

2023年2月21日放送

疾患のゲノムワイド関連解析とマンハッタンプロット

国立国際医療研究センター
ゲノム医科学プロジェクト長 徳永 勝士

疾患の遺伝要因を明らかにすることの意義

はじめに、疾患の遺伝要因を明らかにすることの意義を考えてみましょう。

ある一つの遺伝子の異常が発症に結びつく遺伝病の場合は、その原因遺伝子変異が特定されることによって、診断が確定し、適切な治療法の選択が可能になるというメリットが明らかです。一方で、有病率の高いありふれた疾患、例えば糖尿病や高血圧などは「多因子疾患」に分類され、遺伝要因と環境要因の両方が関わり、しかもその遺伝要因は比較的効果の小さい多数の感受性遺伝子変異が合わさっています。このような疾患においても、個人の遺伝的リスクを推定するなど個別化医療に役立つのみならず、発症や病態進展に関わるパスウェイの解明に繋がり、新たな治療薬の開発を加速することができることがわかっています、

多因子疾患に関わる遺伝要因をゲノム全域から探索するために広く用いられている方法がゲノムワイド関連解析 (genome-wide association study: 以下 GWAS と省略します) です。ゲノム全域に分布する数十万個の一塩基多型 (single nucleotide polymorphism: 以下 SNP と省略します) を一挙にタイピングできるマイクロアレイ技術が確立した 2006 年以降、GWAS 研究は大きく発展しました。当初は数百人規模の検体数で実施した研究が多かったのですが、現在では検体数が数万から数十万人規模となる国際的な大規模共同研究も見られます。

GWAS の原理と解析の概略

ここで GWAS の原理と解析の概略について簡単に説明します。

標準的な GWAS では特定の疾患・病態や形質を示す症例群と示さない対照群について、ゲノム全域に分布する SNP をマイクロアレイ技術で解析した後、個々の SNP のアリル頻度を比較します。マイクロアレイによる SNP タイピングは、現在はタイピング受託会社に外注する 경우가多く

なっています。データを外注先より受け取った後は、品質評価を行って一定の基準に達しない SNP を除外し、さらに集団遺伝学的解析を行って不適切な検体を除外します。

関連解析は、疾患や形質に関わる SNP そのもの、あるいはそれと連鎖不平衡と呼ばれる状態にある SNP を同定する方法であり、非血縁の症例群と非血縁の対照群間における SNP アリル頻度の有意な差として、

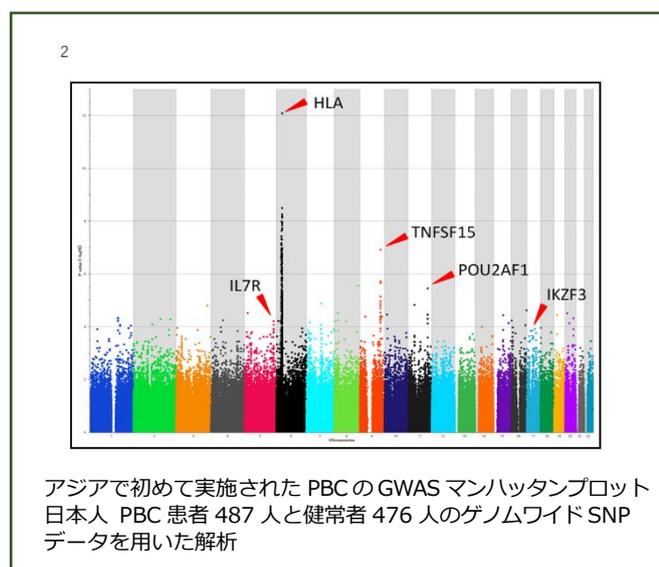
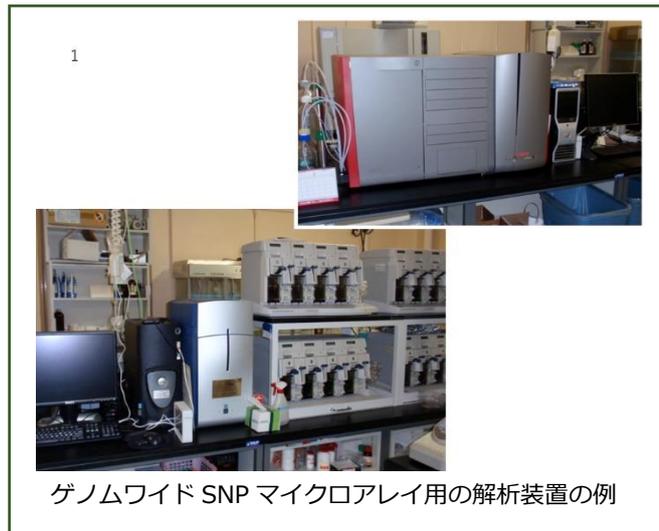
疾患との関連を検出します。すなわち、疾患との関連が検出された SNP の近傍に第一義的な変異が存在する可能性が高いことになります。GWAS によって疾患との関連が検出された後は、より綿密な関連解析マッピングを行ったり、候補変異の機能的な検証を実施することになります。

近年は、ゲノムワイドインピュテーションを行うことが多くなりました。すなわち、タイピングされていない SNP についても、大規模な全ゲノムシーケンスデータを参照パネルとして用いることによって遺伝型を推定し補完することが可能になっています。この結果、擬似的に 5 百万個以上のゲノム全域 SNP について関連解析することができます。なお、非常に多くの SNP を解析してアリル頻度を比較するため、多重比較の補正をどのようにすべきかという課題が生じます。この点については、2008 年頃に、当時の先導的研究者などが集まって議論した結果、 $p=5 \times 10^{-8}$ を統計学的有意水準とする合意が得られ、それ以降の大多数の GWAS において踏襲されています。

マンハッタンプロットについて

次に、「マンハッタンプロット」という用語について説明します。

GWAS の関連解析結果をゲノム全域で俯瞰的に表示できる方法がマンハッタンプロットです。横軸には解析した SNP を染色体上の位置の順番に並べ、縦軸には、それぞれについて得られた p 値の $-\log_{10}$ を表示します。すなわち、p 値の対数のマイナスを表示したものであり、大きい数値、すなわち棒が高いほど統計学的有意性が高いことになります。所々に高い棒が立つ様子をニューヨーク市のマンハッタン島に見られる摩天楼に見立ててマンハッタンプロットと呼ばれています。



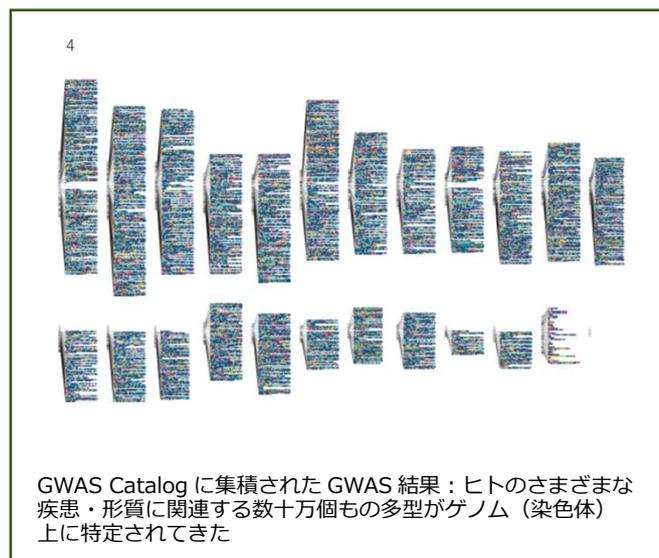
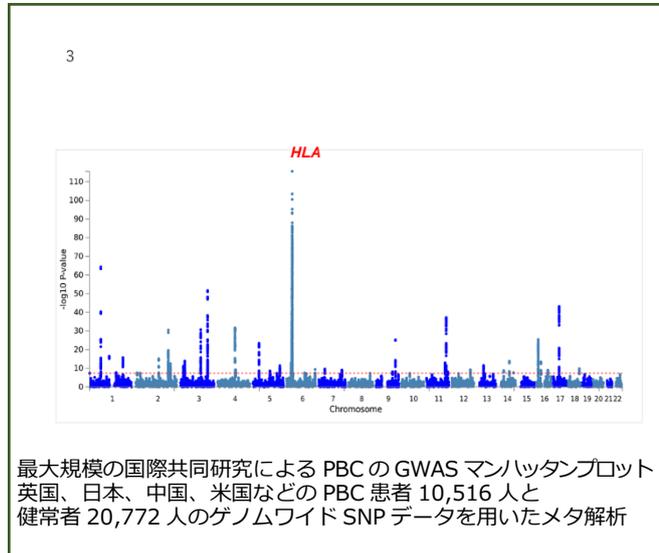
その例として、肝臓の難病の一つ、原発性胆汁性胆管炎 (primary biliary cholangitis: 以下 PBC と省略します) についての国内最大規模の共同研究に参加して行った GWAS のマンハッタンプロットを示しました。日本国内の共同研究では第 6 番染色体上の HLA 領域を筆頭にいくつかのゲノム領域の SNP が関連を示すことがわかります。

多くの多因子疾患・病態や形質では、関連する個々の SNP の効果は小さい場合が多く、これらを検出するためには、症例群、対照群ともなるべく多数の検体を解析することが望まれます。近年は国内外の多施設共同研究が進み、また個々の GWAS 結果を統合的に解析するメタ解析の方法が発達して、大規模な国際共同研究が盛んになっています。上述の PBC についても、イギリスと我々のグループが中心になって、約 1 万人の症例群および約 2 万人の対照群を含む国際メタ GWAS が実施され、56 個ものゲノム領域が検出されました。

なお、これまでに世界で出版された GWAS 結果を収集した最大規模のデータベースである GWAS Catalog によると、2023 年 1 月末時点で 6,245 もの論文が発表され、約 47 万もの有意な関連が報告されています (<https://www.ebi.ac.uk/gwas/>)。

このように、GWAS によって同定された疾患関連 SNP の機能的意義を明らかにすることは、発症のメカニズムを解明し、創薬の標的を探索するためにも有用です。これらの SNP の多くは非翻訳領域に存在することから遺伝子発現レベルに影響を及ぼすと推定されます。さまざまな細胞や組織を用いて遺伝子発現量に影響する多型座位がゲノム網羅的に探索されてきたこと、トランスクリプトーム、プロテオームなどのオミックスデータやメチル化などのエピゲノムデータ、そして 1 細胞解析技術によるデータも集積されつつあること、さらに、多様なバイオインフォマティクス手法の発展が合わさって、疾患関連 SNP の医学生物学的意義の理解も進んでいます。

例えば、大規模な GWAS により検出された多くの SNP について遺伝子セット解析を行うことで疾患発症に重要な生物学的パスウェイが浮かび上がり、発症メカニズムが推定され、創薬のヒントが得られます。また一方、そのパスウェイに影響を与えることが知られている既存薬剤のあ

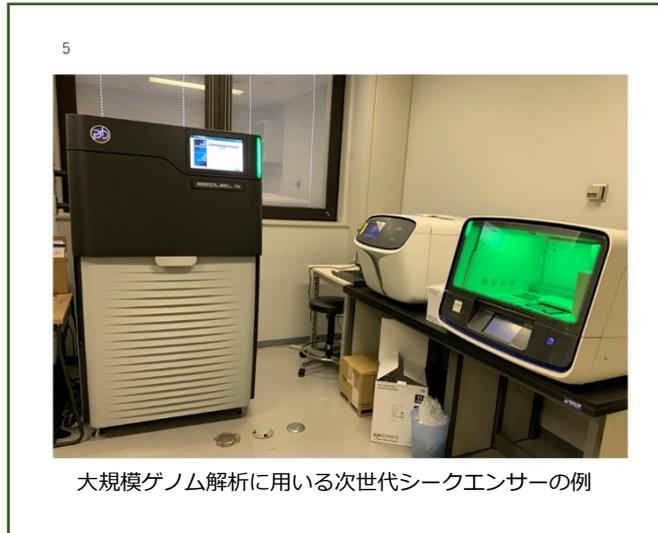


らたな適応疾患を見出すドラッグリポジショニングの報告もあります。上述の PBC の国際メタ GWAS においても、検出された 56 個の SNP について情報解析したところ、いくつかのドラッグリポジショニング候補が同定されました。

ヒトゲノム中で最も多く見られる変異は一塩基レベルのもですが、それら以外にも塩基配列の欠失、重複や挿入、また繰り返し配列などに見られる変異が存在し、いずれも小規模のものから大規模なものまでさまざまです。これらについても関連解析を行うとともに、より頻度の低い変異についても解析するためには、全ゲノムシーケンス解析 (whole genome sequencing: 以下、WGS と省略します) を行う必要があります。最近、

WGS を用いた GWAS も徐々に報告されるようになっており、私達も、感冒薬に起因する稀な過敏症である Stevens-Johnson 症候群について WGS による GWAS を実施しました。その結果、以前に SNP を用いた GWAS で検出、報告してきたものに加えて、稀な変異が集積する遺伝子なども検出されました。

今後も、ゲノム解析技術の発達による大規模ゲノムデータの蓄積、並びにバイオインフォマティクス技術や統計解析手法などの発達によって、ゲノム医療の目覚ましい進展が期待されます。



大規模ゲノム解析に用いる次世代シーケンサーの例

「小児科診療 UP-to-DATE」

<https://www.radionikkei.jp/uptodate/>