自己免疫疾患の病因研究としてのゲノム研究

* はじめに

自己免疫疾患は障害臓器特異性の高低や抗原特異性の高低、男女比、年齢分布、有病率など、様々な点で多彩な疾患の集まりであるが、その多様性を決めている要因の中には遺伝要因がある。本稿では、SNPタイピングによるGWAS(ゲノムワイドアソシエーションスタディ)までの遺伝因子解析による自己免疫疾患の遺伝因子研究の成果を概観したのち、次世代シークエンシング技術をはじめとする技術発展が可能にしつつあるゲノム解析技術が、病因遺伝因子のさらなる解明に向けてどのような役割を果たしつつあるかを解説する。

* 自己免疫疾患の遺伝因子解析の成果
  + 成果のリスト

自己免疫疾患の遺伝性の存在は特定の自己免疫性疾患の家系内集簇や、複数の自己免疫疾患の同一家系内集簇などから疑われてきた。その遺伝因子解析は、他の疾患のそれと同じく、双生児・家系内発症研究による遺伝率の推定に始まり、HLA型を含む標的遺伝子アプローチによる関連遺伝子の同定、家系例・罹患同胞対解析による関連遺伝子座位の検出を経た後、SNPタイピングによるGWAS手法が2000年代半ばから多くの複合遺伝性疾患に適用され、数多くの疾患リスクバリアントが同定された。代表的な自己免疫疾患の解析成果のまとめを表1(…より改変)に引用として示す。なお、表の値は成果の採用基準により、かなり変動することに注意することが必要である。その理由は条件の異なる複数の解析がなされており、それらを統合すること自体がゲノム疫学分野の主要な課題の一つであるという事実とも関係する。実際、遺伝因子解析の成果は日々追加され、その統合も単純ではないため、遺伝因子解析の成果は、データベース化した上で、利用者が用途に応じて使用することが望ましいという考え方に立っている。表に掲載されていない疾患や異なる基準での集計等の用には、GWAS catalog (GWASの成果の検索サイト(1)) OMIM (Online Mendelian Inheritance in Men)単一遺伝性疾患から複合遺伝性疾患までの遺伝因子解析の成果の概説サイト(2) ) Public Health Genomics Knowledge Base(v1.0) (疾患・健康関連のゲノム成果の検索サイト(3))などが利用できる。

表１(4)

* + 成果の特徴と解釈
    - 免疫系遺伝子・薬剤標的候補遺伝子

リスクバリアントを有する遺伝子として挙がったものの中には自然免疫系・獲得免疫系の遺伝子が数多く含まれており、特定の免疫機能のパスウェイ上に複数の関連遺伝子が見出されることもあった。最も大きな存在はHLA領域である。この領域にあるHLA分子をコードする遺伝子には、多くの自己免疫疾患に疾患特異的リスクハプロタイプが見出されている。また、同領域にはHLA分子遺伝子以外にも多数の免疫関連の遺伝子がコードされており、それらに認められる関連もある。非HLA領域にも多数のリスクバリアントが見出されている。表２は商業的遺伝子バリアント情報提供サービス会社(23andMe,DeCODEme)が独自に定めた採用基準で自己免疫疾患リスク用として提供している遺伝子バリアントと自己免疫疾患のリストである(5)。

また多数の疾患リスク関連遺伝子とそのバリアントの同定により、病因的メカニズムの解明が進むだけでなく、治療薬の作用機序や既存治療薬の適用拡大の可能性などについての示唆も得られると考えられている(6)。

* + - 多疾患に関与するバリアントの存在 Pleiotropy

表２を見るといくつかの遺伝子が複数の疾患のリスクバリアントを有していることがわかる(複数の疾患で名前の挙がっている遺伝子をハイライトしてある)。このように、単一の遺伝子のバリアントが複数の疾患に影響することをPleiotropy(多面発現性)と言う。自己免疫疾患が発病機構や免疫異常を共有していることを表していると考えられる。また、このことから、遺伝因子との関連研究の標的として、疾患そのものだけでなく、in vivo, ex vivoな免疫系細胞・免疫反応を標的とすることの有用性が期待され、実際、それらを試料とした各種オミクス実験結果とゲノム解析との統合解析という形で研究が進められている。

* + - アミノ酸置換型バリアントとそれ以外のバリアント

表２には、特定の遺伝子名ではなく” intergenic”とされるバリアントも多数ある。これは、タンパク質をコードするコーディング遺伝子構造ではないところに存在することを意味する。実際、遺伝子名が付与されたリスクバリアントであってもアミノ酸置換を伴うバリアントは少なく、遺伝子の発現効率やオールターナティブ・スプライシングに影響するなど、コードするタンパク質の1次構造以外への影響を介して疾患リスクを上げていると考えられている。このように自己免疫疾患リスクバリアントの大多数はコードタンパク質のアミノ酸置換ではない。アミノ酸非置換型のバリアントの機能性については、「遺伝子の変化」の項に記載する。

* + - 遺伝率はまだ一部しか説明がついていない Missing heritability

双生児研究などから推定された遺伝率が表1に示されているが、これまでに検出された疾患遺伝子座位が説明できる遺伝率はこの値と比べてかなり小さい。この不足分をMissing heritabilityと言う。そもそも双生児研究などから算出した遺伝率の値が高い側にぶれている可能性や、これまでに見出された座位のアレルの相互作用を見積もらないなど遺伝率の算出方法が不足の理由となっている可能性も否定はできないが、やはり、未検出の座位がある可能性は高い。GWASの網にかからなかった座位があるということを意味するが、それには、リスクが強くなくGWASの検出力の弱さに起因するもの、GWASが用いたSNPよりもアレル頻度が低く、結果として連鎖不平衡マッピングで検出されなかったようなレアバリアントに起因するもの、構造バリアントのようにGWASで検出しにくいタイプの遺伝子バリアントに起因するものなどがある。(7)

* ゲノム研究の変化～単純な構図から複雑な世界へ
  + 次世代シークエンサーの登場
    - レアバリアント

GWASが既知SNPを効率的にタイピングする技術を基盤にしていたのに対して、次世代シークエンサー(NGS)は全ゲノムやエクソン領域を網羅的にシークエンシングすることで、未知のバリアントを検出しながらリスクバリアントの探索ができる点が大きく違う。既知SNP方式では、どうしても集団におけるアレル頻度が一定以上あるバリアントのみが対象とならざるを得ないが、NGSを用いた解析では、世界中である個人のみが持つバリアントまでもが対象となる。

* + - De novo バリアント

世界に１つだけのバリアントが検出できるということは、親子をシークエンシングすれば、親は持たず、親の配偶子に起きた突然変異が子に伝わったようなバリアントの検出が可能となる。これをde novo 変異と言う。De novo変異がゲノムワイドに検出できるようになったため、家族歴がない単一遺伝子疾患患者の原因バリアントの同定などが容易となった。実際、家系例による連鎖解析からSNPによるGWASへと比重が移動し、疾患リスク・原因バリアント解析の中心が複合遺伝性疾患となっていたものが、NGSの登場により単一遺伝子疾患研究の成果が再び目立つようになっている。

* + - １細胞シークエンシング

NGSは個々の細胞を試料としてシークエンシングすることもできる。疾患が親から子へ遺伝することに着目し、親から子へ遺伝するのは、個人の全細胞が共有するゲノムであると考えれば、１細胞シークエンシング技術は疾患のゲノム解析に資することはない。しかしながら、NGSはDNA配列を読むだけでなく、エピゲノム・トランスクリプトームの解おうより、ノンコーディングRNA遺伝子が細胞の分化・活動状態の制御においても主要な役割を果たしていることがわかってきた。このように、ゲノムに刻まれる遺伝子はコーディング遺伝子主体であった段階から、ノンコーディングRNA遺伝子を含めたものと変わり、数も機能も飛躍的に増えた段階に変わった。そしてそれらの全容の解明はこれからである。したがってそれらが自己免疫疾患において果たす役割の解明も始ひまっており(9)、まだまだこれからであることがわかる。

図１には、コーディング遺伝子・ノンコーディング遺伝子とその多様な調節機構を示した。DNA上のコーディング遺伝子の発現調節はエピジェネティクスのレベルでの調節がなされ、さらにプロモータ・エンハンサによる転写調節が行われる。そのどちらにもノンコーディングRNAが関与することが知られている。また、転写されたあとにはスプライシングが起き、その結果スプライシングバリアントを作り分けることで１つのコーディング遺伝子から複数のタンパク質分子を生成することができるが、このスプライシングを行うスプライソソームでもノンコーディングRNAが機能を発揮する。さらに作られたmRNA分子の非翻訳領域が作る立体構造が翻訳に至るかどうかに影響を与えることも知られている

特に、この立体構造が細胞の置かれた条件によって変化することで翻訳調節をする仕組みもある(リボスイッチ)。mRNAは核外へと輸送されリボソームにてポリペプチド鎖が翻訳されるが、その過程でmiRNAによる分解制御を受けることがしられている。さらに翻訳されたタンパク質分子の細胞内輸送にもノンコーディングRNAが機能を発揮する。このようにノンコーディングRNAは非常に多彩な役割を持っているのだが、このRNA分子の塩基配列はもともとDNA配列であったから、そこに生じるDNA配列バリアントはこれらすべての機能に影響する可能性がある。(10)

図１　遺伝子バリアントとゲノム・エピゲノム・ncRNA・スプライシング・アミノ酸置換

* + 生殖細胞系列バリアントも体細胞系列バリアントも

従来、疾患の遺伝因子解析と言えば、親から子へ伝えられる生殖細胞系列バリアントの研究のことであった。しかしながら１細胞シークエンシングが可能になり、発生途上におきる体細胞系列変異とその結果としての遺伝的モザイクが引き起こす疾患(11)や、体細胞系列変異を原因とする自己免疫原性の獲得(12)、T細胞受容体、Ｂ細胞受容体に見られるような生理的体細胞ゲノム変化(8)が研究対象となってきており、生殖細胞系列バリアントは遺伝する(hereditary)、体細胞系列バリアントは遺伝しない、と峻別するのではなく、どちらも遺伝子・ゲノム多様性の1形態である(genomic, genetic)であるとして総合的にとらえ、それが疾患をどのように形作っているのかを研究するのが疾患のゲノム解析となっている。

* + ヒトレファレンスゲノム配列のグラフ構造化

21世紀に入ってすぐ、ヒトゲノム配列が発表になった。その配列は唯一１本の配列であった。しかしながら、SNPを中心とする遺伝子バリアントのデータベースを構築するHapMapプロジェクトが進み、NGSによる多人数のリシークエンシングを経て、70億人のヒト集団のゲノム配列は1本のレファレンスゲノムとしてとらえることが難しくなっている。個々の染色体配列は相当な長さに渡って大規模な違いを持っていることがわかっている。それらの違いはAlternative locusとして登録され、それらは1本線ではなく複雑に分岐したグラフ構造として理解されようとしている(13) 。遺伝因子研究はレファレンスゲノムを参照しながら進むので、この変化は自己免疫疾患のゲノム研究にも大きな影響を及ぼすことは間違いなく、特に民族別解析などにその影響が大きいと予想される。

* + 電子カルテとフェノーム解析と PheWAS

最後に、表現型に関する変化について述べる。疾患ゲノム解析では疾患の有無を表現型とすることはほとんどである。しかしながらPleiotropyの項でも書いた通り、免疫系の機能に関するデータを表現型とすることで新たな知見が得られることが期待されている。同様に、疾患の構成要素として各種バイオマーカーや疾患クライテリアの構成要素などを表現型として遺伝子バリアントとの関係を調べることも有用な可能性が大いにある。また、複数の疾患に影響する因子の研究には、複数の疾患の情報を同時に解析することが有用である。このような情報は臨床的に取得され、電子カルテ(EHR:Electrical Health Record)として蓄積されていることが多い。そのように多様なフェノタイプ情報を一括して活用するアプローチがフェノーム解析であり、それらをGWAS的に解析するのがPheWASである(14)。現在のところEHRの情報を活用しての遺伝因子解析に目覚ましい成果が出ているわけではないが、EHR上の情報量は膨大なので、それらを有機的に連携することができれば新たな展開が大いに期待できる。

* 終わりに

GWASによって得られたたくさんの自己免疫疾患の遺伝要因の概説に続いて、ノンコーディング遺伝子の解明の進展、NGS・１細胞解析・電子カルテによる研究対象の拡大・データの膨大化の進行について解説した。図２は、それらを１つの図に表したものである。本稿では以下に列挙したことを述べてきたが、それらの関連を図で確認していただきたい。具体的には、以下が相互に関係しあっている。

1. 疾患遺伝子のためのゲノム解析は、配偶子から受精卵へと受け渡される遺伝子(DNA,ゲノム)を中心に行われること
2. その遺伝子がコーディング遺伝子一辺倒からノンコーディング遺伝子広がっていること
3. 生殖細胞系列のde novo変異が対象になっていること
4. 体細胞系列変異がもたらす様々な変化も対象となっていること
5. 疾患を含む表現型を統合してとらえるフェノームという考え方が有力であり、それを支えるのは電子カルテ等の医療情報であること
6. １細胞解析技術などの進展により、個人の臓器・組織・細胞から細かくデータを取ることができること(中間フェノタイプ)。それは空間的に現象を分解して観察していることに相当すること
7. 受精卵から死まで色々な時刻でのデータ取得も可能であり、時間的にも細かい解析ができること
8. 解析を細かくすることによりマルチオミクスのビッグデータ解析をすることが疾患ゲノム解析の主要なテーマとなっていること

引用文献(等)

1. GWAS catalog <https://www.ebi.ac.uk/gwas/home>
2. OMIM <http://www.ncbi.nlm.nih.gov/omim>
3. Public Health Genomics Knowledge Base(v1.0) <https://phgkb.cdc.gov/GAPPKB/>
4. From Autoimmune diseases — connecting risk alleles with molecular traits of the immune system Maria Gutierrez-Arcelus, Stephen S. Rich & Soumya Raychaudhuri　Nature Reviews Genetics 17, 160–174 (2016) doi:10.1038/nrg.2015.33
5. Eupedia.com <http://www.eupedia.com/genetics/autoimmune_diseases_snp.shtml>
6. Y Okada, D Wu, G Trynka, T Raj, C Terao, K Ikari, Y Kochi, K Ohmura, ...Nature 506 (7488), 376-381
7. Eskin, Eleazar (2015): Discovering Genes Involved in Disease and the Mystery of Missing Heritability. In: Commun. ACM, 58 (10), pp. 80-87, 2015, ISSN: 0001-0782.
8. Identification and Phylogeny of the First T Cell Epitope Identified from a Human Gut Bacteroides Species Maria Elisa Perez-Muñoz , Payal Joglekar , Yi-Ji Shen , Kuan Y. Chang , Daniel A. Peterson Published: December 4, 2015DOI: 10.1371/journal.pone.0144382
9. Biol Proced Online. 2014; 16: 11.The functional role of long non-coding RNAs and epigenetics
10. [Genome Med.](http://www.ncbi.nlm.nih.gov/pubmed/25419237) 2014 Oct 28;6(10):88. doi: 10.1186/s13073-014-0088-0. eCollection 2014.
11. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4276927/>
12. <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0101093>
13. <http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001091>
14. Nature Reviews Genetics 17, 129–145 (2016) doi:10.1038/nrg.2015.36 Published online 15 February 2016