

2.4 SNP による遺伝子マッピング法

前述したように、関連解析、連鎖解析においては、対象疾患に有意に関連、あるいは連鎖する SNP (SNP 群) を見いだすことが第一の目標である。遺伝解析技術の著しい向上に伴い、研究対象とするサンプル DNA を収集する困難さは以前と変わりはないものの、サンプル収集以降、この目標に到達するまでの困難さは以前に比べれば軽減しており、多くの研究者がこの目標に比較的容易に到達するかもしれない。しかしながら、最終的には、どの SNP が真の疾患感受性変異であるかを同定することが必要である (図 2.1、2.5 参照)。他章で詳細に解説されている連鎖不平衡マッピングによる絞り込みは有効な遺伝解析のひとつであるが、連鎖不平衡関係にある複数の SNP の中から、遺伝解析のみで真の遺伝的変異を見いだすことは困難であろう。SNP の病態生理学的な役割を明らかにするための機能解析が必要である所以である。

ひとことで「機能解析」といっても、様々な実験手法 (生物学、分子生物学、細胞生物学など) を駆使する必要があるので、機能解析を行うにあたって、遺伝子機能に影響を与える可能性が高い SNP と、そうでないものとを区別できると、可能性のある SNP のみについての効率よい解析が期待できる。遺伝子の塩基配列からの機能・構造推定 (例えば、転写因子結合領域、細胞膜貫通領域など) と SNP の存在部位情報の組み合わせから、SNP の機能面への影響を予想することは重要である。遺伝子翻訳領域内の SNP のうち、非同義的 SNP (コードしているアミノ酸の種類を変化させるタイプの SNP) は表現型に影響を与える可能性が高いことから、以前より機能解析の対象 SNP として注目されている。最近では、その影響の程度予測を行うアルゴリズムも開発され (Ng and Henikoff 2002)、SIFT (Sorting Intolerant From Tolerant; <http://blocks.fhcrc.org/sift/SIFT.html>)、PolyPhen (Polymorphism Phenotyping; <http://genetics.bwh.harvard.edu/pph/>) などのサイトでは、機能変化を伴う可能性の高い SNP を選別することもできる。あるいは、ヒト遺伝子の発現量多型と関わる SNP についての豊富な知見 (Morley et al. 2004; Cheung et al. 2005 など) を基にして、機能解析を行う必要のある SNP の優先順位をつけることも可能となってきた。しかしながら、例えば、ヒト MDR1 遺伝子の同義的 SNP (アミノ酸の種類を変化させない SNP) であるために、機能変化は引き起こさないと考えられてき

た) が、そのタンパク質の基質特異性に影響を与えることが報告されている (Kimchi-Sarfaty et al. 2007)。このことは、従来のような SNP 位置情報のみからの絞り込みでは、真の感受性変異を見落とす可能性があることを意味している。また、ひとつの SNP がそれ単独で疾患に関わる場合以外にも、SNP ハプロタイプとして、あるいは他の遺伝子上の SNP との相互作用を介して、疾患感受性を規定している場合も十分考えられる (Akagawa et al. 2006; 2007 など)。SNP 単独での効果は弱いことも予想されるので、疾患と関わるヒト型ハプロタイプや相互作用 SNP 群をもつモデル動物やモデル細胞などを作出し、病態解析を行う必要性も高まってくるであろう。

いずれにせよ、対象疾患の感受性変異をヒトゲノム上に「マッピング」するためには、統計遺伝学のおよび実験科学的に、原因 (SNP) と結果 (表現型変化) との関わりについて説明することが求められる。

2.5 エビデンスに基づく治療・診断に向けての課題

さて、対象疾患の感受性遺伝子のマッピングに成功したとする。疾患関連遺伝研究の最終的な目標は、感受性遺伝子を見つけ出すことではなく、研究からの知見を疾患の治療・診断に応用することであろう。遺伝子のマッピングから実際の応用までには、トランスレーション・リサーチと呼ばれる基礎と臨床とをつなぐ研究が必要であることはよく知られている。本節では、基礎研究の立場から、どのような研究が更に必要であるかについて、以下に簡単に紹介しておこう。

通常の科学研究と同様、追試による解析結果の検証は必須である。特に、多因子疾患には、その遺伝的異質性、低浸透率、表現模写という特徴があるため、遺伝子マッピングに用いた集団とは独立の解析集団を用いた統計遺伝学的検証は重要な意味を持つ。この追試は、最初に報告した研究者グループ自身で行うこともあれば、他の研究グループによりなされることもある。このような追試を科学的に有効なものとするためのポイントは、研究対象としたサンプルの由来や表現型（関連解析の場合は、その対照集団の表現型も含む）を正確に報告することであり、また、その報告に合致した表現型をもつサンプルを用いて追試を行うことである。

また、感受性変異と罹患性との因果を検証するには、前向きコホート研究（数万～数十万人規模の健康人集団を追跡調査し、注目する要因と疾病発症との関連を分析する研究）が有効な手段のひとつである。非常に多大の時間、費用がかかる研究ではあるが、感受性変異の相対危険度（Relative risk）や寄与危険度（Attributable risk）も算出することができるため、実際の治療・予防効果を見積もることにもつながる。

感受性遺伝子マッピング研究にも当てはまることであるが、特に、遺伝学的追試や前向きコホート研究は、ひとつの研究室のみで実施できる規模の研究ではないことは留意されたい。そもそも「多因子疾患」を克服するという研究課題自体が非常に大きなものであることを考慮すると、ヒトの遺伝学的研究を進展させるためには、同一、あるいは関連する疾患を研究対象とする複数の研究グループが

有機的に結びつくことが必須であろう。従って、臨床サンプルを用いた研究から感受性変異の同定や検証がなされ、その成果が医療の場に還元される、という理想的な循環を作り出すためには、目的を共有する研究グループ間での共同研究体制の構築が最も大きな課題のひとつである。

2.6 おわりに

多因子疾患の感受性遺伝子をどのようにして同定するかについて、非常に簡単であるが、その流れを解説してきた。詳細については以降の各章を参照されたい。また、本章ではほとんど触れなかったが、多因子疾患では、文字通り、多数の要因が疾患発症に関わることから、要因間の相互作用をも考慮する必要がある。相互作用の検出法（他章参照のこと）のみならず、検出した相互作用をいかにして検証するかが、未解決の課題として残されている。このような課題は、ヒト以外の遺伝学にも共通しており、世界中で様々なアプローチを用いて精力的に取り組まれていることから、いずれ解決することが期待される。相互作用を考慮することで、近い将来、新たな感受性遺伝子マッピング法が作出されるかもしれない。しかしながら、疾患遺伝子解析に DNA サンプルが必要であることは不変である。また、ヒトゲノム中の DNA 変異に対するより深い理解も必須であろう。そして、感受性遺伝子を同定するには、ひとつの学問分野を深化するだけでは十分ではなく、「総合力」を高めることが重要であることを付記して、本稿の締めとしたい。

3. SNP による集団遺伝学的研究

3.1 はじめに

「SNP 解析入門」を手にとられた方のほとんどは、ヒト疾患関連研究への SNP データの応用や、データ解析の実践面にご興味があるのであろうと思われる。本章では、「集団遺伝学」をキーワードとして、SNP などの DNA 多型の生物学的な意味や、ヒト SNP データに対する集団遺伝学的研究からの知見、およびそれらの知見の疾患関連研究における意義などについて紹介する。なお、読みやすいことを優先して、「集団遺伝学」のエッセンスをできるだけ数式を用いずに解説している。数多くの成書が出版されているので（[Cavalli-Sforza et al. 1994](#); [Hartl and Clark 1997](#); [Jobling et al. 2004](#); [Kimura 1983](#); [Klein and Takahata 2002](#); [Nei 1987](#)）、「集団遺伝学」にご興味ある方、あるいはご興味を持たれた方は、そちらの方も併せてご覧いただければと思う。

3.2 集団遺伝学とは

集団遺伝学 (Population genetics) は、現存する生物集団内 (ヒト集団内など) における遺伝子の構成やアレル・遺伝子型頻度の変化などについて数量的に記述することを目的とした遺伝学の一分野であり、主として、生物進化を理解するための学問分野として発展してきた。最近では、分子集団遺伝学 (Molecular population genetics) や進化集団遺伝学 (Evolutionary population genetics) といった名前も用いられているが、その本質は変わらない。生物の進化過程は、換言すれば、生物の辿ってきた歴史のことであり、この歴史性そのものを実験的に再現することはできない。しかしながら、適切な数量モデルが存在すれば、進化過程を推論することは可能である。このような推論を可能とするためのモデル構築が集団遺伝学の目標のひとつであり、従って、集団遺伝学は推測統計学の一種である。

集団遺伝学で最も基本となる量は、集団中の遺伝子頻度 (アレル・遺伝子型頻度) である。ABO 式血液型遺伝子を例として簡単に説明してみよう。ヒトの ABO 式血液型は、1900 年に Landsteiner により発見された血液型で、ヒト集団内において遺伝的多型が観察された最初の遺伝子である。ABO 式血液型のアレルには、主として A、B、O の 3 種類があり、A と B とは互いに優性、O は A、B に対して劣性である。例えば、日本人集団における血液型頻度は、A 型 39%、B 型 22%、O 型 29%、AB 型 10% (およそその比率が A 型 : B 型 : O 型 : AB 型 = 4 : 2 : 3 : 1) であることはご存知のことと思われるが、アレル頻度でみると、A アレル 28%、B アレル 18%、O アレル 54%となる。このアレル頻度はヒト集団間で大きく異なることが知られており、ABO 式血液型遺伝子多型は集団遺伝学的解析の対象として精力的に研究されてきた。このような血液型における多型はタンパク質レベルの個人差として見いだされたものであるが、他のタンパク質多型と同様に、遺伝子をコードする DNA 領域における突然変異により生じたものであり、本質的には DNA レベルでの個人差である。

SNP を中心としたヒト DNA 多型に関する膨大な情報がデータベース化され、かつ、DNA 塩基配列の決定や DNA 多型サイトにおける遺伝子型の決定が比較的容易になされる現在においては、集団遺伝学で取り扱うデータは、DNA レベルでの個人差である場合がほとんどである。他方、ヒトゲノム内の SNP に関するデー

データベース (dbSNP など) が整備され始めたのは高々10年ほど前からであるが、それ以前においても、ヒト集団内の遺伝的多様性に関する量的な理解が十分なされていたことは強調しておきたい。すなわち、ヒトゲノム全域にわたる SNP 情報の取得前後において、ヒト集団を遺伝学的に理解するために必要な数量モデルの根幹部分は大きく変化していない、ということである。集団内の多様性を理解するために必要な概念と共に、これまでに明らかになっている事柄について、以下に概説する。

3.2.1 遺伝的浮動、自然選択、そして「中立的」な突然変異

先に述べたように、ある遺伝子座において観察されるアレルの集団内相対頻度は、集団遺伝学における基本統計量のひとつである。この観点からみると、生物進化は、集団におけるアレル頻度の経時的变化とみなすことが可能である。図 3.1 には説明のための非常に簡単な例を示しているが、ある SNP において2つのアレルが存在し、それらを赤丸、青丸で表すこととする。過去のある世代 (世代数 1) では、赤、青がちょうど5個ずつ、すなわち、各アレル頻度が0.5である場合を考える。その次の世代 (世代数 2) では、赤アレル、青アレルの頻度がそれぞれ0.4、0.6へと変動している。そして、現在 (世代数 T) では、集団内には青アレルが存在せず、観察されるのは赤アレル (頻度 1) のみである。このようなアレル頻度の継代変化を引き起こす主要な要因として、遺伝子頻度の機会的浮動 (遺伝的浮動 [Random genetic drift]) と自然選択 (Natural selection) を挙げる事が出来る。

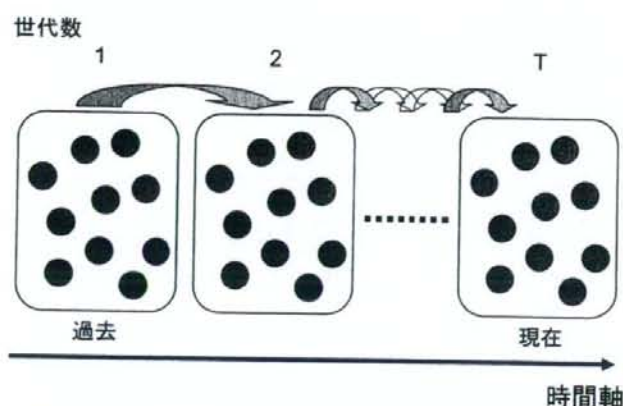


図 3.1 SNP アレルの集団内相対頻度の継代変化

SNP の 2 つのアレルを赤丸、青丸でそれぞれ示している。

遺伝的浮動とは、集団内の個体数が有限であることに起因する現象のことである。すなわち、個体数の有限性により、親世代から子世代へのアレルの伝達（例えば、図 3.1 の世代数 1 から世代数 2 への伝達）は、遺伝子プールと呼ばれる親世代のアレルの集合から、アレルの「ランダムサンプリング（機会的抽出）」により子世代の遺伝子プールを形成することとみなすことができる。このサンプリング過程におけるアレル頻度の確率の変動（遺伝的浮動とよぶ）により、通常、親世代とはわずかにではあるが異なる遺伝子プールをもつ子世代が形成される。一方、あるアレルを持っていると生存上有利、あるいは不利である（相対適応度が高い、あるいは低い）場合、そのアレルに対して自然選択が働くことによっても、アレル頻度の継代変化が引き起こされる。

DNA レベルでの進化機構に関するこれまでの研究から、集団内の分子レベルでの多様性を説明する主要因は、遺伝的浮動と「中立的」な突然変異であると推論されている。「中立的」な突然変異とは、自然選択に対して有利でも不利でもない「中立」（あるいは中立に近い）変異のことである。簡単のために、タンパク質をコードする遺伝子に生じる突然変異を例とする。C. Darwin が 1859 年に著わした「種の起原」では、生物進化の主要因は自然選択にあるという学説が提唱されたが、その自然選択説を分子レベルでの進化に適用すると、個体に生じた遺伝子突然変異の大多数は有害（終止コドンの出現による偽遺伝子化など）であり、生物進化に寄与するのは自然選択に対して有利な突然変異のみである、という予測を

与える（図 3.2）。ところが、実際の観察データからは、自然選択に対して中立的な変異（アミノ酸置換を伴わない DNA 変異など）が多数を占めることが明らかになっている。従って、分子レベルでは、個体の表現型とは関係しない中立突然変異が遺伝的浮動により偶然に集団内に広がる場合がほとんどである、と考えられている（分子進化の中立説とよぶ）。図 3.3 には、新たに出現した中立突然変異の集団（染色体数 50）における頻度が、遺伝的浮動により出現 100 世代後までにどのように変化するかについての 100 回のシミュレーション結果をまとめて示している。例えば、100 世代後を現在とみなすと、新生突然変異アレルの通常運命は集団からの消失であり、非常に少ない場合においてのみ変異アレルが偶然に集団に固定するか、あるいは多型状態で存在することが分かる。

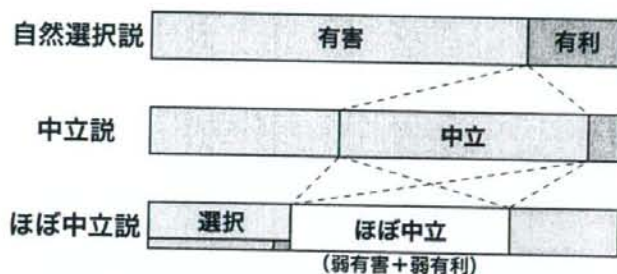


図 3.2 生物進化についての 3 つの学説における新生突然変異についての概念比較図

いずれの説においても、自然選択の影響を強く受ける新生突然変異（有害+有利）の中では、有害なものが多数を占めるという点で共通している。3 者の相違点は、中立、ほぼ中立（弱有害+弱有利）クラスに属する突然変異が全体に対して占める割合にある。

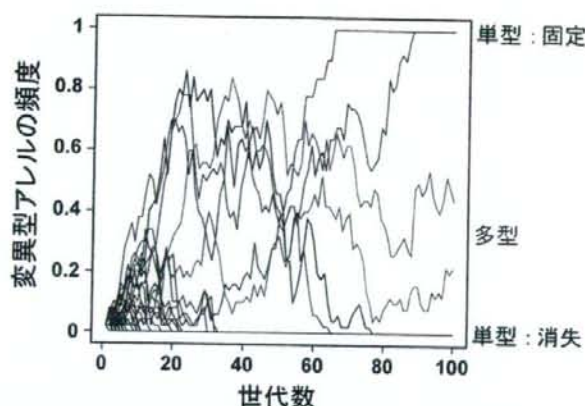


図 3.3 新生中立突然変異の遺伝的浮動によるアレル頻度変化に関するシミュレーション

染色体数 50 (2 倍体ゲノム生物では 25 個体に相当) の集団を仮定し、中立突然変異の初期頻度 $0.02 (= 1/50)$ での変異アレル頻度の経代変化につき、100 世代後までのシミュレーション (100 回試行) 結果を示している。

なお、本章ではあまり触れないが、集団内でのふるまいは中立突然変異と大差ないが、自然選択に対して完全中立ではない「ほぼ中立」と呼ばれるような変異 (自然選択の影響があまり強くない、有害と中立、あるいは有利と中立の中間クラスの突然変異) が比較的多く存在するというモデルも提唱されており (図 3.2; Ohta 2002 など)、集団内の遺伝的多様性の保持機構をより正確に理解するためには、遺伝的浮動と自然選択との相互作用を考慮に入れた、より複雑な進化モデルが必要なのであろう。しかしながら、詳細は後述するが、自然選択の影響を受けている SNP を検出するための帰無仮説として、分子進化の中立モデルが現在では受け入れられている。

3.2.2 集団の有効な大きさ、塩基多様度、そして most recent common ancestor (MRCA)

それでは、DNA レベルで見た場合、ヒト集団はどの程度の多様性を有するのだろうか。多様度を量るための幾つかの指標が提唱されているが、塩基多様度 (Nucleotide diversity: π) が標準統計量として用いられている。塩基多様度とは、遺伝子プール内の任意の 1 対の染色体を比較した時に、2 配列間で異なる塩基の割合の平均値のことである (式 1)。

$$\pi = \sum_{ij}^q d_{ij} x_i x_j \quad (\text{式 1})$$

ここで、 d_{ij} は DNA 配列タイプ i とタイプ j との間で異なる塩基の割合であり、 x_i 、 x_j はそれぞれタイプ i 、タイプ j の集団内頻度である。図 3.4 には、塩基長 20 bp の 3 つの配列タイプが観察された場合を例とし、塩基多様度算出の実際を示しているが、任意交配集団では、 π は塩基配列レベルにおけるヘテロ接合度に相当する。ゲノムワイドな DNA 配列データに基づくヒト集団内の塩基多様度 π の観察値は $\sim 8 \times 10^{-4}$ である（常染色体における観察値であり、性染色体は異なる塩基多様度を有することに留意：The International SNP Map Working Group 2001; Venter et al. 2001）。このことは、任意の 2 つの常染色体を比較した場合、約 1,300 bp 当たり 1 ヶ所という平均頻度で塩基配列の違いが観察されることを意味する。

ex.) 塩基多様度計算の実際

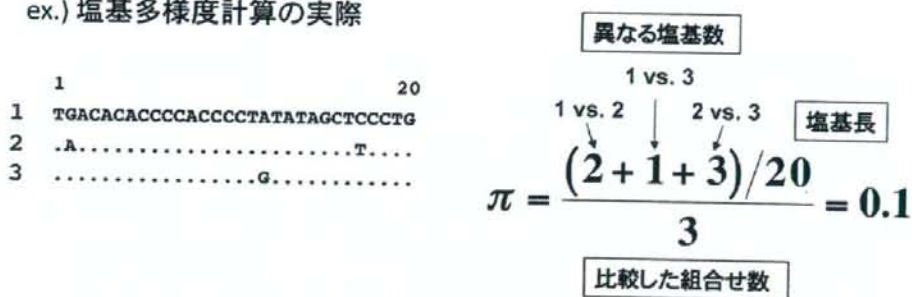


図 3.4 塩基多様度の計算方法

上図は、塩基長 20bp の 3 つのタイプ (1, 2, 3) の塩基配列アラインメントを示している。配列タイプ 2 および 3 においては、配列 1 と一致する塩基サイトはドット (・) で、異なる塩基サイトは観察された塩基で表している。塩基多様度 π は、塩基配列の総当たりの比較から、各比較における異なる塩基の割合の総和 (計算式の分子) を計算し、比較総数 (計算式の分母) で除したものである。

塩基多様度は、下式で表すように、「集団の有効な大きさ N_e 」と「年あたりの塩基突然変異率 μ 」の積と比例関係にあることが知られている。

$$\pi \propto N_e \times \mu \quad (\text{式 2})$$

「集団の有効な大きさ」は、理想集団（全個体の次世代に対する遺伝的寄与率が等しい任意交配集団）では集団内の個体数とほぼ一致するが、通常、例えばヒト集団では全人口（2008年現在約66億人）よりも小さい。ヒトなどの2倍体生物では $\pi = 4N_e\mu$ と記述できることから、集団の有効な大きさは $N_e = \pi/(4\mu)$ より見積もることができる。ヒトの場合、その平均塩基突然変異率はおおよそ 2×10^{-8} （1年あたり1塩基サイトあたり）と推定されており（Nachman and Crowell 2000）、ヒトの進化史を通じての集団の有効な大きさは、 $N_e = 8 \times 10^4 / (4 \times 2 \times 10^{-8}) = 10^4$ 個体程度である。他の推定方法からもヒト集団の有効な大きさは 10^4 オーダーであることが示されており（Klein and Takahata 2002）、ヒトの有効集団サイズは決して大きくない。

さて、本項の最後に「遺伝子系図学」について簡単に触れる。DNAが半保存的複製することから自明なように、現存する世界中のヒトのゲノムは「共通祖先ゲノム」に由来する。時間の流れを現在から過去へと遡って、現在から共通祖先までの道筋をたどったもの、すなわちゲノム間の類縁関係を表したものを系図と呼ぶが、遺伝的多様性情報に基づくと、系図上、現在に最も近い共通祖先（Most recent common ancestor: MRCA）の存在時期が推定可能である。図3.5には、簡単のために、「遺伝子の系図（遺伝的組換えのない場合）」を例として示している。現在の遺伝子プールには黒丸アレルのみが存在するが、それらの全ては、過去のある時点において白丸アレルから突然変異により出現した1つの祖先アレルに由来すると仮定し、アレル頻度の継代変化は遺伝的浮動によるとする（図3.5の例では、系図上のMRCA〔図中の赤黒丸〕と祖先アレルとが異なることに注意）。ヒトゲノム中の異なる遺伝子座に位置する遺伝子はすべて、図3.5と同様の系図を有するが、アレル頻度の確率的変動量が世代ごと、遺伝子座ごとに異なるために、各遺伝子のMRCAの存在時期（Time to MRCA: TMRCA）は同一ではない。その一方で、ヒトなどの2倍体生物の場合、各遺伝子のTMRCAは、平均 $4N_e$ 世代時間前であるという期待値が与えられている。ヒトの世代時間を20年とすると、平均 $TMRCA = 4 \times 10^4 \times 20 = 8 \times 10^5 = 80$ 万年となる。同様の確率論的モデルを用いると、中立突然変異により出現したアレルの「平均年齢」を推定することも可能である（式3: Kimura and Ota 1973）。

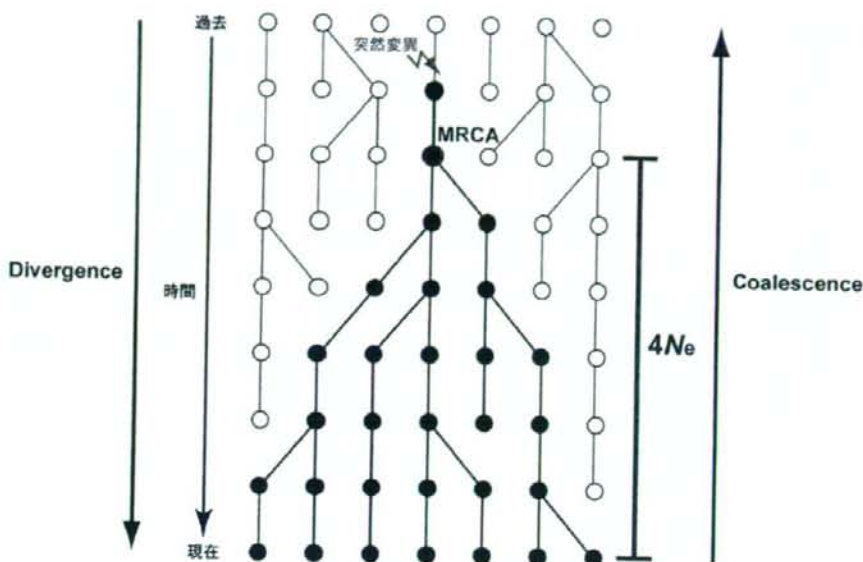


図 3.5 集団内に存在するすべての遺伝子間の進化的関係を表す「遺伝子の系図」

白丸および黒丸はそれぞれ非変異型、変異型遺伝子を表し、過去のある時点において（緑色矢印）、遺伝子突然変異により変異型（黒丸）が出現したとする。遺伝子間を結ぶ線分は遺伝的関係性を表し、現在から過去へと時間を遡った場合（Coalescence 過程）、現在の集団内に存在するすべての遺伝子の遺伝的由来を説明する線分を赤色で強調している。現存する全遺伝子に共通する現在に最も近い共通祖先（Most recent common ancestor: MRCA: 赤黒丸）の存在時期（Time to MRCA）は、平均 $4N_e$ 世代時間前であることが期待される（ N_e は集団の有効な大きさ）。

$$\bar{t}(x) = -4N_e \left(\frac{x}{1-x} \right) \log_e x \quad (\text{式 3})$$

ここで、 x は集団内のアレル頻度であり、平均年齢 \bar{t} は世代時間で測っている（ \bar{t} と世代時間との積が物理的時間である）。ここで注意しておきたいことは、TMRCA やアレル年齢の推定は、仮定しているモデルが必ずしも実際のデータに適合するわけではなく、その上、推定自体が非常に大きな不確実性を伴っていることである。例えば、確率的誤差のために、いずれの推定においても、平均値と同程度か、あるいはそれよりも大きな標準偏差（すなわち N_e 世代時間オーダーの標準偏差）が与えられる。しかしながら、この系図という概念は、後述するような疾患関連変異の遺伝を考察する上で有効なものである。

3.3 疾患関連研究としての集団遺伝学の意義

前節で概説したように、集団遺伝学はヒト集団を遺伝学的に理解するためには必要な学問領域である。その研究のなかから、例えば、全ての現代人は今から約15-20万年前にアフリカ大陸に住んでいた祖先集団に由来する (Horai et al. 1995 など)、といったようなことまで推論可能になってきた。本節では、ヒト集団を遺伝学的に理解することの意義について、特に、疾患感受性遺伝子の同定などのゲノム医学研究に対して集団遺伝学がどのように貢献し得るかについて簡単に紹介する。

メンデル型の遺伝様式を示す単一遺伝病は、その原因となるひとつの疾病遺伝子 (突然変異などによる変異型遺伝子のこと) を両親より受け継ぐことにより発症する。例えば、常染色体劣性遺伝の場合は両親のいずれから、常染色体優性では少なくとも両親のどちらかから、疾病遺伝子を受け継ぐことが発症要因となる。このようなメンデル遺伝性疾患の場合は、家系サンプルを収集し、それらのDNAを遺伝学的に精査することにより、家系内発症の要因となっている創始者突然変異 (Founder mutation) を同定することが、現在では比較的容易になってきている。すなわち、図 3.5 に示したような「疾患遺伝子の系図」を研究成果として得ることが十分可能となってきているということである。その一方、メンデル型の遺伝様式を示さない、例えば多因子疾患 (疾患発症に関与する遺伝因子が複数存在する疾患) においては、その遺伝的異質性、低浸透率 (前章などを参照のこと) などの要因から、「疾患感受性遺伝子の系図」を描くことは容易ではない。しかしながら、前節で述べたゲノムの同祖性から明らかのように、複数の疾患発症者の間で、発症に関わる感受性変異が少なくともひとつは共有されているとする仮定は十分成立する。また、前述したヒトの平均塩基突然変異率 (1年あたり1塩基サイトあたり 2×10^{-8}) を考慮に入れると、ヒトの進化史 (約15-20万年間) において、全く同一の疾患感受性アレル (SNPアレル) が独立に複数のゲノムで生じることが非常に低い確率でしかあり得ないことから、患者間で共有している感受性アレルのほとんどは同祖であるとみなしてよい (図 3.6)。この疾患感受性アレルの同祖性という概念に基づいて、患者-対照関連解析では、両群における共有の度合いの違いを統計学的に分析していると考えることが出来る。なお、ありふれ

た疾患の発症を説明するための仮説として、Common Disease-Common Variant (CD-CV)仮説（集団頻度の比較的高いアレルが疾患発症に関わるとする仮説：Lander 1996; Reich and Lander 2001）や Common Disease-Rare Variant (CD-RV)仮説（集団頻度の低いアレルをより重視した仮説：Liu et al. 2005 など）が提唱され、これらの仮説に基づいて患者-対照関連解析が精力的に行われているが、患者群、対照（健常者）群いずれにも観察されるアレルのなかで、患者群での保有率が有意に高いものが疾患発症に関わるといふ、感受性遺伝子同定研究の背景には、アレルの同祖性に関する集団遺伝学的解析からの成果が潜んでいるのである。

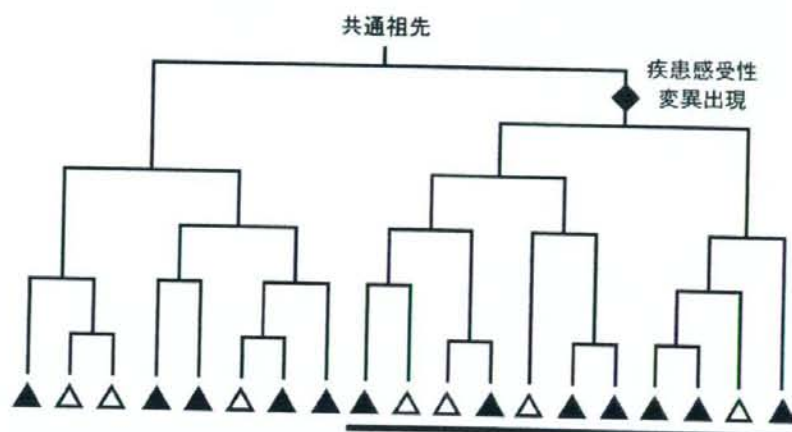


図 3.6 多因子疾患感受性遺伝子の系図

黒三角および白三角はそれぞれ疾患罹患患者、非罹患患者を表し（簡単のために、1 倍体ゲノムとして表している）、疾患発症に関わる感受性遺伝子の系譜を模式的に示している。赤色で強調している罹患患者および非罹患患者（系図の右側クラスター）は、発症リスクを高める疾患感受性変異をひとつ共有しているが、浸透率が低いなどのために、感受性変異（赤色）を共有するすべての人が疾患を発症するわけではない。

では、どのような SNP アレルが疾患発症に関わると考えられるのであろうか。Bustamante らは、約 11,000 遺伝子に対するヒト集団内 DNA 多型解析およびヒト-チンパンジー比較ゲノム解析を行い、ヒト集団内には、アミノ酸置換を伴う SNP が過剰に存在していることを報告している（Bustamante et al. 2005）。加えて、進化的に負の自然選択の影響を受けている遺伝子のなかに、メンデル遺伝性疾患に関わる遺伝子が比較的多く含まれることが示され、負の弱い自然選択を受けている遺伝子が多因子疾患の感受性遺伝子候補になり得るといふ考察がなされている。

その一方、生物としてのヒトを特徴づけることに関わる遺伝子には進化的に正の自然選択が働いている場合が多く（例えば、FOXP2 遺伝子と口頭伝達能：Lai et al. 2001; Enard et al. 2002 など）、そのような遺伝子に生じた変異や、実際に正の自然選択を受けている変異は疾患発症の要因となり得ることもよく知られている（Helgason et al. 2007 など）。表 1 には、ヒトの疾患や形質に関わると考えられている遺伝子のうち、主として集団遺伝学的解析から自然選択が働いていることが示唆されている代表的なものを示している。現在では、ゲノム全域での体系的アソシエーション・スタディ（Genome-wide association study; GWAS）により疾患感受性変異を見いだすというアプローチが広く用いられており、GWAS により探査する SNP や遺伝子に働く自然選択についての先験的な知識は、原則的には必要ない。しかしながら、GWAS などにより見いだされた疾患関連遺伝領域から真の感受性 SNP を同定したり、感受性 SNP の疾患への関わり方を明らかにしたりする上で、ヒト集団内で観察される SNP に対する進化集団遺伝学的解析結果は示唆的な役割を果たすことが期待される。というのは、分子進化の中立モデルから逸脱した SNP や遺伝子は、何らかの病態生理的な機能に関わる可能性を秘めていると考えられるからである。次節では、分子進化の中立性から逸脱した SNP を見いだすための研究手法や、その現状を簡単に紹介する。

表 1 自然選択の痕跡が観察された代表的な遺伝子、およびそれらの遺伝子が関与するヒト疾患・形質

遺伝子名 (シンボル)	関連する疾患・形質	参考文献
	疾患	
アンジオテンシノーゲン (AGT)	高血圧	Nakajima et al. (2004)
インターロイキン 1A (IL1A)	喘息	Akey et al. (2004)
インターロイキン 4 (IL4)	喘息	Rockman et al. (2003); Sakagami et al. (2004)
インターロイキン 13 (IL13)	喘息	Zhou et al. (2004)
カルパイン 10 (CAPN10)	2型糖尿病	Fullerton et al. (2002)
チトクローム P450 3A サブファミリー (CYP3A)	高血圧	Thompson et al. (2004)
マトリックスメタプロテアーゼ 3 (MMP3)	冠動脈心疾患	Rockman et al. (2004)
	形質	
ABCC11	耳あか型	Yoshiura et al. (2006)
EDAR	毛髪 (太さ)	Fujimoto et al. (2008)
SLC24A5	体色 (色素沈着)	Lamason et al. (2005)
ダッフィ血液型 (FY)	マラリア抵抗性	Hamblin et al. (2002)
ラクターゼ (LCT)	ラクトース (乳糖) 耐性	Bersaglieri et al. (2004)

3.4 全ゲノムSNPスキャンからの疾患感受性候補遺伝子の抽出

前節までに述べてきたように、ヒト集団内の遺伝的多様性を説明する主なものは「中立的」な SNP や遺伝子であるが、正または負の自然選択の影響下にある SNP・遺伝子の存在も報告されている。そして、自然選択の影響により分子進化の中立性からずれた SNP や遺伝子が、病態生理学的な役割を担う可能性があることから、現在では、SNP や遺伝子の「集団遺伝学的な変わりもの（外れ値）探し」が精力的に行われている。「変わりもの」探しのために様々な手法が用いられているが、それらに共通しているのは、ヒト集団内の遺伝的多様性をいろいろな視点から数量化し、その経験値分布に基づき統計学的に見いだした「外れ値」は、分子進化の中立モデル（すべての SNP が進化的に中立であることを仮定するモデル）から逸脱している可能性が高い、とする概念である（図 3.7）。

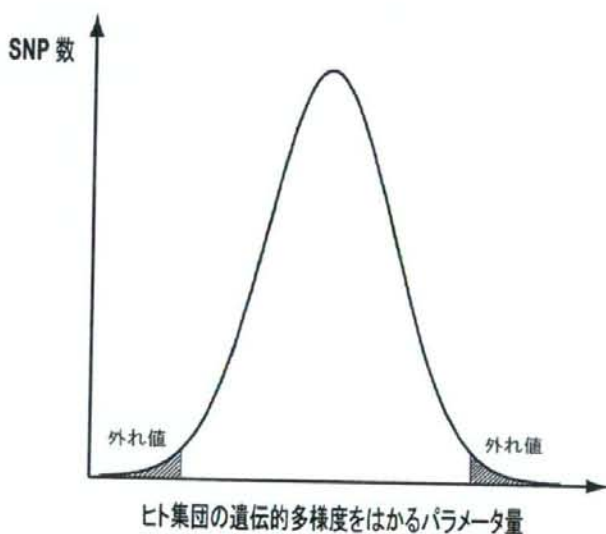


図 3.7 ヒトゲノム中に存在する「外れ値」SNP の同定方法

ヒトゲノムを網羅する SNP のうち、(1) ヒト集団全体、あるいは複数の地理的集団でのアレル・遺伝子型頻度が調べられており、(2) 各 SNP における個々人の遺伝子型などの情報が入手可能な SNP が研究対象とされる。ヒト集団内の遺伝的多様性の程度につき、SNP ごと、あるいは遺伝子領域ごとに、種々の角度から数量化し、それらの経験値分布に基づき統計学的な「外れ値」SNP や「外れ値」遺伝子を見いだす。

このような研究が進展した大きな理由のひとつは、ヒトゲノム SNP やハプロタ

イブ構造などに関するデータベースの充実を挙げることが出来る。国際 HapMap プロジェクトからのヒトゲノム・ハプロタイプ地図(<http://www.hapmap.org/>)以外にも、以下に記すような各種データベースが解析に用いられている：dbSNP (<http://www.ncbi.nlm.nih.gov/SNP/>)、Perlegen (<http://www.perlegen.com/>)、SeattleSNPs (<http://pga.gs.washington.edu/>)、Celera (<http://www.celera.com/>)、ALFRED (<http://alfred.med.yale.edu/>)。別の言葉で説明を加えるならば、従来のように分子進化の中立モデルを仮定し、モデルからの逸脱度を統計学的に見積もる方法では、遺伝的浮動に伴う確率的誤差などが「外れ値」SNP の検出力を押し下げているのに対して、ゲノム網羅性の高い SNP データなどが利用可能となったことにより、実際の大量観察データに基づくモデルフリーの解析が可能となり、「外れ値」SNP 候補を同定することが容易になってきた、ということである。多くの遺伝学関連雑誌で、ヒトゲノムでの「外れ値」SNP 探しについての総説が出版され(例えば、Biswas and Akey 2006; Nielsen et al. 2007 など)、特集も組まれていることから(例えば、2006 年の Nature Reviews Genetics 10 月号 [Excoffier and Heckel 2006; Marjoram and Tavaré 2006] など)、興味ある方はそちらも合わせてご覧いただきたい。なお、自然選択の影響を受けてきた遺伝子を探索するためのアプローチのひとつである比較ゲノム解析、特にヒトにとって「進化の隣人」とよばれるチンパンジーとの比較ゲノム解析からの成果については、他の総説などをご参照されたい (Biswas et al. 2006; Kehrer-Sawatzki and Cooper 2007; Nielsen et al. 2007 など)。

「外れ値」SNP や「外れ値」遺伝子探しという観点からみて、ヒト集団内 SNP に関する大量データに基づく解析も万能ではない。特に、負の自然選択の影響下にある SNP や遺伝子を見いだすことは十分に出来ていないのが現状である(負の自然選択は純化選択[Purifying selection]ともよばれ、有害な変異を集団から取り除くための選択である)。この理由としては、元来、機能的に重要な遺伝領域(例えば、遺伝子座)の大多数には進化的に負の自然選択が働いていることが知られており、その中から注目すべき「外れ値」を統計学的に抽出することが困難であるということが挙げられるであろう。しかしながら、前述したように、そのような選択が働いている SNP や遺伝子は、単一あるいは多遺伝性疾患の要因候補となり得ると考えられており、これらの「外れ値」も研究対象とする必要がある。最近、負の自然選択を検出することも可能な Tajima's D とよばれる統計量 (Tajima