

染色体と遺伝子について

Point

- ✓ 核ゲノムは約 31 億塩基対あり、細胞核内で 23 対 46 本の染色体を形成する。
- ✓ ヒトの遺伝子は約 25,000 種類あり、からだを作る設計図の役割を担っている。
- ✓ 疾患の責任遺伝子と症状との関連を検討するために様々な解析技術が発展してきた。
- ✓ 遺伝子変異には様々な変化の種類があるが、同一遺伝子でも変化の部位や種類の違いで表現型には大きな差が出る。

基礎的な知識

ヒトゲノム

「ゲノム (genome)」とはヒト DNA の全配列のことで、「遺伝子 (gene)」と「接尾語 “-ome”: すべて (all) を表すギリシャ語」を用いた造語である。21 世紀初頭のヒトゲノム計画により、ヒトゲノムの全容が解明された。これ以前のゲノム研究では、遺伝子の一つひとつ調べる必要があったが、ヒトゲノム計画終了後、また、次世代シーケンサーの開発を契機としてゲノム解析が高速化し、ヒトゲノム全体を解析することが比較的容易になった。それにより、集団における遺伝子の多様性やそれと病気との関連性についての研究が幅広く行われるようになった。

ゲノムを構成する DNA (deoxyribonucleic acid: デオキシリボ核酸) は、A (アデニン)・T (チミン)・C (シトシン)・G (グアニン) という 4 種の塩基がヌクレオチドに結合して、ポリヌクレオチド鎖を形成してできている。ヒトゲノムではこのような 2 本のポリヌクレオチド鎖が A-T, C-G で対になって結合し、二重らせん構造となっている。そしてこの塩基の配列が遺伝子を決定しているため、DNA は遺伝情報のもとということになる。

DNA はむき出しで存在しているわけではなく、糸巻のようにヒストン八量体 (タンパク質の 1 つ) に巻きつき、それが数珠状に連なった状態になっている。細胞分裂の時以外は核全体に広がっているが、細胞分裂時はそれが凝縮し、「染色体」となり、顕微鏡で観察可能な状態になる (☞ [付録 A-1] 参照)。

ヒトの染色体は 23 対 46 本あり、対のうちの一方は父由来で、他方は母由来の染色体となる。染色体全体で約 31 億塩基対の遺伝情報が存在しており、1 本の染色体には平均で約 1,000 個の遺伝子があり、ヒト遺伝子の総数は約 25,000 個である。これらの染色体全てに含まれる遺伝子や DNA をまとめて核ゲノムと呼び、細胞質内のミトコンド

リアゲノムとは区別される。

また、DNA 上にタンパク質を規定する遺伝子があるが、DNA が転写されて RNA (ribonucleic acid: リボ核酸) が合成される。RNA では T が U (ウラシル) に代わり、構造も一本鎖になる。そして、RNA の塩基配列により、合成するタンパク質のアミノ酸配列が決定される (☞ [付録 A-2] 参照)。

遺伝子

多くの遺伝子が 1 つ以上の異なる産物を作り出すことが可能で、約 25,000 種類のヒト遺伝子から数十万種類のタンパク質が作られる。また、個々のタンパク質は 1 つではなく複数が関連し合うことで、多彩な機能を発揮している。

大部分の遺伝子にはタンパク質の設計図にあたる領域 (エクソン) と、タンパク質の合成を管理・制御する領域 (イントロン) とからなる。これにより、“いつ” “どこで” “どんな” タンパク質を合成するかがコントロールされる。全 DNA の中で遺伝子に関連する領域はごく一部であり、残りの多くはタンパク質の産生には関与しない。実際、タンパク質の設計図であるエクソンは全ゲノム DNA 中の 2% であり、その他の調節因子など働きを持つ領域 3% と合わせても 5% 程度とされる。その他の領域は、ヘテロクロマチン 7%、反復配列 45%、機能が十分に解明されていない領域 43% となっている。

遺伝子から mRNA (コード情報を保持する RNA) が転写される際には、遺伝子の上流 (5' 末端) に存在するプロモーター領域が転写の適切な開始の役割を担っている。プロモーターによって、様々な組織や細胞において特定の遺伝子の発現パターンが調節されている。

染色体

ヒト体細胞の染色体が 46 本であることがわかったのは 1956 年である。国際ヒト染色体名規約 (International System for Human Cytogenetic Nomenclature: ISCN) では、1~

表 1-1 染色体の群分け

群	特色
A 群 (1~3 番)	大型の染色体で、セントロメアが中心部にあるため短腕と長腕の長さがほぼ同等。2 番は短腕の方が短い。
B 群 (4~5 番)	大型の染色体でセントロメアは末端近くにある。4・5 番は非常によく似ており、区別が困難。
C 群 (6~12 番, X)	中型の染色体でセントロメアはやや末端部寄り。いずれも形や長さが似ているため、識別が最も難しい群。
D 群 (13~15 番)	中型の染色体で、末端部にセントロメアがあり、短腕にサテライトがあるが見かけ上は長腕のみに見える。
E 群 (16~18 番)	やや小型の染色体で、セントロメアはほぼ中央~端部寄り。
F 群 (19~20 番)	小型の染色体で、セントロメアはほぼ中央にある。両者は区別しづらい。
G 群 (21~22 番, Y)	最も小さな染色体でセントロメアは末端部にあり、Y 染色体以外は短腕にサテライトがあるが、見かけ上は長腕のみに見える。後に 21 番の方が 22 番より小さいことが判明した。

22 番の染色体を“常染色体”，X と Y を“性染色体”と分類した。染色体は顕微鏡下で目視すると、動原体（セントロメア）がくびれている。そしてセントロメアを中心に短い方を短腕（p），長い方を長腕（q）という。

染色体を G 分染法にて染色すると、濃淡で染め分けられ、縞模様の特徴的なパターンを持つ。このパターンは DNA 配列の特徴とほぼ相関しており、これを指標にいくつかの領域に分類している。動原体から腕の末端方向に向かって順番に 1, 2, ……と領域の番号や、各領域のバンドの番号が住所のようにつけられている。

染色体は長さや動原体の位置などを基準に表 1-1 のように群分け（A~G 群）されている。

遺伝子の異常と疾患

遺伝性疾患は表 1-2 に示すように多くの種類がある。例えば、ある 1 つの遺伝子の異常（単一遺伝子異常）で起こる疾患については、その遺伝子（責任遺伝子）を調べれば変異の有無がわかるため、発症予測につながる。このような責任遺伝子は、患者とその家族のゲノムの違いを解析することで明らかにされてきた。

これ以外にも、多数の因子（遺伝子や環境要因）が重なり合うことで発症する多因子遺伝性疾患もある。これらについては、ある疾患を発症した複数の患者（血縁でない者）

表 1-2 遺伝性疾患の分類

単一遺伝子疾患	単一遺伝子の異常による疾患。別名、メンデル遺伝病。この遺伝形質は常染色体か性染色体（X や Y）上の座位によって決定されている。基本的には以下の 5 種類に分類される； ・常染色体優性遺伝（AD） ・常染色体劣性遺伝（AR） ・X 連鎖優性遺伝（XLD） ・X 連鎖劣性遺伝（XLR） ・Y 連鎖（YL）
多因子遺伝性疾患	多数の遺伝子と環境要因の相互作用によって発症する疾患。 ・生活習慣病など
染色体異常症	染色体の数的・構造的異常。正常細胞と異常細胞が混在するモザイクタイプもある。
微細欠失・重複症候群	顕微鏡下で検出困難なレベルの微細な染色体の欠失や重複で、隣接する複数の遺伝子の量的不均衡により生じる。また、近接する類似した DNA 反復配列によって惹起されるものをゲノム病と呼ぶ。
ゲノム刷り込み関連疾患	遺伝子の DNA 塩基配列の変化によらず、メチル化などの修復によって遺伝子の発現が異なる現象をゲノム刷り込み（インプリンティング）現象という。インプリンティング遺伝子はいずれの親由来かによって発現の有無が異なるため、メンデル遺伝形式に従わない。
ミトコンドリア遺伝病	ミトコンドリア DNA 異常に起因する。ミトコンドリア DNA 異常だけでなく、各染色体遺伝子異常によるものもある。

と患者ではないヒトの遺伝子を比較検討することで、関連遺伝子が解明されてきた（ゲノムワイド関連解析）。

さらに、遺伝子に変異があっても、疾患を発症しないこともある。そこで、実際の細胞で発現される RNA やタンパク質を解析する技術も開発され、それぞれトランスクリプトーム解析（RNA）、プロテオーム解析（タンパク質）という。

そして、ゲノム解析が進むにつれ、ヒトそれぞれでゲノム DNA の塩基配列に差があることもわかるようになった。例えば、抗腫瘍薬の効きやすさや副作用の出方などは関連するゲノム DNA の中の 1 塩基の変化によって起こることがある。これを SNP（single nucleotide polymorphism: 一塩基多型）と呼び、個人レベルのゲノムの違いは遺伝子多型といわれる。

変異の種類とその影響

遺伝子変異の種類にもいくつかあり、まず、ミスセンス変異がある。これは DNA 配列中の 1 つのヌクレオチドが

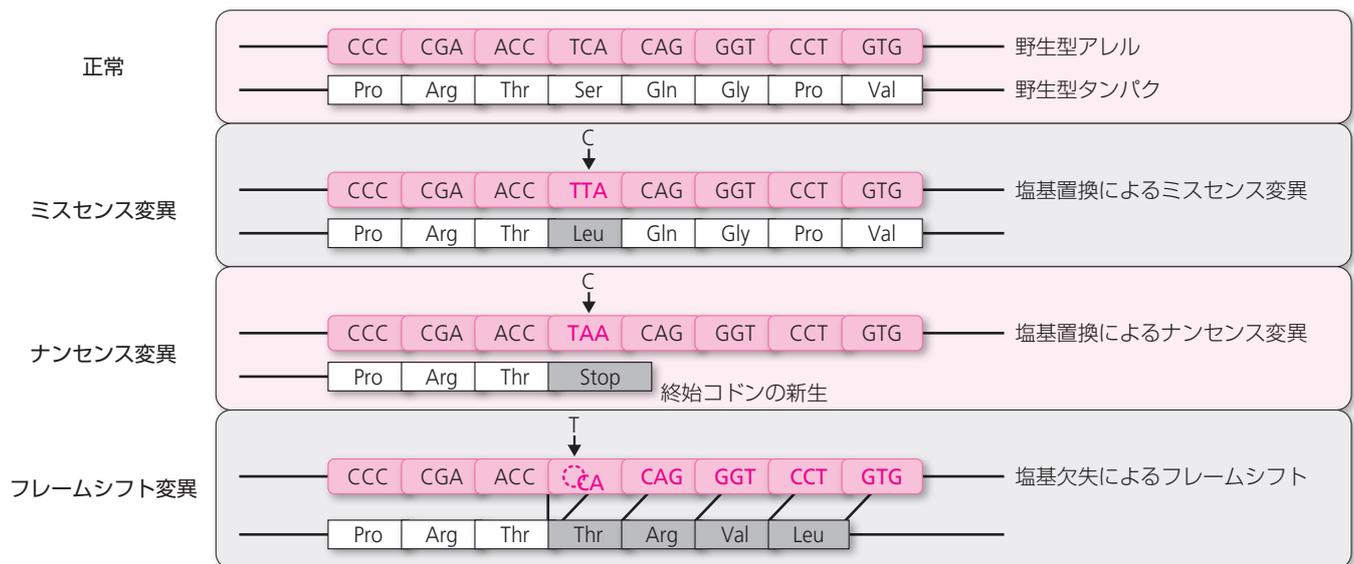


図 1-1 遺伝子変異の種類

別のものに置き換わり、アミノ酸を規定する3塩基からなるコドンが変化することで、遺伝子産物であるアミノ酸の1つが別のアミノ酸に変わる現象である。これにより、正常とは異なるアミノ酸が産生されることで遺伝子のコード鎖の意味が変わるため、ミスセンス変異と呼ばれる(図1-1)。

また、ナンセンス変異とは点変異が起こることで、正常なアミノ酸をコードしていたコドンを終始コドンに変換する現象である。これにより、mRNA自体が分解され翻訳が進まなかったり、タンパク質が生成されても短縮されるた

め不安定になり細胞内で分解されたりといった現象が起こる。特に遺伝子の前半でナンセンス変異が起こった場合に影響が出やすい。

少数の塩基の挿入もしくは欠失した塩基が3の倍数でなければコドンの読み枠が変わり、すぐ下流に終始コドンができやすくなるが、これをフレームシフト変異と呼ぶ。一方で、3の倍数で起こった塩基の挿入・欠失をインフレーム変異と呼ぶ。そのほか、より大きな領域の遺伝子の欠失・逆位・融合・重複も考えられる。

〈関沢明彦，廣瀬達子〉

Whole exome sequencing/Whole genome sequencing

C O L U M N

ゲノムの塩基配列決定法として1977年にサンガー法が登場した。サンガー法は目的とする領域をPCRで増幅させ、その領域を読んでいくもので、現在でも標準的な塩基解読法である。現行では約1,000塩基程度まで一度に読むことができるが、1反応につき1つのDNA断片しか解読できない。一方、2005年に登場した次世代シーケンサー（第2世代シーケンサー）は、従来のサンガー法とは異なり、大量のDNA断片を約100~300塩基程度の短い配列で並行して読んでいく（massive parallel sequencing: 大量並行シーケンシング）。この技術革新により大量の塩基配列データを出力することが可能となり、網羅的なゲノム解析が可能となった。

Whole exome sequencing と Whole genome sequencing

次世代シーケンサーによる塩基解読法として、ゲノム全体を解読する“全ゲノムシーケンス（whole genome sequencing: WGS）”と、タンパク質をコードするエクソン領域だけを解読する“全エクソームシーケンス（whole exome sequencing: WES）”（単に“エクソーム”とも呼ばれる）がある。ヒトゲノムのうち、タンパク質をコードしているのは2%未満であり、その領域にヒトメンデル遺伝性疾患の原因となる変異の85%が存在するといわれている。そのため、エクソン領域を中心に解読を行うエクソーム解析は、メンデル遺伝性疾患の原因となる変異を同定するのに効率のよい方法である。

次世代シーケンサーを用いた解析の工程

実際の解析の流れを図1に示す。初めにゲノムDNAを目的に応じて200~600塩基の長さランダムに断片化（超音波法、制限酵素法などがある）する。断片化されたゲノム断片の両端を修復後、アダプター配列などを付加する。エクソームを行う場合は、エクソン領域の配列を選択的に抽出するため、解読したい領域に相補的なオリゴDNA（ベイトと呼ばれる）とハイブリダイゼーションさせることで目的の配列だけを抽出する（ゲノム分画、キャプチャー）工程が加わる。次に、それぞれのゲノム断片の検出シグナルを強くするために、そのDNA断片を基盤上で増幅させ（Illumina社のGAllx/HiSeqなどでは、フローセルと呼ばれるガラスの基盤上でbridge PCRにより同じDNA配列を増幅させクラスターを形成させる）、それを次世代シーケンサーで解読していく。出力された配列情報（例: fastq など）を、基準となる参照配列にマッピングさせることで、対象の塩基配列を解読していく（図1, 2）。

出力されたバリエーション（参照配列とは異なる塩基配列）の中から、病気の原因となる変異を同定するには、解析する対象疾患の頻度や遺伝形式を考慮した解釈が必要である。例えば、常染色体優性遺伝で、新生変異（*de novo*）が原因で起こる疾患の場合、患者と両親を同時に解析させるトリオ解析が有効である。通常エクソームで検出される *de novo* の変異は1トリオあたり0~3個といわれ

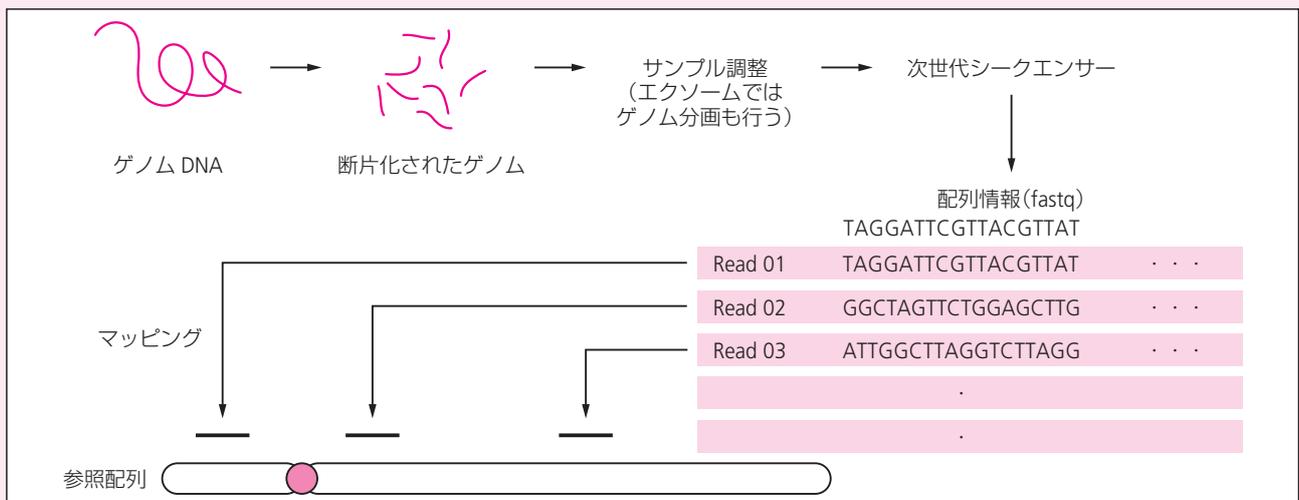


図1 全エクソーム解析の流れ

ゲノムDNAを断片化した後、サンプル調整を行い、次世代シーケンサーで塩基解読を行う。出力されたリード（塩基配列情報）を参照配列に当てていく（マッピング）。参照配列と異なる配列部分を、バリエーションとしてコールさせる。

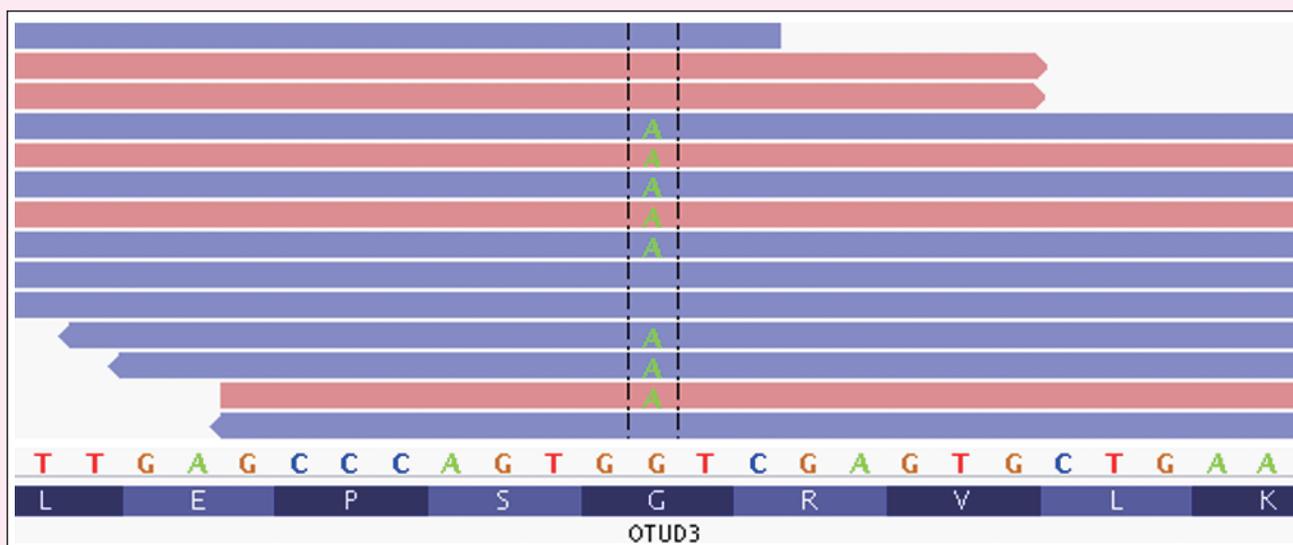


図2 次世代シーケンサーのリードの見え方の例

リードを可視化できるツールの一つ Integrative Genomics Viewer (IGV) で実際の結果を一部示した。上段にリード，下段に参照配列とそれに相当するアミノ酸配列が表示されている。参照配列に対し，順行性にマップされたリードが青色で，逆行性にマップされたリードが赤色で表示されている。例として，*OTUD3* 遺伝子上にあるヘテロ接合性の一塩基置換を (G>A) を示した。リファレンスと異なる塩基の箇所だけに，置換後の塩基 A が表示されている。

ている。 *de novo* のバリエーションがすべて病気を起こすわけではないので，それぞれのバリエーションに関して適切に解釈する必要がある。その際，コントロール群におけるバリエーションの頻度情報 (gnomAD, ExAC, EVS, 1000 Genome, HGVD) や，病原性を予測するソフト (SIFT, PolyPhen-2, MutationTaster) などが用いられている。

今後の課題と展望

現在の主流であるエクソーム解析による変異同定率は，(解析する対象疾患にもよるが) 概ね 30~40%といわれている。原因となる変異を同定できない理由として，エクソン領域以外の塩基配列解読はできない，エクソン領域であってもキャプチャーされない領域があることなどが挙げられる。この点に関しては，WGSにより解決できる部分がある。また，第2世代のシーケンサーは short read と呼ばれ，短い配列を参照配列に当てていく読み方が主流であり，繰り返し配列や，複雑な構造異常，未知の配列などの解読は難しい。近年，1分子シーケンサーとも呼ばれる第3世代シーケンサーが登場した。これは1分子のDNAを連続して数十 kb 程度解読

できる方法である。この方法により，第2世代シーケンサーでは解読が困難であった繰り返し配列や，染色体破砕 (chromothripsis) と呼ばれる複雑なゲノムの構造異常等を検出することが期待されている。また，上記以外の原因となる変異を同定できない理由として，想定した遺伝形式が間違っている場合などがあり，罹患の有無，臨床症状の評価，正確な家系図の把握は極めて重要である。

参考 Web サイト

gnomAD: <http://gnomad.broadinstitute.org/>
 ExAC: <http://exac.broadinstitute.org/>
 EVS: <http://evs.gs.washington.edu/EVS/>
 1000 Genome: <http://www.internationalgenome.org/>
 HGVD: <http://www.hgvd.genome.med.kyoto-u.ac.jp/>
 SIFT: <http://sift.jcvi.org/>
 PolyPhen-2: <http://genetics.bwh.harvard.edu/pph2/>
 MutationTaster: <http://www.mutationtaster.org/>

〈三宅紀子〉