

S2 群(ナノ・量子・バイオ) - 6 編(バイオインフォマティクス)

2 章 分子生物学の基礎的事項

(執筆者：福岡 豊)[2018年2月受領]

概要

Watson と Crick の DNA 二重らせん構造の発見以来、分子レベルで生命現象を扱う分子生物学が発展してきた¹⁾。親から子に受け継がれる遺伝情報の担体が DNA と呼ばれる分子であり、その情報が塩基の配列で表現されていることが明らかにされた。様々な遺伝子の配列が解読され、研究者間での情報共有のためにデータベース化されてきた。ヒトは 30 億の塩基に及ぶ膨大な情報を持っており、その解析にコンピュータが応用されるようになった。また、多くの分子の相互作用によって、生命現象が営まれていることも明らかになってきた。複雑な現象を解き明かすためにも、シミュレーションなどを用いたアプローチが用いられるようになった。このように、生命科学の発展に、計算論的アプローチは不可欠なものとなっている。

そのような背景のもとに、バイオインフォマティクスやシステムバイオロジーという研究分野が発展してきた。中心となるのは工学技術であるが、遺伝子やタンパク質に関する情報を扱うので分子生物学の知識が不可欠となる。本章では、これらの分野の研究に必要な最低限の事項に絞って説明する。分子生物学に興味のある読者や実際にこれらの分野の研究を行いたい読者は参考文献^{1, 2, 3)}を参照されたい。

【本章の構成】

2-1 節では、分子生物学のセントラルドグマと呼ばれる DNA RNA タンパク質の一連の反応について説明する。前述のように、タンパク質をはじめとする多数の分子の相互作用によって生命現象が営まれている。2-2 節では、タンパク質間の相互作用について説明する。ヒトの DNA 配列は、平均すると 1000 塩基に 1 か所程度の割合で標準的な配列と異なる点があると言われている。2-3 節では、このような遺伝子の変異について説明する。2-4 節では、タンパク質をコードしない non-coding RNA について説明する。エピジェネティクスとは、塩基配列の変化を伴わないが、遺伝的背景に影響を与える要因の総称である。2-5 節では、代表的なエピジェネティクスについて述べる。2-6 節では、進化とゲノムの関係を説明する。

参考文献

- 1) 中村桂子, 松原謙一 (監訳): “細胞の分子生物学 (第 6 版)”, ニュートンプレス, 2017.
- 2) 井出利憲: “分子生物学講義中継 part 1”, 羊土社, 2002.
- 3) 中村桂子, 松原謙一 (監訳): “細胞生物学 (原書第 4 版)”: “南江堂, 2016.

S2 群 - 6 編 - 2 章

2-1 セントラルドグマ^{1,2)}

(執筆者：福岡 豊)[2018 年 2 月受領]

生物の遺伝情報を担っているのは DNA (デオキシリボ核酸：Deoxyribonucleic Acid) と呼ばれる分子である。DNA の基本要素はヌクレオチドであり、糖 (デオキシリボース)・リン酸・塩基の 3 つの部分から構成される。塩基にはアデニン (A)、シトシン (C)、グアニン (G)、チミン (T) の 4 種類がある。以下では、4 種類の塩基を A、C、G、T と表記する。DNA はいわゆる二重らせん構造を持っている。多数のヌクレオチドが繋がったものが一本鎖 DNA であり、2 つの一本鎖 DNA が塩基の部分で結合して二重らせん構造となる。その際、A と T が特異的に結合し、G と C が特異的に結合する。これを相補性という。二重らせんの一方の鎖 (ストランド) における 4 種類の塩基の並び方 (塩基配列) が決まると、他方のストランドの塩基配列は相補性により自動的に決定される (図 2.1)。

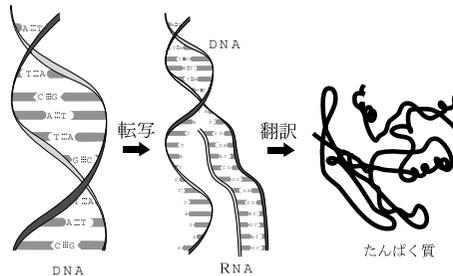


図 2.1 セントラルドグマ

DNA 上の一部の領域の塩基配列情報は RNA (リボ核酸：Ribonucleic Acid) と呼ばれる分子にコピー (転写) される。この際、二重らせんの一部がほどけて、RNA が合成される (図 2.1)。転写された RNA の塩基配列情報に基づきタンパク質が合成される。RNA からタンパク質が合成される反応を翻訳という。翻訳に使われる RNA をメッセンジャー RNA と呼ぶ (以下では mRNA と書く)。RNA も糖 (リボース)・リン酸・塩基という 3 つの部分から構成され、塩基の種類も 4 種であり、DNA とよく似た構造を持つ。しかし、RNA では T ではなくウラシル (U) が使われる。DNA と RNA を総称して核酸という。

この DNA RNA タンパク質という情報の流れは、大腸菌などの単細胞生物からヒトをはじめとする高等生物まで共通である。これを分子生物学のセントラルドグマという。近年、エイズウイルスなどの RNA ウイルスが RNA から DNA を合成する酵素を持っていることが明らかとなった。このような酵素を逆転写酵素という。逆転写という例外を除けば、細胞内での反応過程はセントラルドグマに従っている。

2-1-1 染色体と遺伝子

生物の細胞は DNA を含んでいる。大腸菌などの原核生物には細胞核がなく、プラスミドと呼ばれる環状の構造を持った DNA が細胞内に存在する。これに対し、真核生物の細胞には核があり、DNA はその中に格納されている。核内の DNA は染色体と呼ばれる構造を持

つ．染色体の数は、それぞれの生物に固有であり、ヒトでは 22 種類の常染色体と 2 種類の性染色体がある．染色体の両端はテロメアという特殊な構造を持ち、そのほぼ中央にセントロメアという別の構造がある（図 2.2A）．テロメアとセントロメアの間が短い方を短腕といい、もう一方を長腕という．染色体上の位置は、短腕のテロメア側を上流、長腕のテロメア側を下流として表す．また、一本鎖 DNA についても分子の構造に基づき向きが定義されており、5' 側が上流、3' 側が下流である（図 2.2B）．この 5 や 3 という数字はデオキシリボースに含まれる炭素の番号であり、分子構造によって一意に定まる．

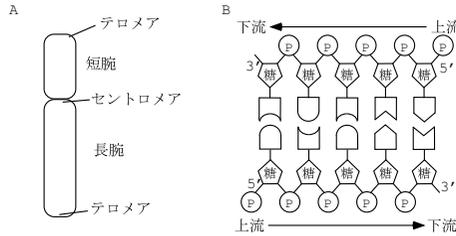


図 2.2 染色体 (A) と DNA の向き (B)

遺伝情報は DNA の塩基配列により表現されている．遺伝子とは遺伝的な形質を規定する因子であり、mRNA に転写された後、タンパク質に翻訳される部分である．遺伝子あるいは DNA 断片の長さは塩基対 (base pair: bp) を単位として測る．原核生物の遺伝子は単純な構造を持ち、転写された RNA がほぼそのままタンパク質に翻訳される．一方、真核生物（特に高等生物）の遺伝子には、エクソンとイントロンと呼ばれる 2 種類の領域がある．転写後にイントロンの部分を切り離して、エクソンのみをつなぎ合わせた mRNA が作られる（図 2.3）．この反応をスプライシングといい、反応前後の mRNA をそれぞれ Unspliced mRNA と Spliced mRNA と呼ぶ．高等生物では、1 つの遺伝子から異なる Spliced mRNA が作られることがある．すなわち、エクソンの組合せ方は固定的ではなく、異なる組合せが可能である．これを選択的スプライシングと呼ぶ．この仕組みによって、1 つの遺伝子から異なるアミノ酸配列を持つタンパク質を合成することが可能となる．

遺伝子はアルファベット 3～5 文字程度のシンボルで表される．これは遺伝子シンボルと呼ばれるもので、英語表記したときの単語の一字目を組み合わせることが多い．例えば、EGF という遺伝子は Epidermal Growth Factor という正式名称を持つ．生物種によって慣例的な命名法・表記法がある．1 つの遺伝子に複数の名称・シンボルがつけられていることがある．正式名称以外は、シノニム (Synonyms) や Aliases と呼ばれる．よく似た遺伝子が複数ある（このような一群の遺伝子をファミリーという）場合には、名称によるシンボルの後に 1, 2 や A, B などをつける．また、複数の名称・シンボルがあると混乱が生じるので、データベースでは ID を用いて管理されている．しかし、米国 (Entrez Gene ID) と欧州 (Ensembl) で異なる ID が使われるなど、完全に統一されているわけではない．

ヒトは 22 種類の常染色体をそれぞれ 2 本ずつと 2 本の性染色体の計 46 の染色体を持っている．性染色体は男性では X と Y、女性では X と X である．このようにヒトでは同じ常染色体を 2 本ずつ持っている．これはゲノムを 2 セット持つことを意味しており、このよう

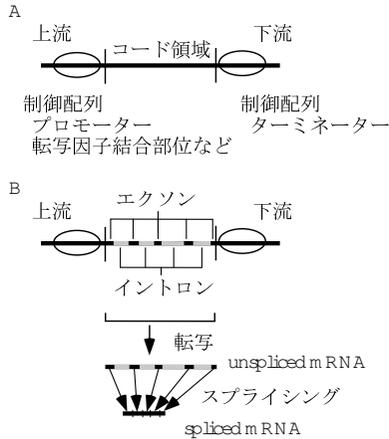


図 2.3 コード領域と制御配列 (A), エクソンとイントロン (B)

な生物を 2 倍体という。実際には、父親と母親に由来する遺伝子を持つことになるので、完全に同じ配列を持つ遺伝子であるとは限らない。そこで、染色体上の同じ位置にある遺伝子を対立遺伝子（アレル）と呼ぶ。対立遺伝子のなかには同じ配列を持つものや異なる配列を持つものがある。

対立遺伝子の片方が疾患を引き起こす異常な遺伝子であった場合でも、異常の現れ方には何通りかがある。異常な遺伝子を持つ場合に直ちに異常が現れるとき、その遺伝子は優性といわれる。一方、異常な遺伝子が 2 つ揃わないと疾患が現れないときは、その遺伝子は劣性といわれる。すなわち、片方が優性で、もう一方が劣性のときは、優性の性質が表れる。このように表に現れる性質（形質と呼ぶ）を表現型（フェノタイプ）という。遺伝子の優性、劣性と表現型の関係は、発見者の名を冠してメンデルの法則と呼ばれる。

ヒトの 1 つの細胞中にある DNA を 1 本につなげるとおよそ 2 m になる。これを細胞の核内に納めるにはコンパクトに折りたたむ必要がある。そのために、DNA はヒストンと呼ばれるタンパク質複合体（2-2 節参照）と、クロマチンという複合体をなしている。ヒストンに約 150 塩基の DNA が巻き付いた構造をヌクレオソームと呼び、クロマチンの基本構成単位となっている。クロマチンは真核生物に共通する構造である。

2-1-2 遺伝子の発現

ある遺伝子が mRNA に転写され、タンパク質に翻訳されることを遺伝子の発現という。セントラルドグマに従って、転写に続き翻訳という順序でタンパク質が合成される。それぞれの反応過程について説明する。

転写は RNA ポリメラーゼという酵素によって行われる。RNA ポリメラーゼは、リボヌクレオチド（リボース・リン酸・塩基）を重合させて RNA を合成する。mRNA だけでなく、後述する tRNA なども転写によって合成される。単純化のために、以下では mRNA の転写に話を限定する。

転写の過程は、原核生物と真核生物で異なる。大腸菌などの原核生物は核を持たないので、転写は細胞質で行われる。大腸菌の RNA ポリメラーゼは、5 つのサブユニットから構成され、このうちの 하나가 DNA の配列を認識して結合する。転写が始まる点を転写開始点というが、その上流に RNA ポリメラーゼが結合する部位がある。この部位をプロモータという。すなわち、原核生物の転写は、プロモータに RNA ポリメラーゼが結合することによって始まる。特定の配列（ターミネータ）を認識すると転写が終了し、RNA ポリメラーゼが DNA から離れる。

真核生物では、まずプロモータに転写因子が結合する。転写因子は、遺伝子の転写の促進、抑制に関わるタンパク質であり、DNA の配列を認識して結合する。プロモータは特定の配列を持つことが多いが、完全に同一な配列でなくても転写因子は結合する。転写因子の結合に続いて、RNA ポリメラーゼが DNA に結合し、転写が始まる。真核生物には、3 種類の RNA ポリメラーゼが存在するが、mRNA の転写に関わるのは RNA ポリメラーゼ II である。真核生物においてもターミネータを認識すると転写が終了する。転写直後の mRNA は、両端に特定の分子が結合される。その後、2-1-1 項で述べたように、スプライシングという過程で、イントロンが切り取られ、成熟 mRNA (Spliced mRNA) となる。これらの一連の反応は核内で行われる。成熟 mRNA は核膜孔を通して細胞質に運ばれる。

原核生物、真核生物ともに、翻訳は細胞質にあるリボソームで行われる。真核生物のリボソームは原核生物より大きいですが、翻訳の過程は似ている。タンパク質を構成するアミノ酸は 20 種類あり (2-1-3 項参照)、4 種類の塩基でコードするためには 3 塩基が必要となる。この 3 塩基の組をコドンと呼ぶ。⁴³ で 64 通りのコドンが 20 種のアミノ酸に対応している。コドンに対応するアミノ酸が tRNA (transfer RNA) によって運ばれて、リボソームでペプチド結合される。多数のアミノ酸が結合されたポリペプチド (タンパク質) はフォールディングと呼ばれる過程を経て、一定の立体構造を持つようになる。

必要となるタンパク質の種類や量は、細胞の種類や状態によって異なる。そこで、遺伝子発現量の調節のための様々な仕組みがある。代表的なものとして、転写因子による調節が挙げられる。転写因子もタンパク質であるので、遺伝子から合成される。自身の転写を抑制する転写因子が見つっている。また、クロマチン (2-1-1 項参照) の状態によっても発現量が変化する。ヒストンが凝集している状態をヘテロクロマチンといい、DNA に転写因子が結合できないので遺伝子は発現しない。逆に、ヒストンの間隔が開いている状態をユークロマチンといい、転写因子などが結合できるので、遺伝子の発現が可能である。なお、遺伝子の発現量 (すなわち、タンパク質の量) は mRNA 量にほぼ比例すると考えられている。

2-1-3 タンパク質

タンパク質の基本構成要素はアミノ酸であり、一部の例外を除いて 20 種類のアミノ酸がタンパク質の構成要素として使われている (表 2.1)。4 種類の塩基で 20 種類のアミノ酸をコードするために、3 つの塩基の組合せで 1 つのアミノ酸を表現している。この 3 つの塩基の組合せをコドンと呼ぶ。コドンとアミノ酸の対応は、生物種によらず一定である。この関係を遺伝暗号と呼び、対応関係の表 (コドン表) があれば、mRNA の配列からアミノ酸の配列を決定できる。また、20 種類のアミノ酸は、親水性のものと疎水性などの異なる性質を持つものに分類される (表 2.1)。

表 2.1 アミノ酸

アミノ酸	記号	コドン	性質
アラニン	A (Ala)	CGA, GCC, GCG, GCU	疎水性
システイン	C (Cys)	UGC, UGU	疎水性
アスパラギン酸	D (Asp)	GAC, GAU	親水性, 酸性
グルタミン酸	E (Glu)	GAA, GAG	親水性, 酸性
フェニルアラニン	F (Phe)	UUC, UUU	疎水性
グリシン	G (gly)	GGA, GGC, GGG, GGU	疎水性
ヒスチジン	H (His)	CAC, CAU	親水性, 塩基性
イソロイシン	I (Ile)	AUA, AUC, AUU	疎水性
リシン	K (Lys)	AAA, AAG	親水性, 塩基性
ロイシン	L (Leu)	UUA, UUG, CUA, CUC, CUG, CUU	疎水性
メチオニン	M (Met)	AUG	疎水性
アスパラギン	N (Asn)	AAC, AAU	親水性
プロリン	P (Pro)	CCA, CCC, CCG, CCU	疎水性
グルタミン	Q (Gln)	CAA, CAG	親水性
アルギニン	R (Arg)	AGA, AGG, CGA, CGC, CGG, CGU	親水性, 塩基性
セリン	S (Ser)	AGC, AGU, UCA, UCC, UCG, UCU	親水性
スレオニン	T (Thr)	ACA, ACC, ACG, ACU	親水性
バリン	V (Val)	GUA, GUC, GUG, GUU	疎水性
トリプトファン	W (Trp)	UGG	疎水性
チロシン	Y (Tyr)	UAC, UAU	疎水性

タンパク質は、50 以上のアミノ酸がペプチド結合によって鎖状に連結した生体高分子である（アミノ酸の数が 50 未満の場合はペプチドと呼ばれる）。アミノ酸の分子構造によって、鎖の向きが決められる。アミノ基の側を N 末端といい、mRNA の 5' 側（上流側）に相当する。3' 側（下流側）は C 末端と呼ばれる。アミノ酸の配列は N 末端から C 末端に向けて書く。

鎖状のタンパク質が合成されると、異なる部位にあるアミノ酸と結合して折りたたまれ（フォールディング）、一定の立体構造をとる。その際に、各アミノ酸の性質（親水性 / 疎水性など）や介在分子（分子シャペロン）の存在が重要である。タンパク質はほかのタンパク質と相互作用しながら機能を果たす。相互作用の相手は、形状（すなわち、立体構造）によって認識される。したがって、タンパク質が正常に機能するためには、一定の立体構造をとる必要がある。遺伝子の配列に異常（突然変異など、2-3 節参照）があるとアミノ酸配列が変化して、立体構造に変化生じる。このようなタンパク質は正常な機能を果たすことができなくなり、疾患の原因となることがある。

このように、機能を考えるうえで立体構造は重要であるが、特定のタンパク質の立体構造を決定するのは容易ではない。分子内の各原子間に働く力を計算して立体構造を予測する分子動力学や、結晶化して X 線回折像から構造を同定する方法が用いられている。また、構造が既知なタンパク質のアミノ酸配列から構造を予測する方法も用いられる。この方法では、しばしば部分構造の予測も行われる。タンパク質には ヘリックス（らせん構造）や シー

トなどの特徴的な部分構造が存在する．このような部分構造を構造モチーフと呼ぶ．タンパク質全体の立体構造の予測が難しい場合でも，構造モチーフに対応するアミノ酸配列が予測できる場合がある．

タンパク質の構造は，一次構造から四次構造までの階層で考えることが多い．タンパク質を構成するアミノ酸の配列を一次構造と呼ぶ．二次構造は構造モチーフ，三次構造は立体構造である．四次構造とは，複数数のタンパク質が結合した構造（複合体）を指す．2つのタンパク質（サブユニット）からなる複合体をダイマー（2量体）という．2つが同じ場合はホモダイマー，異なる場合はヘテロダイマーと呼ぶ．3つ以上に関してもトリマー（3量体），テトラマー（4量体）などの名前がついている．

参考文献

- 1) 中村桂子, 松原謙一 (監訳): “細胞の分子生物学 (第6版),” ニュートンプレス, 2017.
- 2) 中村桂子, 松原謙一 (監訳): “細胞生物学 (原書第4版),” 南江堂, 2016.

S2 群 - 6 編 - 2 章

2-2 タンパク質間の相互作用

(執筆者: 福岡 豊) [2018 年 2 月受領]

タンパク質の機能には、酵素、輸送、シグナル伝達、受容体、転写因子、構造タンパク、モータータンパクなどがある¹⁾。これらの機能は、多数のタンパク質の相互作用(Protein-Protein Interaction: PPI)によって営まれている。したがって、正常な機能を営む PPI のネットワークについて様々な解析が行われ、データベース化されている²⁾。また、代謝経路に関する一連の PPI をパスウェイと呼んでいたが、近年では代謝以外の相互作用も同様にパスウェイと呼ばれている。パスウェイは PPI と同義語であるが、前者の方が経路というニュアンスが強い。パスウェイについても多くのデータベースが存在する。京都大学による KEGG (Kyoto Encyclopedia of Genes and Genomes)³⁾は代表的なパスウェイデータベースである。

タンパク質の相互作用の方式は何通りに分類できる。まずは、複数のタンパク質が結合して複合体を形成する方式があげられる。ほかには、タンパク質が分子スイッチの役割を果たすシグナル伝達系と呼ばれる相互作用が存在する。

前述のように、複数のタンパク質が特異的に結合した集合体をタンパク質複合体と呼ぶ。代表的なタンパク質複合体として、ヒストンやリボソームが挙げられる。ヒストンは 4 種類のタンパク質が 2 つずつ、計 8 分子が結合してできるタンパク質複合体である。このような分子を 8 量体(オクタマー)という。直径 10 nm 程度のドラム型をしており、ドラムの周囲に 147 塩基対の DNA が 2 周巻き付いて、ヌクレオソームを形成する。ヒストンは DNA を核内に収納するのに重要な役割を果たす。近年、ヒストンを構成するタンパク質からアミノ酸の鎖(尾部という)が出ており、その部分がメチル基などの化学修飾を受けることが分かった。このような化学修飾によって、クロマチンの状態(ユークロマチン、ヘテロクロマチン)に変化が生じ、遺伝子発現の制御に関与していることが明らかになっている¹⁾。

リボソームは、80 種類以上のタンパク質と数種類の rRNA (ribosomal RNA) の巨大な複合体である。大小 2 つのサブユニットからなる。2-1 節で述べたように、タンパク質の合成はリボソームで行われる。まず小サブユニットが mRNA に結合し、続いて大サブユニットが結合する。リボソームが mRNA 上を移動しながら、N 末端から C 末端に向けてタンパク質が合成される。

もう一つの相互作用の方式であるシグナル伝達系は、細胞内の情報伝達機を担っている。多くの細胞は、細胞膜上に受容体と呼ばれる機構を持っている。ほかの細胞から分泌されたホルモンなどが受容体に結合し、その情報が細胞内に伝わる。細胞膜上の受容体に結合する分子を細胞外シグナルと呼ぶ。受容体には特異性があり、特定の細胞外シグナル分子しか結合しない。

細胞内では何段階かの過程(カスケード)を経て、特定の遺伝子の転写調節などの変化が生じる。最終的に生じる細胞内の変化も細胞外シグナル分子ごとに決まっている。細胞内での過程は、分子スイッチによる多段階の反応系である。分子スイッチとして働く分子を細胞内シグナルまたはセカンドメッセンジャーと呼ぶ。段階を経るごとに関与するセカンドメッセンジャー(酵素など)の数が増大する。すなわち、シグナル伝達系には増幅作用がある。セカンドメッセンジャーのうち、タンパク質はリン酸化によって、その活性が切り替わるものが多い。このようにタンパク質をリン酸化する酵素をキナーゼ(タンパク質リン酸化酵素)

と呼び、ほかのタンパク質を活性化する働きを持つ。一方、タンパク質脱リン酸化酵素も存在し、タンパク質を不活性化し、その働きを抑制する。

ここでは、タンパク質の相互作用の代表的な 2 つの方式について説明した。物理的な実態としては生化学反応であるが、信号の伝達、増幅、スイッチなどの概念が適用でき、システムの観点から生命現象を解析することができる。このような動向については、6 章で説明する。

参考文献

- 1) 中村桂子, 松原謙一 (監訳) : “細胞生物学 (原書第 4 版)”, 南江堂, 2016.
- 2) D. Szklarczyk, J. H. Morris, H. Cook, M. Kuhn, S. Wyder, M. Simonovic, A. Santos, N. T. Doncheva, A. Roth, P. Bork, L. J. Jensen, and C. von Mering : “ The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible, ” *Nucleic Acids Res*, vol.45, pp.D362-D368, 2017.
- 3) M. Kanehisa and S. Goto : “ KEGG: Kyoto Encyclopedia of Genes and Genomes, ” *Nucleic Acids Res*, vol.28, pp.27-30, 2000.

S2 群 - 6 編 - 2 章

2-3 遺伝子の変異

(執筆者：福岡 豊)[2018年2月受領]

DNA の配列は、生物種ごとにほぼ一定である。ある生物種の標準的な DNA 配列の全体をゲノム (genome) という。例えば、生物種としてのヒトのゲノムをヒトゲノムという。各個人 (個体) の DNA には、標準的な配列と異なる部分が数多く存在する。そのような部分のうち、人口の 1% 以上の頻度で観測されるものを多型と呼び、それより頻度の低いものを変異という¹⁾。多型や変異には疾患や薬剤応答に関連するものがあり、個人のゲノム解析が進められている。

2-3-1 多 型

多型の大多数は一塩基多型 (SNP) である。SNP は文字通り、一塩基が別のものに置き換わった多型である。このように別の塩基に置き換わることを置換という。一塩基のみの置換なので、その影響は小さい。SNP がエクソンに存在したとしても、影響を受けるアミノ酸は 1 つ以下である。表 2.1 に示されるように、アラニンやグリシンなどの 8 つのアミノ酸では、コドンの 3 番目に 4 種類の塩基を含んでいる。例えば、GC の後はどの塩基でもアラニンに対応する。したがって、そのような位置に SNP が存在しても、合成されるタンパク質には影響がない。このように、アミノ酸の変化を伴わない置換を同義置換という。一方、コドンの 1 番目や 2 番目に SNP が存在する場合には、アミノ酸の変化を伴う。これを非同義置換という。この場合でも、1 つのアミノ酸のみの変化なので、合成されるタンパク質の性質を大きく変える可能性は低い。したがって、SNP は進化の過程で淘汰されることなく、子孫に受け継がれる可能性が高い。

ほかの種類が多型も存在する。ゲノム中には、数～数十塩基程度の繰り返し配列が存在する部位がある。この繰り返し数が異なる多型がある。このような多型のうち、2～7 塩基程度の配列の繰り返し数が異なるものをマイクロサテライト多型という。また、標準配列には存在しない塩基が挿入されている多型や塩基が欠失している多型もある。しかし、挿入や欠失が生じると、その部位以降のアミノ酸すべてに変化が生じるため、合成されるタンパク質の性質が大きく異なることになる。そのため、淘汰の可能性が高くなり、SNP と比べて、このような多型の数はかなり少ない。

2-3-2 突然変異

生物やウイルスの核酸 (DNA あるいは RNA) の配列に生じる変化を遺伝子突然変異という。このほかに、染色体の数や構造に変化が生じる染色体突然変異 (ダウン症などの原因となる) がある。遺伝子突然変異の原因としては、化学物質や放射線による DNA の損傷や DNA の複製ミスなどが挙げられる。置換、欠失、挿入の何れも突然変異によって生じたものである。遺伝子突然変異には、淘汰に有利でも不利でもない中立的なものがあり、そのような変異が集団に広がると多型となる。生存にとって有利な場合も、集団に広がることになる。このような変化は進化につながる。一方、生存に著しく不利な場合は淘汰されてしまい、広まることはない。突然変異体 (ミュータント) とは、表現型に変化が生じた個体または細胞のことである。遺伝子機能の解析などに用いられる。

ウイルスは、厳密な定義では生物に当てはまらないが、生物と似た性質を持っている。ウイルスの突然変異で問題となるのは、遺伝子配列の変異によって薬剤に対する耐性を獲得することである。突然変異が頻繁に生じるウイルス（エイズウイルスなど）では、薬剤耐性を獲得するまでの期間が短く、多剤を併用するなど治療に工夫が必要である。

染色体突然変異には、染色体の部分的な異常（転座など）と染色体数が異なる異数体がある。多くのがんでは、染色体に異常があることが知られている。通常は各常染色体を一对（2本ずつ）持つが、がん細胞では染色体数が増減することがある。また、ある染色体の一部が別の染色体の一部と融合する転座が、白血病などがんの原因となることが分かっている。

参考文献

- 1) 菅野純夫 (監修), 服部成介, 水島-菅野純子: “よくわかるゲノム医学 (改定第 2 版),” 羊土社, 2015.
- 2) 井出利憲: “分子生物学講義中継 Part 1,” 羊土社, 2002.

S2 群 - 6 編 - 2 章

2-4 non-coding RNA

(執筆者: 福岡 豊) [2018 年 2 月 受領]

2-1 節で述べたように、RNA は DNA によく似た分子である。DNA から転写され、タンパク質の合成に使われる mRNA がその代表である。mRNA はタンパク質の情報をコードしているので、coding RNA といわれる。ヒトなどの高等生物では、ゲノム内で遺伝子が占める領域は非常に小さく、数%にすぎない。その他の領域は機能を持たないジャンク領域と考えられていた。近年、そのような領域から転写される RNA に機能を持つものが数多く見つかっている。このような機能性 RNA はタンパク質をコードしないので、non-coding RNA (ncRNA) と呼ばれる¹⁾。

non-coding RNA には多くの種類が存在するが、比較的古くから知られていたものに tRNA と rRNA などがある¹⁾。tRNA は transfer RNA のことであり、タンパク質合成の場であるリボソームにアミノ酸を運ぶ働きを持つ。タンパク質の合成に使われるアミノ酸には 20 種類があるので、それぞれに対応した tRNA が存在する。tRNA は mRNA のコドン認識する部位 (アンチコドン) と対応したアミノ酸と結合する部分を持つ。rRNA (ribosomal RNA) はリボソームを構成する RNA である。リボソームの重量のおよそ 60% が rRNA、残りの 40% がタンパク質である。rRNA は細胞内に最も多く存在する RNA である。snRNA (small nuclear RNA) は、スプライシングに関与する RNA である。snoRNA (small nucleolar RNA) は核内の核小体に局在する RNA で、メチル化などの化学修飾に関与する。

miRNA (micro RNA) は、1993 年に発見された RNA である²⁾。近年の研究によって、miRNA は発生、分化、細胞増殖、アポトーシス、代謝パスウェイなどに関与していることが示唆されている^{3,4)}。また、がんとの関連も注目されている³⁾。名前が示すように、miRNA は 22 塩基程度の長さを持つ小さな RNA である。DNA から転写された直後には 100 塩基程度の長さがあるが、数段階の処理を経て成熟した miRNA となる。miRNA は RNA 干渉 (RNAi) という機構を作用機序として、遺伝子の発現を調節 (主に抑制) する。1 種類の miRNA が多種類の遺伝子の発現調節に関わっている。一方、1 種類の遺伝子は複数の miRNA による調節を受ける。このように、miRNA- 遺伝子間には、多対多のネットワークが形成されている。

miRNA と同じように RNA 干渉を通じて遺伝子の発現抑制を行うものに siRNA (short interference RNA) がある。siRNA はほかの non-coding RNA と異なり、DNA から転写されるのではなく、細胞外から移入される。自然界ではウイルス感染の防御機構として働いていると考えられている。近年では、人為的に RNA を細胞に導入し、特定の遺伝子の発現を抑制する方法として利用されている。siRNA は、siRNA と相補な配列を持つ遺伝子の発現を抑制する。

参考文献

- 1) 中村桂子, 松原謙一 (監訳): “細胞生物学 (原書第 4 版):” 南江堂, 2016.
- 2) R.C. Lee, R.L. Feinbaum, and V. Ambros: “The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementary to *lin-14*,” *Cell*, vol.75, pp.843-854, 1993.
- 3) E.A.C. Wiemer: “The role of microRNAs in cancer: no small matter,” *Eur. J. Cancer.*, vol.43, pp.1529-1544, 2007.
- 4) V. Ambros: “The functions of animal microRNAs,” *Nature*, vol.431, pp.350-355, 2004.

S2 群 - 6 編 - 2 章

2-5 エピジェネティクス

(執筆者：福岡 豊)[2018年2月受領]

ヒトゲノム計画の段階ではゲノムの配列が決定できれば、ヒトの生命現象についてすべて説明できると考えられていた。しかし、現在でも不明な点がたくさん残されている。これは、遺伝子配列の以外の情報が生命現象に影響を与えるからである。そのような要因の一つとして、エピジェネティクスが挙げられる^{1,2)}。正確には、エピジェネティクスとは、塩基配列の変化を伴わないが、細胞間で継承される遺伝的背景を研究する分野のことである。ここでは、そのような配列の変化を伴わないが、遺伝子の発現に影響を与える要因を総称するものとする。

全く同じ塩基配列を持つクローン動物でも体表の柄などに違いが生じる。ヒトでも、一卵性双生児において性格の違いが見られる。エピジェネティクスは、このような現象に関係している。エピジェネティクスには、DNA やヒストンの化学修飾をはじめとして、様々なものがある。すべてを網羅することが本稿の目的ではないので、以下では代表的なもののみを説明する。

エピジェネティクスの代表例として、DNA のメチル化が挙げられる。DNA メチル化とは、4 種類の塩基の一つである C (シトシン) を構成する炭素にメチル基が付加されることを指す。この反応は、CG という塩基配列を持つ部位 (これを CpG と書く) で生じやすい。2-1 節で述べたように、真核生物では遺伝子の上流には転写因子が結合する部位がある。一方、遺伝子の上流には、CpG が高密度で存在する CpG アイランドと呼ばれる部位がある。CpG アイランドで DNA のメチル化が生じると、転写因子の結合に影響を与え、遺伝子の発現が低下する。このように、DNA メチル化は、遺伝子発現の調節機構として働いている。がん細胞で DNA メチル化の異常が高頻度で観察されている。異常の一例として過剰なメチル化があり、広範囲な遺伝子の不活性化のため、がん抑制遺伝子が機能しなくなることが知られている。

もう一つの機構であるヒストンの化学修飾には、メチル化、アセチル化、リン酸化、ユビキチン化がある^{3,4,5,6)}。それぞれ、メチル基、アセチル基、リン酸、ユビキチンが付加される反応である。ヒストンを構成する 4 種類のタンパク質は尾部 (アミノ酸の鎖が外部に出ている部分) を持つ。化学修飾は、尾部にある特定のアミノ酸に生じる。例えば、ヒストン H3 (4 種類のタンパク質の 1 種) の N 末端から 4 番目のリシンにメチル基が 1 つ付加されることを H3K4me1 と書く。リシンには最大で 3 つのメチル基が付加される。先のリシンに 3 つのメチル基が付加される場合、H3K4me3 と書く。一方、アセチル化は H3K9ac のように書く。

このような化学修飾によって、クロマチンがユークロマチンからヘテロクロマチンに変化したり、逆の変化をしたりする。これをクロマチンリモデリングという。ヘテロクロマチンでは、ヌクレオソームが凝集しているので、遺伝子が発現しにくい。ユークロマチンの状態では、ヌクレオソームの間隔が開いており、転写因子などが結合できるので、遺伝子発現が生じる。このように、ヒストンの化学修飾の状態によって、遺伝子発現が調節されている。このような化学修飾の状態をヒストンコードと呼ぶ。

参考文献

- 1) 中村桂子, 松原謙一 (監訳): “細胞生物学 (原書第 4 版),” 南江堂, 2016.

- 2) 押村光雄, 伊藤 敬 (編) : “ 発生・細胞分化を決定するエピジェネティクスと遺伝子発現制御, ” 実験医学, vol.21, no.11, 2016.
- 3) 堀越正美 (編) : “ クロマチンと遺伝子機能制御, ” シュプリンガー・フェアラーク東京, 2003.
- 4) 牛島俊和, 眞貝洋一 : “ エピジェネティクスキーワード事典, ” 羊土社, 2013.
- 5) 井出利憲 : “ 分子生物学講義中継 Part 1, ” 羊土社, 2002.
- 6) 井出利憲 : “ 分子生物学講義中継 Part 3, ” 羊土社, 2004.

S2 群 - 6 編 - 2 章

2-6 進化とゲノム

(執筆者：福岡 豊)[2018年2月受領]

生物の進化に関しては、ダーウィンの進化論が有名である。ダーウィンの進化論は、突然変異と自然選択が基礎となっている¹⁾。ダーウィンの時代には、遺伝子の実態は明らかではなかったため、ゲノム配列との関係は論じられていない。遺伝子の実態が DNA 上の遺伝子であることが明らかになった後、遺伝子突然変異の大半は自然選択にとって有利でも不利でもないとする分子進化の中立説が唱えられた²⁾。

2-3 節で述べたように、遺伝子突然変異にはアミノ酸の配列に変化を生じさせないものがある(同義置換)。このような変異は、当然ながら中立な変異である。分子進化の中立説では、生物にはたえず遺伝子突然変異が生じるが、大半は中立的な変異であり、ある生物のゲノムにはそのような変異が蓄積されていると考える。実際に同義置換と非同義置換の生じる速度を推定してみると、前者のほうが大きいという結果となる。一方で、遺伝子突然変異のなかには、自然淘汰に不利なものも生じるが、それらが集団に広がる速度は、中立な変異に比べると著しく遅いと考えられる。

生物の進化とゲノムは関連が深く、進化的に近い生物種ほど、ゲノム配列の相違点が少ない。例えば、ヒトとチンパンジーを比べた場合、ヒトとマウスよりも相違点は少ない。逆に、ゲノム配列の相違から生物の進化を探ることができる。その場合、ゲノム全体を比較するのが困難な場合もあるので、特定の遺伝子の配列を比較することが多い。

系統樹とは、複数の生物間の類縁関係を樹形図として表現したものである。枝の長さが短いほど、2つの生物が類似していることを表している。従来は、生物の解剖学的構造(脊椎の有無)などの表現型を用いて、類縁関係が調べられてきた。近年では、ゲノム(あるいは特定の遺伝子)の配列の類似性を用いて、系統樹が描かれる。このようにして描いた系統樹を分子系統樹という。分子進化の中立説は分子系統樹の基礎となっている。

現在、重要な機能を持つ部位の塩基配列は、様々な生物種において保存されていると考えられている。機能にとって重要でない部位では、突然変異が起きても、自然淘汰に影響がないと仮定しているので、分子進化の中立説に基づく考えである。

参考文献

- 1) C.R. Darwin, 八杉龍一(訳):“種の起源<上><下>,”岩波文庫,1990.
- 2) M. Kimura :“The Neutral Theory of Molecular Evolution,” Cambridge University Press, 1983. (邦訳 “分子進化の中立説,” 紀伊國屋書店, 1986.)