

遺伝統計学の世界によろこ

大阪大学大学院医学系研究科・教授

岡田 随象

なぜヒトは病気にかかるのでしょうか？なぜヒトによって異なる病気にかかるのでしょうか？もちろん、生活習慣や怪我、加齢など、様々なイベントによって病気が誘発されます。一方で、生まれながらに特定の病気へのかかりやすさに個人差があることが知られています。それが、遺伝です。遺伝統計学(statistical genetics)は、ヒトの遺伝情報と、病気などの形質情報の関わりを、統計学の観点から検討する学問分野です。本日は皆さんに、遺伝統計学の世界をご紹介します。

ヒトの遺伝情報は30億対の塩基配列で構成されるヒトゲノム配列として、身体を構成する一つの細胞の中に保存されています。46本の染色体に分かれ、各染色体の中に、A、T、G、Cの4種類の塩基で構成された塩基配列が細かく折りたたまれています。ヒトゲノムプロジェクトによりヒトゲノム配列の全容が明らかになったのは、2000年代初頭のことでした。ヒトゲノム配列上で、機能的な働きを有する連続した短い配列を、遺伝子と呼びます。ヒトゲノムプロジェクトにより、ヒトゲノム配列上に何個の遺伝子が、どの場所に位置しているか、の全容が明らかになりました。

ヒトゲノム上の塩基配列は、個人間で少しずつ異なります。ヒトゲノム配列の個人差を、多型(polymorphism)と呼びます。最も代表的なのは、一塩基単位で変異が生じる、一塩基多型(single nucleotide polymorphism; SNP)です。一卵性双生児や親子の例に見るように、ゲノム配列が似ていると見た目(=身体的特徴)も似ることが知られています。現在では、ゲノム配列が似ていると、外見だけでなく、病気のかかりやすさや薬の効きやすさも似ていることがわかっています。そのため、ヒトゲノム多型の情報に基づいた、疾患や医療の研究が行われています。

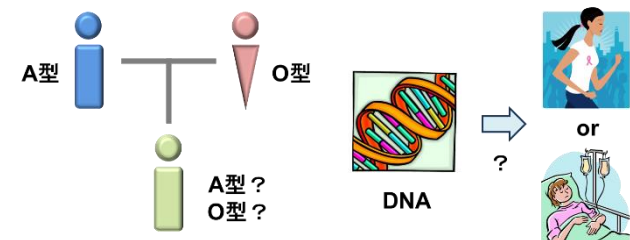
2003年に、国際共同研究プロジェクトである International HapMap Project により、欧米/アジア/アフリカ人集団 270名におけるヒトゲノム多型の概要が解読されました。約 200~300 万の SNP

が明らかになり、これまで手探りで行っていたヒト疾患感受性遺伝子の同定を網羅的に行うことが可能になりました。プロジェクト参加国の中で、日本は一番の貢献を果たしています。2010年には、1000 Genomes Project により、複数集団 2,500名における、約 1億個の SNP が明らかになり、集団中での頻度の低いレアバリエントのカタログが公開されています。

ヒトゲノム配列の解読コストは著しい低下を

ヒトゲノム解析(遺伝統計解析)とは？

・「**遺伝情報**」と「**形質情報**」との結びつきを、統計解析を通じて評価すること。



・**遺伝情報**...生物の細胞の中に、DNAとして保存されている。

ヒト個人間ではほとんど同じだが(99.9%)、**少しずつ違う**。

・**形質情報**...ヒトから得られる情報。**個人間で違うもの**が対象となる。

(血液型、性別、身長・体重、病気 etc.)

示しています。当初は 100 億円/サンプルと高額なコストが必要だった全ゲノム配列解読 (=全ゲノムシーケンス)も、現在では 10 万円/サンプル以下で実施することが可能になり、将来的には 1 万円/サンプルも達成されると考えられています。ヒトゲノム上の数十万か所の代表的な一塩基多型をマイクロアレイ技術を用いて測定する場合は、5000 円/サンプル程度となっており、ヒトゲノム研究の参入障壁は、ますます下がっています。

ヒトゲノム配列解読コストの低下は、ヒトゲノム研究の大規模化を引き起こしています。英国の UK バイオバンクは、50 万人分の SNP マイクロアレイデータを中心とした大規模ゲノムデータを、数千種類の形質情報と共に一般公開しています。数百万人を対象としたヒトゲノム研究プロジェクトも進行中です。以前は、一部の研究施設に所属する限られた少数の研究者だけがヒトゲノム情報にアクセスできる時代もありました。しかし現在では、世界最先端・最大規模のデータを、誰でも解析することができる時代が到来しています。

ゲノムデータの規模が大きくなると、ゲノム解析を通じて多くのことが解明されるようになりました。代表的な成果の一つが、ゲノムワイド関連解析 (genome-wide association study; GWAS) を通じた病気の感受性遺伝子の同定です。ゲノムワイド関連解析は、遺伝統計学の解析手法の一つで、数千人から数十万人を対象に、ヒトゲノム全体を網羅する数千万箇所の SNP のタイピングを実施し、対象形質との関連を評価する手法です。2002 年に日本の理化学研究所で世界に先駆けて実施された解析手法でもあります。2000 年代後半以降、世界中の研究施設でゲノムワイド関連解析が精力的に実施され、現在までに、1,000 以上の形質に対し約 5,000 報のゲノムワイド関連解析が報告されています。糖尿病やがんといった病気から、身長や肥満などの身体測定値、血液検査値などの臨床検査値、食生活習慣など、多彩なヒトの形質情報に関連する遺伝子変異が多数報告されています。新型コロナウイルス感染症の重症化についても、パンデミックの早期に国際共同研究を通じてゲノムワイド関連解析が実施され、重症化に関連する遺伝子変異が見つかりました。

このように、ヒトゲノム解析技術の発展に伴い、大規模ヒトゲノム研究を通じて病気に関連する遺伝子変異が数多く同定された一方で、ゲノム解析の成果をどのように活用すれば病気の病態解明や新規創薬、個別化医療に貢献できるのかは、ほとんどわかっていない状況です。30 億塩基対のヒトゲノム配列はヒトの設計図なわけですから、そのメカニズムを完全に解明すれば、生命現象を再構成できるはずですが、しかし実態としては、そのほとんどは未解明です。どうやって、ヒトゲノム配列に隠された生命現象を引き出すか、が私の一貫した研究テーマになります。

ヒトゲノム研究の大規模化と感受性遺伝子同定

UK バイオバンク

(<https://www.ukbiobank.ac.uk/>)



ゲノムワイド関連解析 (GWAS)

が同定した感受性遺伝子

(<https://www.ebi.ac.uk/gwas/>)



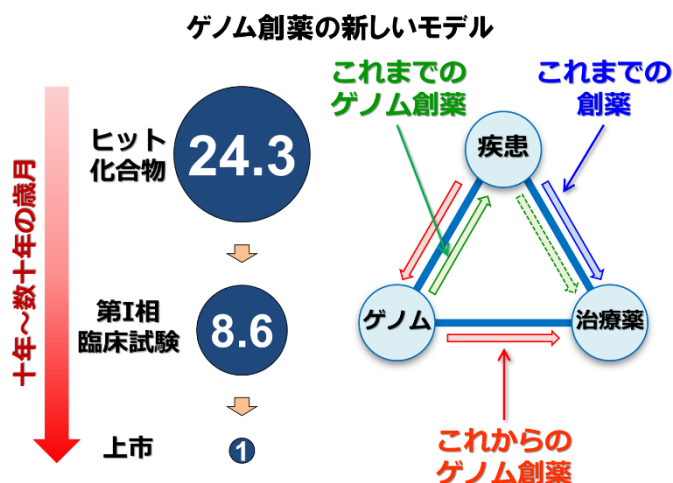
- UK バイオバンクは 50 万人の SNP データ、45 万人の全エクソーム、15 万人の全ゲノムシーケンスを実施し、ゲノム・表現型情報を世界中に安価で分譲することで研究成果を創出している。
- 世界中の研究施設でゲノムワイド関連解析が実施され、2021 年までに、1,000 以上のヒト形質に対し、5,000 以上の論文が報告されている。

私が心がけていることとして、「実験技術・情報解析技術の進歩は、常に私たちの予想を上回る速度で進んできた」という客観的な事実があります。ヒトゲノム解析の分野においても、私が研究に参画した 15 年前からは想像ができないほど、技術革新が進んでいます。おそらく今後も、想像を超える速度で進んでいくものと期待されます。最先端の実験技術・情報解析技術を積極的に導入していくことが、生命科学研究の発展には必要です。また、単に実験機器や解析ツールを購入するだけでなく、その原理を正確に理解することが、革新的な研究には不可欠と考えています。

ゲノム情報から病気の病態を解明する方法として、オミクス解析が知られています。オミクス情報とは、「ゲノム」、「エピゲノム」、「プロテオーム」など、生体情報の網羅的データ(= -ome)をさらに統合して得られる情報を指します。多彩なオミクス情報を分野横断的に統合することで、生命現象につながる知見が得られることが知られています。異なるサンプル集団由来のオミクス情報であっても、共通の情報単位に変換することで、横断的な統合が可能になります。遺伝子というカテゴリに変換することに加え、ヒトゲノム標準配列に沿ったベクトル情報に変換するオミクス解析方法の開発が盛んです。多数の病気に対するゲノムワイド関連解析の結果を、色々な細胞組織におけるエピゲノム情報と横断的に統合することで、病気と細胞組織のつながりのネットワークが明らかになります。例えば、肥満の遺伝的背景における中枢神経系の関与や、免疫関連疾患における免疫細胞の関与などが、明らかになっています。

新しいオミクス情報層の開拓も進んでいます。大阪大学に着任してからは、微生物叢の研究に取り組みました。微生物叢は、宿主であるヒトや動物と共生関係にある多種多様な微生物の集まりで、宿主であるヒトと様々な相互作用を持ち、「第二の臓器」と呼ばれています。微生物叢の個人差は腸管の病気に限らず、2型糖尿病や心血管障害、がんなど多くの病気の原因に関与していることが知られています。微生物叢研究の分野でも、新しい実験技術が革新的な研究成果を生み出しています。従来は、微生物叢に含まれる細菌のゲノム配列の一部のみを解読する 16S rRNA 解析が行われていましたが、現在では、微生物叢の全ゲノム情報を網羅的に解読する、メタゲノムショットガンシーケンス解析が主流になりつつあります。私たちは、メタゲノムショットガンシーケンス解析を活用することで、自己免疫疾患の患者さんの腸内微生物叢の特徴的なメタゲノム情報を同定することに成功しました。最近では、腸内に含まれる細菌に加え、ウイルスの研究も進めています。

遺伝統計学の可能性の一つに、ゲノム創薬があります。新規創薬には長い時間と多大な開発費用がかかりますが、その効率が年々低下しており、創薬プロセスの効率化の重要性が指摘されています。病気の研究から治療



薬を探す「これまでの創薬」に加え、病気のサンプル由来のゲノム情報から治療薬を探す「これからのゲノム創薬」が今後は必要です。海外の製薬企業の研究では、病気のゲノム情報が創薬プロセスを効率化することが報告され、注目が集まっています。私たちは、ゲノムワイド関連解析などの大規模ゲノム解析の成果に基づき、病気の治療薬候補の探索が効率化できることを報告してきました。特に、既存の治療薬の他の病気への適用拡大する、ドラッグ・リポジショニングに貢献できると考えています。一般公開されているゲノム解析の結果から、治療薬候補を探索する遺伝統計解析ソフトウェアの開発を行っています。

最後にお話しするのが、ゲノム個別化医療への挑戦です。世の中には沢山の病気がありますが、同じ病気であっても、人によって病態が異なることが経験的に知られています。全員に画一的な標準治療を行うことも大事ですが、個人の体質に応じて最適化された治療法の確立が期待されています。それが、個別化医療です。2015年のオバマ大統領の一般教書演説で、“Precision Medicine Initiative”という言葉が注目を集めました。特に個人のゲノム情報の違いを活用した個別化医療の提言に重きを置いた点が画期的でした。ゲノム個別化医療の時代の幕開け、という解釈もできます。ゲノム個別化医療の研究分野では、特定の遺伝子変異に着目したり、ヒトゲノム全体に分布する無数の遺伝子変異を統合することで、個人の病気の発症リスクをゲノム情報に基づき予測することが可能となりつつあります。近い将来、自分がどんな病気になりやすく、またなりにくいのかを人生の早期の段階で判定して、個人の特性にあわせた最適なライフスタイルや医療を受けられる時代が来ると考えられます。その社会実装において必要となるプロセスを、遺伝統計学を通じて実現していきたいと考えています。

基礎医学研究の更なる発展に必要なのは、若手人材の育成です。特に日本では、遺伝統計学やバイオインフォマティクスの研究分野の人材不足が指摘されています。私たちの教室では、「遺伝統計学・夏の学校@大阪大学」と題して、夏休みに三日間のサマースクールを開催しています。遺伝統計学の座学から、プログラミングや実際のデータを使ったゲノムデータ解析演習まで、初心者を対象とした幅広い内容となっています。お陰様で、2022年度は300名の方に参加して頂くことができました。講義演習資料は教室のホームページで一般公開しておりますので、興味を持たれた方は、是非ともご覧ください。(http://www.sg.med.osaka-u.ac.jp/school_2022.html)

遺伝統計学 夏の学校 @大阪大学 ONLINE

参加費：無料

夏の大阪で、ゲノム解析の世界に触れてみませんか？

日時：令和4年8月26日(金)～8月28日(日)
(原則として、三日間通じて参加をお願いします)

場所：オンライン形式でのWeb開催
(新型コロナウイルス対策の観点より、本年度はオンライン形式での開催になります)

対象：大学学部生以上
(大学院生・社会人の方も歓迎です。保護者の了解が得られれば高校生も歓迎です)

形式：講義・ゲノムデータ解析演習
(演習に使用するパソコンは、参加者各自で用意して頂く必要があります)
(使用する解析用データ一式は、夏の学校で用意し、設定資料と共に事前配布します)

定員：100名 (応募者多数の場合、所属や学年を考慮して調整させて頂く事があります)

時間帯	一日目 (8/26)	二日目 (8/27)	三日目 (8/28)
午前 (9:00~12:00)	遺伝学入門 統計学入門	統計ソフトR入門	ゲノムデータ 解析演習 ③・④・⑤・⑥
午後 (13:00~16:30)	Linux入門 プログラミング入門	ゲノムデータ 解析演習①・②	

(講義演習内容は、変更となる場合があります)

URL : http://www.sg.med.osaka-u.ac.jp/school_2022.html
申込み : office@sg.med.osaka-u.ac.jp (担当：岡田 随象)

共催：東京大学大学院医学系研究科遺伝情報学、理化学研究所生命医科学研究センターシステム遺伝学チーム
 後援：大阪大学バイオインフォマティクス・イニシアティブ、合原ムーンショットプロジェクト
 AMEDゲノム医療実現推進プラットフォーム事業GRIFIN