

# 1 ゲノム時代の分子病態学

## ◆まとめ

1. ヒトゲノムはハプロイドあたり約 30 億塩基対からなり、約 2 万種類のタンパク質をコードする遺伝子を含む。
2. ゲノム・エピゲノムの遺伝的多型はさまざまな疾患の発症素因に関わると考えられる。
3. ゲノム・エピゲノムの後天的変異はがんの本質的な発症機構であり、その解明と新たな治療薬開発が加速している。
4. 疾患に関連するゲノム・エピゲノム情報の解明によって疾患分類自体が大きな変革を遂げるとともに、臨床の場におけるゲノム・エピゲノム分子診断が重要となる。

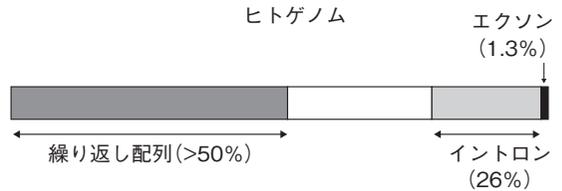


図 1-1 ヒトゲノムの構造

ヒトゲノムはハプロイドあたり約 30 億塩基対からなり、繰り返し配列が過半数を占める。産生されるタンパクのアミノ酸配列を規定するエクソン領域は全体のわずか 1.3% 程に過ぎない。

いる。

たとえば近年、細胞内タンパク質を制御する機能ユニットとしての短い RNA に着目が集まり、約 22-26 塩基の microRNA (miRNA) がさまざまな PCG mRNA の翻訳制御をしていること、また miRNA 自体が現在も進化・拡大している機能ユニットであることが明らかになった。たとえば miRBase (<http://www.mirbase.org/index.shtml>) の Release 19 では計 2,042 種類の miRNA がヒトで発現していることが報告されている。また miRNA 以外にも piRNA などのさまざまな機能性 short noncoding RNA が同定されている。

さらにこうした低分子 RNA のみならず、long non-coding RNA (lncRNA) も機能的意義がありそうである。たとえば lncRNA が染色体上の特定の部位にクロマチン制御因子をリクルートしたり、染色体の高次構造を制御していることがわかってきた<sup>3)</sup>。さらに驚くべきことにヒトゲノム全体の約 75% が何らかの形で転写されていることが明らかになったのである<sup>2)</sup>。こうして、我々のゲノムは PCG 以外にもさまざまな機能ユニットが次々と存在することが明らかになっており、タンパク質を中心とした遺伝情報の考え方はおそらく近未来に大きな変革を迎えるであろう。しかし本稿では、現在の疾患ゲノム解析の主体である PCG を中心として、ゲノム時代の病態理解の変化について議論する。

## A. ヒトゲノムの俯瞰図

ヒトゲノムは一体細胞あたり 1 番染色体から 22 番染色体までが 2 ペアおよび性染色体が 2 本 (男性は XY, 女性は YY) の計 46 本の染色体からなる。ハプロイドあたりの総塩基数は約 30 億 (3 Gbp) に及ぶが、その過半数は SINE, LINE などの繰り返し配列からなり、これらの機能的意義はまだ明らかになっていない。実際にタンパク質をコードするエクソン領域はゲノム中のわずか 1.3% に過ぎず、イントロン領域を加えてもタンパク質をコードする遺伝子 protein-coding gene (PCG) がゲノム全体で占める割合は 1/4 程度である<sup>1)</sup> (図 1-1)。PCG の総数はデータベースによって異なっており、最新の ENCODE (Encyclopedia of DNA Elements) プロジェクトではヒトの PCG 総数は 20,687 と予想されている<sup>2)</sup>。

かつては「遺伝子」という概念は「ゲノム中のタンパク質をコードする単位」として定義されていた。すなわち 1 種類のタンパク質をコードするエクソン+イントロン領域と、それに隣接する転写調節領域 (プロモーター, エンハンサー) などを含めて遺伝子と呼ばれていた。つまりゲノム中の遺伝情報の担い手はあくまでタンパク質を作るユニットと考えられていたのである。しかしこのような遺伝子の理解は再考を迫られて

## B. ゲノム変異と疾患病態

ゲノム・エピゲノム変化 (異常) と疾患の関わりを

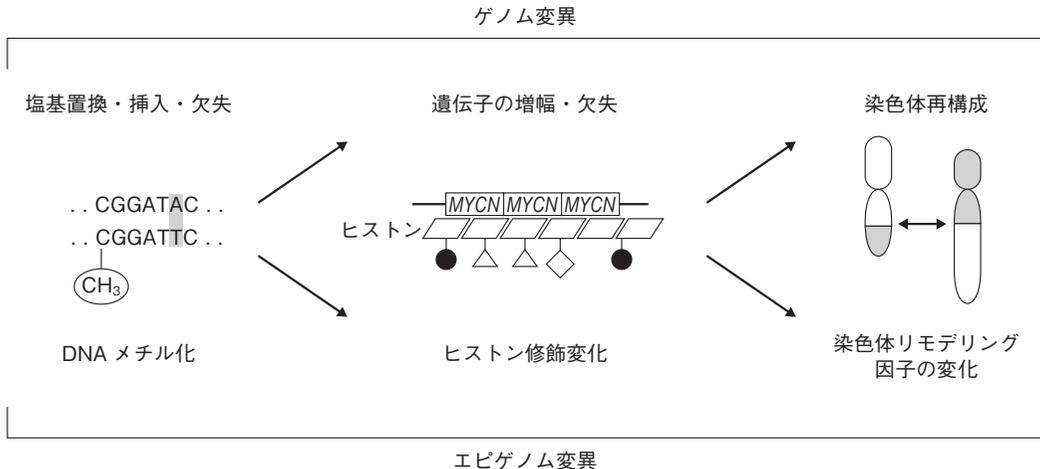


図 1-2 ゲノム・エピゲノム変異

ゲノム・エピゲノムの変異・多型はさまざまなレベルに及ぶ。たとえば一塩基置換あるいは一塩基のメチル化から、染色体全体の再構成などさまざまな変異・多型が健常者あるいは疾患で生じている。

考える場合、それら変化が生殖細胞系列に生じて遺伝的背景として働いているのか（たとえば生活習慣病の罹患素因）、あるいは後天的に生じた体細胞変異として疾患発症に寄与するのか（たとえば発がんのドライバー変異）を常に区別して評価する必要がある。

ゲノムの先天的・後天的変化は多岐に及ぶが（図 1-2）、たとえば一塩基の置換が正常範囲の多型として遺伝的に受け継がれるものは一塩基多型 single nucleotide polymorphism (SNP) と呼ばれ、各個人ゲノムにおいて数百塩基対毎に存在すると予想されている。つまりハプロイドあたり数百万から 1 千万種類ほどの SNP が存在しており、個人間の SNP の共通性は血縁（進化の過程での分岐）の近さのバロメーターになると考えられる。さらに、たとえば EGFR 阻害薬である gefitinib による重症間質性肺炎がアジアの中でも特に日本にだけ多いことなどは、日本人の一部に固有の間質性肺炎罹患関連 SNP があることを示唆しているといえる。

医療福祉を考えた場合、SNP の生活習慣病における意義は大きい。高血圧や糖尿病などの生活習慣病に家族集積性が強いことから、これら生活習慣病は遺伝的背景の上に特定の生活嗜好（過食、塩分過摂取など）が重なると高頻度に発症すると考えられており、その遺伝的背景として SNP が大きな役割を果たすと予想されている。高頻度に認め

られる生活習慣病は高頻度に存在する SNP に起因するという「common disease, common variant 仮説」に基づき、これまでアレル頻度の高い（集団の 1% 以上に存在する）SNP に関する解析が大規模になされてきた。実際、糖尿病において人種固有あるいは人種を越えて存在する罹患関連 SNP が同定されてきたが、数千例という大規模な SNP 解析を行っても、SNP による寄与率で説明可能な 2 型糖尿病は 1 割に満たない。したがって近年は、アレル頻度の低い稀な SNP が生活習慣病発症に重要な役割を果たすのではないかと考えられるようになり<sup>4)</sup>、次世代シーケンサーを用いた全ゲノム解析による生活習慣病リスク多型の同定が試みられている。

また一塩基多型・変異のみならず、数塩基の挿入・欠失もしばしば認められ、たとえば CAG 配列のリピート数が異常に多くなると発症する疾患として、Huntington 舞蹈病、脆弱 X 症候群、Friedreich 失調症、筋緊張性ジストロフィーなどが知られている。

こうした塩基置換・挿入・欠失が、後天的に一部の細胞にのみ生じる例として最も重要なものは腫瘍である。がん細胞ゲノムを解析すると、多くの場合数万種類の体細胞変異が見つかり、そのうちアミノ酸置換をもたらすものは数百種類であることが多い。この中には変異の結果発がん能を獲得する遺伝子が生じることがあり、たとえば低分子量 G タンパク質である RAS ファミリー (KRAS/HRAS/NRAS) の活性型アミノ酸

置換は、ヒト悪性腫瘍で最も高頻度に見つかる発がん原因分子である。また EGFR キナーゼにおいてはインフレームな塩基欠失の結果、数アミノ酸が取り除かれ、その結果 EGFR の酵素活性が恒常的に上昇して肺がん発症を導くことが明らかになっている。さらには遺伝子増幅・欠失もがん細胞でしばしば認められ、直接発がんに寄与する遺伝子増幅として神経芽腫における MYCN、乳がんにおける ERBB2 が知られている。

染色体の部分的な増幅・欠失、染色体再構成さらに異数性などは、先天的変化としても、あるいは後天的異常としても観察される。我々のゲノムには予想外にコピー数多型が多い（全ゲノムの約 1 割の領域）ことが発見され<sup>5)</sup>、その生活習慣病あるいは精神疾患への関連が示唆されている。また先天的な染色体異数性としては 21 番染色体が 3 本存在する Down 症候群、性染色体が XXY となる Klinefelter 症候群などが有名である。

一方こうした染色体の構造異常はがん細胞においてしばしば認められる。染色体の短腕・長腕全体あるいはその一部の欠失・増幅は極めて多くのがんで認められており、たとえば 5q 欠失症候群のように、骨髄異形成症候群内の特定のサブグループを形成するものもある。さらに 9 番染色体と 22 番染色体の相互転座 t(9;22) により生じる BCR-ABL1 融合遺伝子、2 番染色体内の微小逆位 inv(2) (p21p23) による EML4-ALK 融合遺伝子のように転座点が発がんの直接的な原因となるものも報告されている。

### C. エピゲノム異常と疾患病態

ゲノムの配列に直接変化を与えないが、細胞分裂を超えてゲノムの転写制御を司るシステムをエピゲノムと呼び、それを解析する手法をエピゲノミクスと呼ぶ。具体的なエピゲノム制御機構は多岐にわたり（図 1-2）、ゲノム中の CpG 配列にメチル基が付与されることはよく知られている。この DNA メチル化はほとんどの場合遺伝子発現の抑制をもたらす。ヒトゲノム中で CpG 配列が密集している領域を CpG アイランドと呼ぶが、遺伝子のプロモーター領域に多く存在することが知られる。がんにおいては CpG アイランドのメチル化が亢進し、がん抑制遺伝子の発現低下がしばしば認められるとともに、CpG アイランド以外のゲノムの脱メチル化が亢進し、染色体の不安定性を誘導すると考えられている。

またゲノムは H2A/H2B/H3/H4 からなるコアヒストンの周りを一周してコンパクトな構造を維持している。ヒストン（特に H3, H4）はしばしばアセチル化、メチル化、リン酸化などの修飾を受け、その結果ゲノム DNA の高次構造を変化させる。たとえばヒストンのアセチル化は DNA のヒストンからの遊離を促し、転写の促進をもたらす。一方ヒストンのメチル化は、メチル化されるアミノ酸の部位によって転写の促進あるいは抑制のどちらも誘導可能なことが知られる。ヒストンメチル化酵素である MLL がさまざまな白血病で異常融合タンパク質となっていること、ヒストン脱メチル化酵素である KDM5C が腎臓がんでしばしばアミノ酸置換を有することなどは、ヒストン修飾が発がんにいかに関与かを物語っている。また ATP 依存性に染色体の高次構造を制御するタンパク質群をリモデリング因子と呼び、たとえば SWI/SNF 複合体のメンバーの異常が多くのがん種で発見されている。

一方こうしたエピジェネティック変化の生活習慣病における意義はまだ十分には解析されていない。しかし妊婦の低栄養（あるいは過剰栄養）が胎児のエピゲノムに影響を与え、出生後の糖尿病発症率を有意に上昇させることなどは今後の重要なテーマであろう<sup>6)</sup>。

### D. 医療の革新

こうしたゲノム・エピゲノム解析は、我々の疾患病態の理解を変えるだけでなく、疾患の診断法・治療法にも変革をもたらすであろう。特に変異あたりの疾患寄与率が大きいがんの場合、遺伝子変異の解明はがんの分類自体も変えると予想される。

遺伝的素因としての多型についても、たとえば CYP2C9 遺伝子と VKORC1 遺伝子の SNP によってワーファリンの代謝率に変化が生じ、血中濃度に影響が出るということが明らかになった。これらを受けて CYP2C9 と VKORC1 の SNP によってワーファリン投与量を調整することが推奨されている<sup>7)</sup>。糖尿病あるいは他の生活習慣病において、治療への介入が推奨されるに至った SNP はまだないが、次世代シーケンサーを用いた低アレル頻度 SNP の大規模解析の結果、寄与率の大きな罹患関連 SNP が発見されれば、生活習慣への介入なども現実のものになるであろう。ただしそのような SNP は個人の重要な遺伝情報であるので、SNP 情報の秘匿性を担保する法的整備も喫緊の課題である。

がんの診断・治療については、既にゲノム情報に基づく治療介入が広く行われはじめている。これまでのように、肺がんや胃がんという発症臓器（およびその病理型）に基づく治療選択から、「どのような発がん分子によって発症しているか」に基づく治療法選択へとパラダイムシフトが生じている<sup>8)</sup>。たとえば2012年の米国におけるがん治療ガイドラインでは、(1) 肺がんについてはまずEGFR変異とEML4-ALK融合遺伝子の有無を検査すること、そして(2) 何れかが陽性であればそれぞれに対応するキナーゼ阻害薬治療を第一選択とすること、が明記されている(National Comprehensive Cancer Network: <http://www.nccn.org/index.asp>)。

同様に白血病のうちBCR-ABL1融合遺伝子陽性サブタイプに対してのみABL1阻害薬は利用され、RARA遺伝子融合型白血病に対してのみオールトランスレチノイン酸治療が選択される。こうして発がん原因に基づく薬剤選択が現実のものになり、その傾向は、発がん原因がさらに解明されるにつれ、より加速して行くであろう。

またこのような流れは、疾患のゲノム・エピゲノム診断の重要性をさらに増すことになる。たとえば近未来のがん医療においては、どのようながん種であっても最初の診断時に、少なくとも治療薬が存在する（あるいは臨床試験が行われている）標的分子の全てについてゲノム解析がなされるであろう。ヒトゲノムには

約2万種類のPCGがあるが、その中で実際の発がんに寄与するものは数百個であろう。これら全てについてがん化変異が存在するかをまず検討し、その結果に基づいて治療薬が選択される時代になると予想される。また既にこのようなアプローチを商業ベースで実施して成功している会社もある(<http://www.foundationmedicine.com/index.php>)。ゲノム情報に基づく病態の理解は、我々の疾患概念を変え、患者の診断・治療の在り方を大きく様変わりさせているのだ。

#### ■文献

- 1) International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature*. 2004; 431: 931-45.
- 2) Pennisi E. Genomics. ENCODE project writes eulogy for junk DNA. *Science*. 2012; 337: 1159, 61.
- 3) Nagano T, et al. No-nonsense functions for long noncoding RNAs. *Cell*. 2011; 145: 178-81.
- 4) Manolio TA, et al. Finding the missing heritability of complex diseases. *Nature*. 2009; 461: 747-53.
- 5) Redon R, et al. Global variation in copy number in the human genome. *Nature*. 2006; 444: 444-54.
- 6) Gluckman PD, et al. Effect of in utero and early-life conditions on adult health and disease. *N Engl J Med*. 2008; 359: 61-73.
- 7) Wang L, et al. Genomics and drug response. *N Engl J Med*. 2011; 364: 1144-53.
- 8) Mano H. ALKoma: a cancer subtype with a shared target. *Cancer Discov*. 2012; 2: 495-502.

<間野博行>

## 2 ゲノミクスと病態解析

### ◆まとめ

1. 機能ゲノム科学の進歩は、癌などの生活習慣病の診断、治療、予防に新たなツールをオーダーメイドに提供しつつある。
2. 新世代高密度化マイクロアレイ解析は、病態解析に有用である。

ヒトゲノム全 DNA 配列の解読が完了後、世界の研究者の関心は構造（塩基配列）からゲノムに記されている情報（遺伝子の機能）の読解（機能ゲノム科学 functional genomics）へと、すでにポストゲノム（シーケンス）時代へ突入している。このヒトゲノム計画によって大きく進展が期待されるのは、ヒトの病気の原因解明・診断・治療といった医療分野である。ヒトゲノム計画の成果により、診断から使用する薬の製造までのすべての過程は大きく影響を受け、近い将来には“ありふれた生活習慣病”に対しても個々の患者の遺伝的体質に合わせた処方、治療計画がなされる、いわゆるオーダーメイド医療が提供される可能性が高まっている。さらには、それら疾患を予防するためのオーダーメイド予防対策が提供されてくると考えられる。ゲノム塩基配列の変化によらない遺伝子発現の調節機構をエピジェネティックな制御と呼ぶが、この領域はポストゲノム研究として注目を集めている。さらに、RNAi や non-coding RNA など、遺伝子発現を制御する機能をそなえた新たな RNA が数多く発見されてきている。本稿では、分子病態学におけるゲノミクスの最前線ともいえる DNA チップ・マイクロアレイ技術の応用を中心に解説する。

### A. マイクロアレイ技術と種類

DNA から読み取られ、タンパク質合成の鋳型となっている RNA はヒト遺伝子では3万~4万種と考えられている。その RNA のなかから、発現する RNA の組み合わせや、それらの発現量の違いによって、組織、細胞に特異的な RNA の発現パターンが現れ、それが種々の病気に罹患することで、健常時とは異なる発現パターンを示すようになる。つまり、3万~4万個の

RNA の発現パターンを観察することができれば、より正確な細胞の状態を理解することができ、ひいては、癌などの生活習慣病の診断、予防へつなげることが期待されている。

マイクロアレイ技術は、発現アレイ、一塩基多型 (SNP) 解析, comparative genomic hybridization (CGH) など、その目的に応じて使い分けられ応用されてきている。発現解析に使用するマイクロアレイには、cDNA だけでなく 20~120 mer のオリゴヌクレオチドをプローブ DNA としてスポットするタイプがある。Affymetrix 社から提供されているものは後者によるもので、我々も主に発現解析に利用してきた。すべての遺伝子断片が1枚のガラス板上に固定されており、このプローブと呼ばれる遺伝子断片と、ターゲットと呼ばれる mRNA を逆転写酵素で相補的 DNA (cDNA) に変換したものをハイブリダイゼーションすることによって、細胞や組織内で発現している遺伝子情報を網羅的に検出することが可能となっている (図 1-3)<sup>1-3)</sup>。

Affymetrix 社の GeneChip とスタンフォード大学の Brown 研究室が開発した cDNA マイクロアレイの2つの方式が主に使用されている。GeneChip は数十万個のオリゴヌクレオチドプローブをフォトリソグラフィ技術と固相反応化学技術を使用して、基板上で 20~25 mer のオリゴヌクレオチドを人工的に合成することにより作製される。このオリゴヌクレオチドは、あらかじめ遺伝子の特異的な塩基配列を特定するためにコンピューターを用いて位置や長さなどがデザインされている。特に、特定遺伝子と完全に相補的になるようにデザインされたプローブパーフェクトマッチ (PM) だけでなく、ミスマッチ (MM) と呼ばれる非特異的な塩基配列もプローブとして配置することによって、非特異的なクロスハイブリダイゼーションの定量値をシグナル値から減算でき、定量性を向上させていることも大きな特徴である。たとえば、GeneChip Human Gene ST Array では mRNA と遺伝子間の長い非コード RNA (lincRNA) の両方の転写産物を測定するためのプローブを含む全転写産物アレイとしてデザインされている。このような全転写産物をカバーするアレイデザインにより、mRNA の完全な発現プロファイル