 万個の SNP を一気に決定できる方法論が一般化し、全ゲノム関連解析 (genome-wide association study : GWAS, ジーワスと発音することが多い) が 2010 年頃より盛んに行われるようになった [このような網羅的解析の場合、genetics という用語ではなく、genomics (ゲノム学) を用いる]。そして、GWAS は、驚くべき革命的な結果をもたらし、それまで同定に失敗していた精神疾患のみならず、多くの複雑疾患の疾患感受性遺伝子 (“リスク” 遺伝子)

の同定に成功した (なお、メンデル遺伝形式をとる単一遺伝子疾患では、変異が発症の決め手となるため、原因遺伝子あるいは責任遺伝子と呼ばれる。他方、複雑疾患で疾患と関連する遺伝子は、疾患感受性遺伝子と呼ばれる。後述するリスクの概念で理解でき、リスク型をもつと疾患の発症危険率が上昇する)。

本総説では、GWAS の結果を理解するため、その基盤となる遺伝子多型 (SNP)、GWAS で用いる方法論 (マイクロアレイ) の生データから算出できるコピー数変異 (Copy Number Variant : CNV) とはどのようなものかについて概説する。また、「GWAS」の「AS」部分、Association Study・関連解析のイメージを説明する (Study は普通「研究」と訳されるが、日本では GWAS の場合、関連「解析」と訳されることが多い。GWAS を初めて行った理化学研究所の先生が使用しているからではないか、と推測する)。そして、最後に蓄積されつつある精神疾患の GWAS・CNV 研究の知見を振り返る。

なお、「臨床家の先生にわかりやすい総説にする」ことを最大の目的としているため、多少不正

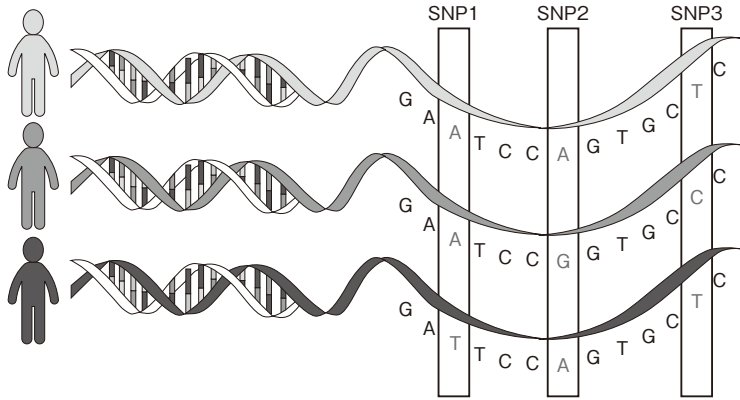


図1 一塩基多型 (Single Nucleotide Polymorphism : SNP)

確な部分もあることをお許し願いたい。

## I. SNP・CNV と関連解析

### 1. SNP

SNP (スニップと発音する。複数形の場合は、SNPsで、スニップスと言う) は、最もありふれた遺伝子多型であり、GWASで主に用いられる「遺伝子配列の少しの違い」である。正確には、「ゲノム配列において、一塩基の変異が、ある集団の中で1%以上の頻度で見られるもの」をSNPと言う(図1)。例えば、AがGに変わるSNPは、A>Gと表現し、頻度の高い方(メジャーアレル：個人によってゲノム配列に差がある場所における、1つの染色体上の特定の配列で、頻度の高い方)を左側に記載する。おおよそ、500~1,000 bp (base pair) に1個存在するため、そして変異型の頻度がある程度高いため、マーカーとして用いられる。この「マーカー」の意義においては、SNPは機能的に意味をもつ必要はない。もちろん、位置により、変異型が遺伝子の発現を変化させる場合もあるし、アミノ酸配列を変える場合もありうるが(非同義置換)、ほとんどのSNPsは、機能的意義をもたない。したがって、GWASで同定される「有意なSNPs」は、それ自体に生物学的意義はないことが多く、その近傍に「真の」疾患感受性遺伝子多型があることに留意が必要である。

### 2. CNV

ヒトのゲノムは、両親から一対のゲノムを受け継ぐことで2コピーをもつ。しかし、特定の領域が欠失(deletion)して1コピー以下となったり、重複(duplication)して3コピー以上になる場合がある。このような「変異」でかつ、おおよそ1 kb (=1,000 bp) 以上の場合をCNVと呼ぶ(図2)。病態生理に関連しない場合もあるが、病的意義をもつ場合は稀なことが多いので、解析においては1%以下の頻度のCNVに限定した解析を行うことが多い。

### 3. 関連解析

SNPやCNVが疾患と関連するかどうかを検査する統計的処理にあたる。例えば、この理解においては、タバコと肺がんの関係がわかりやすい。タバコが肺がんのリスク因子であることは、周知の事実であるが、これは、タバコを吸う・吸わないの2群と、肺がん(+ )と(-)の2群を比較する2×2表を考える。

SNPの場合も同じで、アレル、疾患の有無で2×2表をつくり(図3)、自由度1の $\chi^2$ 検定で、アレル頻度に関する検定を行う(その他の解析法もあるが、ここでは詳細は述べない)。

この際注意が必要なのは、数十万個の検定を行うため、有意水準の設定に特段の注意を払わなけ

ればならないことである。一般の統計解析でよく用いられる有意水準は、 $P < 0.05$ が多いが、GWAS の場合は  $P < 5 \times 10^{-8}$  の有意水準を用いる (genome-wide significance とされる)。そしてそれ以下にならないと有意としないことが多い。これは SNPs 数などから算出された大まかな基準であるが、経験的にこれ以下であれば、追試されることが多く、広く用いられている。

なお、この関連性の全体像を把握するため、マンハッタンプロットが使用される。X 軸には染色体・bp 順に SNP を並べ、Y 軸には、 $P$  を  $-\log_{10}(P)$  で示してプロットしたものである。有意になればなるほど上へプロットされ、まるでマンハッ

タンのようにみえるため、そのような洒落た名前がつけられた (図 4)。ちなみに、genome-wide significance は、 $-\log_{10}(5E-8) = 7.3$  (Y 軸) にあたる。

他方、効果の大きさ (エフェクトサイズ) を考えることも極めて重要であり、binary な形質の場合は、学生時代に習ったオッズ比 (=  $ad/bc$ ) で示すことが多い (図 3)。

もちろん、民族差などを考慮するため、それを示す主成分や、時には性別・年齢などを補正因子としてロジスティック回帰分析をする。しかし、 $P$  値やオッズ比は、アレルワイズの検定と同様に示すことができる。

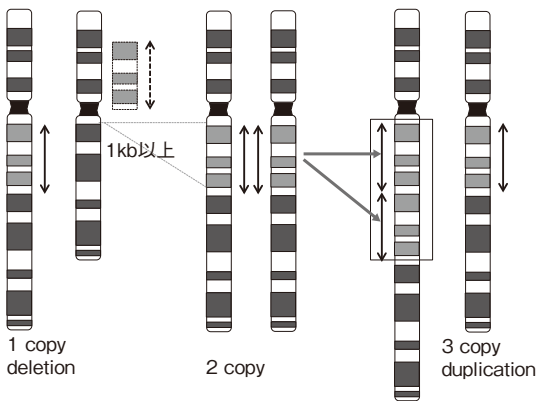


図2 コピー数変異 (Copy Number Variant : CNV)

図 1 の SNP2

	マイナーアレル (例えば "G")	メジャーアレル (例えば "A")
疾患 (+)	a	b
疾患 (-)	c	d

オッズ比 :  $\frac{a/b}{c/d} = ad/bc$

図3 SNP 関連解析における 2×2 表

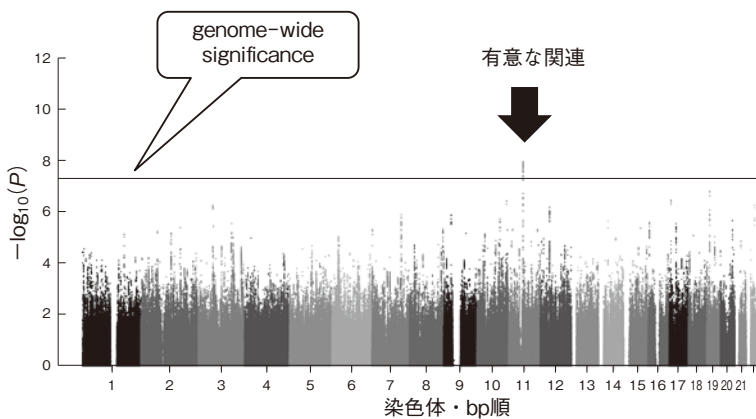


図4 マンハッタンプロットの例 (文献 2 より改変して引用)

表1 PGCによるCNV解析の結果

染色体	start (bp)	end (bp)	座位 (遺伝子)	タイプ	オッズ比 (95%信頼区間)	領域の P 値
22	17400000	19750000	22q11.21	deletion	67.7 (9.3~492.8)	$5.70 \times 10^{-18}$
16	29560000	30110000	16p11.2, proximal	duplication	9.4 (4.2~20.9)	$2.52 \times 10^{-12}$
2	50000992	51113178	2p16.3(NRXN1)	deletion	14.4 (4.2~46.9)	$4.92 \times 10^{-9}$
15	28920000	30270000	15q13.3	deletion	15.6 (3.7~66.5)	$2.13 \times 10^{-7}$
1	144646000	146176000	1q21.1	deletion + duplication	3.8 (2.1~6.9)	$1.50 \times 10^{-6}$
3	197230000	198840000	3q29	deletion	INF	$1.86 \times 10^{-6}$
16	28730000	28960000	16p11.2, distal	deletion	20.6 (2.6~162.2)	$5.52 \times 10^{-5}$
7	72380000	73780000	7q11.23	duplication	16.1 (3.1~125.7)	$1.68 \times 10^{-4}$

位置は hg18 を使用, family-wise correction で有意なもののみを記載。  
(文献3より改変して引用)

## II. 最新の精神疾患 GWAS, CNV 研究の知見

精神疾患のなかで、最も精力的に研究されているのは統合失調症であろう。その遺伝率は80%と高いため、遺伝子から病態生理へアプローチする意義が高いことが理由である。本稿では、統合失調症におけるGWASとCNV解析の結果をサマリーする。その他の疾患は、他のレビューを参照されたい<sup>1)</sup>。

### 1. 統合失調症のGWAS

対象となっているサンプル数は精神疾患のなかで最大であり、かつ、最も多くの疾患感受性遺伝子が同定されており、現在は100個以上の領域が有意な関連を示している。

特に、GWASのメタ解析を行う Psychiatric Genomics Consortium (PGC) が行った研究では、追試研究を含め15万人程度の症例-対照サンプルなどを用い、108個の領域で有意な関連を報告した<sup>4)</sup>。80個程度は、過去に知られていない新規遺伝子であったが、既知の候補遺伝子のなかで最も勇気づけられたのは、ドパミンD<sub>2</sub>受容体遺伝子で有意な関連を得たことである。このことは、今まで経験的知見から仮説を立てたドパミン仮説が一部正しいことを意味し、われわれが処方している抗精神病薬が、病態生理に基づいたものである可能性が示された(すべての患者にあてはまるわけではないことは、後述する)。

ただし、これら「有意」となったSNPsのエフェクトサイズが、極めて小さいこともまた意義深い。そのオッズ比の中央値は1.08であり、リスクアレルをもつ人は、もたない人に比べて統合失調症になるリスクが8%上昇するのみである。したがって、一つ一つの「リスク」SNPの有無が診断などの意味をもたないことは一目瞭然である。これは、他の主要精神疾患である双極性障害やうつ病の場合でもまったく同じことがいえる。そして、リスクとなっている各SNPは、相加的に疾患に寄与し、何千~何万個が疾患の発症にかかわっていることが推察されている<sup>4)</sup>。

### 2. 統合失調症のCNV解析

病態生理に関連するCNVは稀であることが多い。言い換えれば、deletionやduplicationのような遺伝子数の多寡に影響するような効果の大きな変異は、そのエフェクトサイズが大きいことを暗に意味し、遺伝的適応度や自然選択などに深く関連する。現在までに最もインパクトある統合失調症CNV解析は、やはりPGCが行った研究である<sup>3)</sup>。複数の領域が有意であり(この場合はSNPで用いた有意水準 $5 \times 10^{-8}$ ではなく、false discovery rateなどを用いた多重比較補正を行っている)、そのオッズ比は4~70とSNPのそれに比較して大きいことが見て取れる(表1)。

しかし、ここでもCNVをもっているからと

いって統合失調症発症に必ず至るわけではなく、あくまでリスク論であることに注意が必要である(リスクCNVをもつ健常対照者も存在することから自明である)。

### おわりに

最近では遺伝子検査に関連する広告をインターネットで目にする機会が増えた。これらは Direct to Consumer (DTC) 遺伝子検査といわれるものであるが、今回紹介したような、GWAS の方法論・結果を応用している。そうした会社では、生活習慣病や体質など、多くの形質を情報のターゲットとするが、精神疾患(統合失調症や双極性障害など)のリスク情報はセンシティブであるため、扱うことはほぼない。しかし、例えばある業者ではアルツハイマー型認知症やアルコール依存症・心的外傷後ストレス障害のリスク情報を消費者に提供しようである。

ある日、こうしたDTCの「リスク」結果を手にした患者が、例えば「×××障害のリスクがあるんですが」などと医師に尋ねるかもしれない。そのとき、「精神疾患・関連表現型」に関しての「SNPを基盤にした結果」であれば、本稿で述べたエフェクトサイズのことを思い出していただきたい。そして、「そのSNPだけで決まるものでもないし、そもそも1つのリスクをもっていただけ」として診断に使えるものではまったくないのですよ」と自信をもって患者に伝えていただきたい。これは将来にわたっていえることであると確信する。なぜなら、ほぼ確実に、診断に期するほどの大きなエフェクトサイズをもつ「精神疾患リスクSNP」が今後発見されることはないからだ(もし存在するならば、すでに見つかっているはずである。また、他の身体的疾患に関して、同じようなことをきかれた場合も、ほぼ同様の返事で対応可能である)。

もちろん、エフェクトサイズの大きいCNVや、single nucleotide variant (SNV: 全ゲノムシーケンス解析や全エクソンシーケンス解析で同定されることが予想される)はその限りではないが、そうした「変異」に関して、遺伝カウンセリングなしにサービスの提供をする無茶な業者はおそらくない(と信じたい)。

今後遺伝情報は、医療のなかでも貴重な情報として利用されるであろう。そして、そうした包括的な情報が患者に最適な医療を提供できることは確実である。オバマ前アメリカ大統領が推し進めた precision medicine initiative はまさにそれであるが、今後もこの方向性が推進されていくことを切望する。そして、臨床医も遺伝子解析に少しでも関心をもっていただき、研究結果にアンテナを張りつづけていただければ幸いである。所詮結果は、2×2表のような単純なものから算出されたものであるのだから。

なお、本論文に関連して開示すべき利益相反はない。

### 文 献

- 1) 池田匡志: 双極性障害のゲノム研究—全ゲノム関連解析を中心に—. *Legato*, 3 (4); 193-197, 2017
- 2) Ikeda, M., Takahashi, A., Kamatani, Y., et al.: A genome-wide association study identifies two novel susceptibility loci and trans population polygenicity associated with bipolar disorder. *Mol Psychiatry*, 23 (3); 639-647, 2018
- 3) Marshall, C. R., Howrigan, D. P., Merico, D., et al.: Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat Genet*, 49; 27-35, 2017
- 4) Schizophrenia Working Group of the Psychiatric Genomics Consortium: Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, 511; 421-427, 2014

## Psychiatric Genomics : Genetic Association Analysis Using Single Nucleotide Polymorphism (SNP) and Copy Number Variant (CNV)

Masashi IKEDA

*Department of Psychiatry, Fujita Health University School of Medicine*

Recent advances in genomic analysis have provided revolutionary results for susceptibility genes of complex diseases including psychiatric disorders. In this review, for better understanding of the results based on the genome-wide association study (GWAS), which is a main stream in the genomic analysis, (1) the overview of single nucleotide polymorphism (SNP) and copy number variant (CNV), and (2) the concept of the genetic association study are summarized. This review also comprises a short summary for the recent results of psychiatric genomics.

<Author's abstract>

<**Keywords** : schizophrenia, genome-wide association study, psychiatric genetics>

---