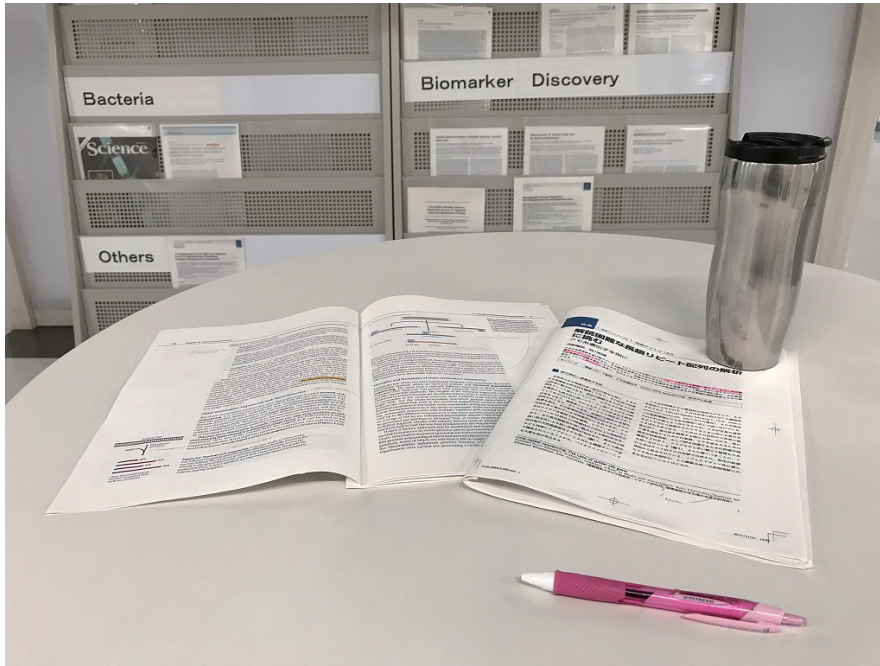


# ゲノム配列の読み取りと解読

## Genome Sequencing and Interpretation



慶應義塾大学先端生命科学研究所・ゲノム分子生物学  
Genomic Molecular Biology, Institute for Advanced Biosciences, Keio University

2024年11月20日

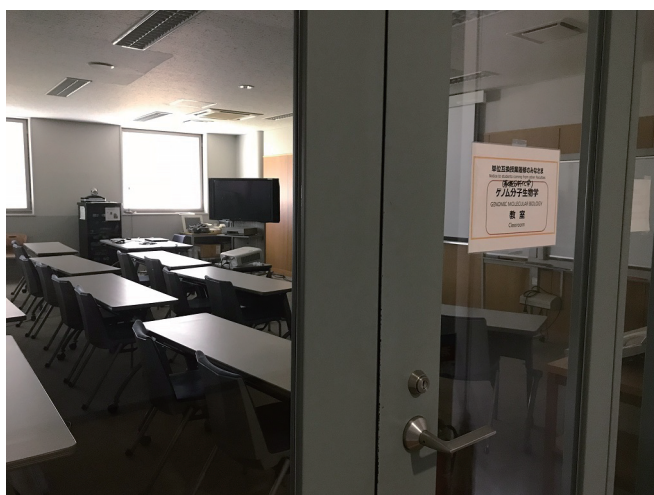
## 本テキストを利用するに当たっての注意

本テキストは慶應義塾大学鶴岡タウンキャンパス (TTCK) で行われているゲノム分子生物学 (環境情報系) の履修者有志 (慶應義塾大学、鶴岡工業高等専門学校または山形大学の在學生) によって作成されたものです<sup>1</sup>。生物学の専門家によって作成されたものではありませんので、重要なポイントについては元となった書籍”Genomes 4”[1] でしっかり確認して下さい。

ただ学生が作ったものとはいえ、独自に作成した豊富な図を用いて初学者の視点から分子生物学を分かりやすく解説することを試んでいます。

また本テキストはウィキ (Wiki) 方式で毎年履修者によって改良され、どんどん良い物になってゆくでしょう。本テキストを受動的に利用するのではなく、ぜひ執筆活動に積極的に参加する姿勢で利用して頂きたいと思います。

齋藤 輪太郎



---

<sup>1</sup>一部、ChatGPT や Gemini 等の生成 AI を利用しています。

## 目次

1	ゲノム配列を読み取る意義	5
2	DNA と RNA の化学的構造	8
3	DNA の複製メカニズム	10
4	遺伝子地図の作成法	13
5	ゲノム配列を読み取る技術	16
5.1	技術の変遷	16
5.2	Chain-termination sequencing	16
5.3	次世代シーケンサー	20
5.3.1	次世代シーケンサー用のライブラリ調整	20
5.3.2	イルミナーシーケンシング	20
5.3.3	SOLiD	20
5.3.4	PacBio シーケンシング	26
5.3.5	ナノポアシーケンシング	26
5.4	反復配列の読み取り	29
5.4.1	ゲノムに存在する反復配列	29
5.4.2	反復配列を読み取る際の問題	29
6	ゲノム配列の機能解析法	31
6.1	ゲノム配列中の遺伝子領域の探索	31
6.1.1	Open Reading Frame (ORF) の探索	31
6.1.2	相同性検索を用いた既知の転写産物情報・タンパク質情報の活用	32
6.1.3	確率モデルの活用	33
6.2	遺伝子の機能阻害による機能解析	33
6.2.1	相同組換えによる遺伝子のノックアウト	33
6.2.2	RNAi を用いた遺伝子のノックダウン	34
6.2.3	制御可能な Cas9 エンドヌクレアーゼを用いた遺伝子の改変	34
7	転写産物の解析法	38
7.1	遺伝子発現解析の意義	38
7.2	プロモーターによる転写開始の仕組み	39
7.3	エンハンサーによる転写制御	39
7.4	マイクロアレイ	40
7.5	RNA-seq(内容確認・校正中)	42
7.6	CAGE 法	42

<b>8</b>	<b>タンパク質解析法</b>	<b>43</b>
8.1	細胞内に存在するタンパク質を解析する意義	43
8.2	SDS ポリアクリルアミドゲル電気泳動 (SDS-PAGE)	43
8.3	クロマトグラフィー	45
8.4	質量分析	46
8.4.1	概要	46
8.4.2	MALDI(マトリックス支援レーザー脱離イオン化) 法	48
8.4.3	MS/MS 分析	48
8.4.4	四重極型質量分析計	49
8.4.5	トップダウン法とボトムアップ法	50
8.4.6	全長タンパク質のままでも質量分析は可能?	50
8.5	免疫染色	51
<b>9</b>	<b>細胞内分子間相互作用</b>	<b>52</b>
<b>10</b>	<b>分子間相互作用解析法</b>	<b>55</b>
10.1	実験的方法	55
10.1.1	アフィニティークロマトグラフィー (内容確認・校正中)	55
10.1.2	共免疫沈降法 (内容確認・校正中)	58
10.1.3	酵母ツーハイブリッド法 (内容確認・校正中)	58
10.1.4	ファージディスプレイ法	58
10.2	情報科学を用いた方法	59
10.2.1	遺伝子の発現相関 (内容確認・校正中)	59
10.2.2	系統プロファイル法	59
10.2.3	オペロン解析 (内容確認・校正中)	61
<b>11</b>	<b>ゲノム科学の進歩が生活・社会に与える影響</b>	<b>61</b>



# 1 ゲノム配列を読み取る意義

まずは、基本的な用語とその意味を眺めてみよう:

**DNA** デオキシリボ核酸と呼ばれる生体内の巨大分子。地球上のほとんどの生物の遺伝情報がここにコードされている。遺伝情報の継承と発現を担っている。

**遺伝子** 遺伝情報の単位。

**ゲノム** 遺伝情報の総体。

その上で、ゲノム配列を読み取る意義は何だろうか。何故、20世紀の終わりより様々な種のゲノム配列の読み取りが進められたのだろうか。特にヒトを含めた特定の生物の「全」ゲノム配列を読み取る意義は何だろうか。

いろいろな考え方ができるが、おおざっぱに言えば、対象となる生物がどのような遺伝情報に基づいて生命現象を引き起こしているのかという分子生物学的特性を理解するとともに、その生物がどのように進化してきたかを知る事ではないだろうか。

医学的な応用を考えるとその意義が分かりやすい。ヒトゲノムを読み読み取った上で、疾患の原因となる遺伝子を特定することにより、各個人に対して遺伝性の疾患のリスク評価が可能になる。また各個人の遺伝子型に合った治療法の選択すなわち、**個別化医療**が発展する。さらに、疾患の原因遺伝子が生体内でどのように機能するのかを理解することによって、それに基づいた新しい治療法や薬剤の開発につながる。

また身近なところで、食料生産に対する応用も考えられる。野菜や穀物、果物を含む食用の植物や、食用の家畜のゲノム配列に基づいた遺伝子解析を通じて、生産性・耐病性・栄養価・味等に関連した遺伝子マーカーの開発が可能となり、より質の高い食用植物や家畜の選択・育種・品種改良ができるようになる。

一方で、特に学術的には、ゲノム配列の解析は、生物種間の進化的関連性を明らかにするための貴重な手段であり、複数種間で共有されている遺伝子の比較により、生物種の類縁関係や共通の祖先を推定することができる。ヒトと同じ霊長類であり、高度な知能を持つ種であるチンパンジーとヒトのゲノム配列の違いは、計算方法にも依存するが、2%に満たないと言われている [3]。従ってヒトとチンパンジーは極めて近縁種であることが伺える。

ヒトより進化的にもっと離れていると思われるメダカに目を向けると、メダカの特定の染色体(LG22)上には633個の遺伝子が存在し、そのうち588個が人間と共通の働きを持った遺伝子と推定されている [4]。さらに、ヒトとメダカの間で、同じ機能を持つと思われる遺伝子の並び順が同一に保たれているゲノム領域が多数発見されており、ヒトとメダカの共通の祖先の痕跡であると考えると興味深い。もっと言うてしまえば、ヒトと大腸菌は遙か昔、共通の祖先から進化的に分かれてきた、という痕跡も指摘されているくらいである [5]。

このように、ゲノム配列の読み取りは学術的にも実用的にも大きな可能性を秘めている。それを念頭に本テキストではゲノム科学について学習し、その活用法を考えていきたい。

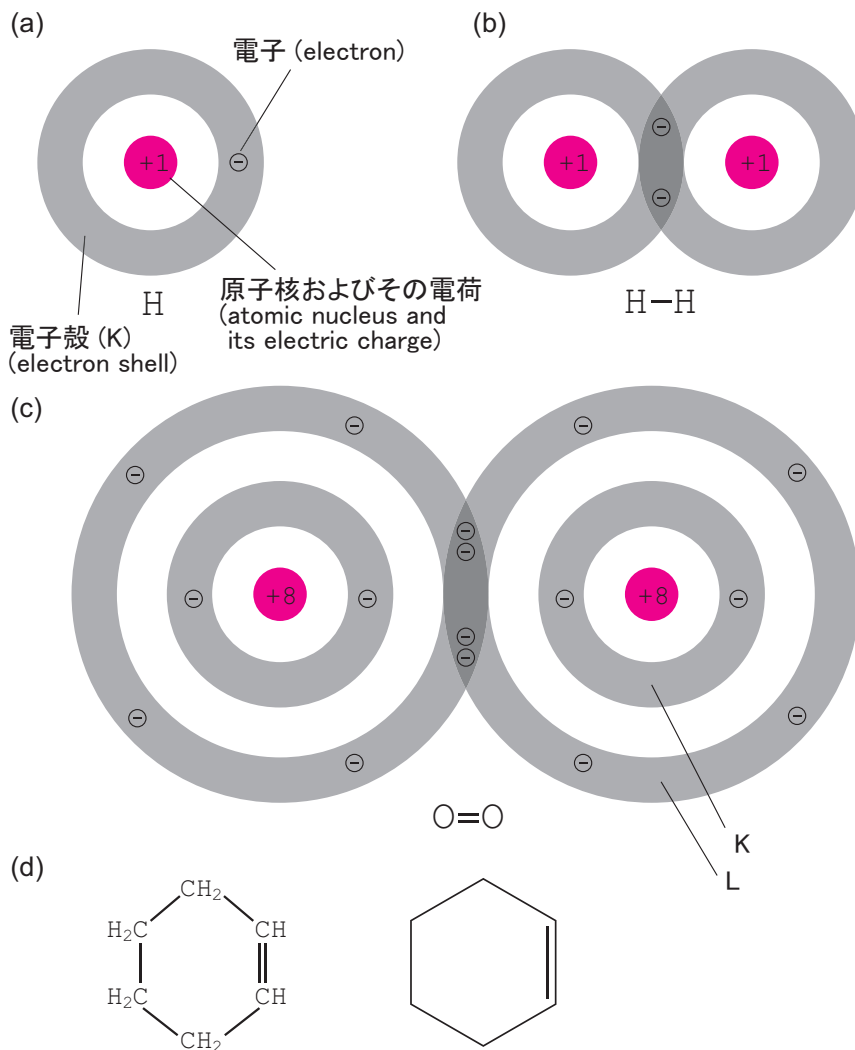


図 1: 共有結合の例

(a) 水素原子 (H) の模式図。電荷が+1の原子核周辺に電荷が-1の電子が存在するため、全体の電荷は0となる。但し、原子核に最も近い電子殻 (K 殻) は本来2個の電子を収納可能なため、1つ”空席”が存在することになる。(b) 水素分子 (H<sub>2</sub>)。原子同士が単結合で分子を作っている。この場合、1対の電子を共有する (単結合) ことで、K 殻には2個の電子が入り、かつ全体の電荷が0となって安定する。構造式 (structural formula) では、H-H のように単結合を表す。(c) 酸素分子 (O<sub>2</sub>) で見られる二重結合。構造式では、O=O のように二重結合を表す。K 殻の1つ外側の L 殻には8個の電子を収納可能で、2対の電子を共有することで、電荷が0になり、かつ電子殻が埋まって安定する。(d) シクロヘキセン (cyclohexene) の構造式。環状となっている。右のように、C と H はしばしば省略される。

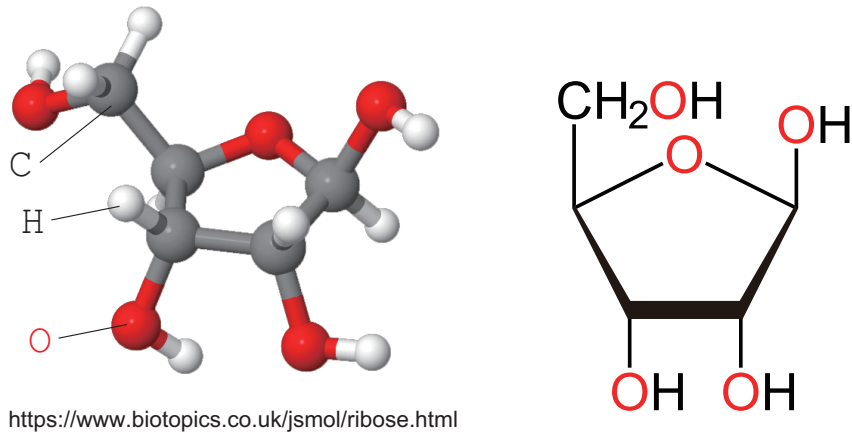


図 2: ハース投影式 (Haworth projection)

リボース (ribose) の立体構造模型 (左) および対応するハース投影式。太い線の部分が、読者に近い方の結合を表す。

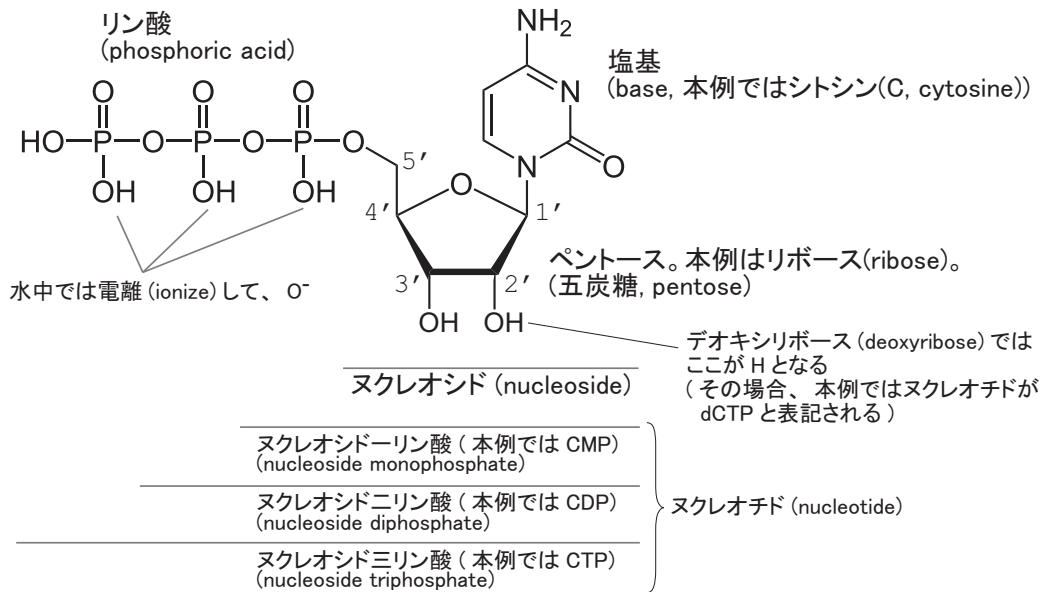


図 3: ヌクレオチドの構造の概略

CTP の例を示す。複数の CTP が繋がって核酸になるときは、外側の 2 つのリン酸が外れる。図で示した 3 箇所のヒドロキシ基 (OH) は水中 (中性条件) では電離して  $H^+$  が外れて  $O^-$  となる。

## 2 DNA と RNA の化学的構造

DNA (deoxyribonucleic acid, デオキシリボ核酸) も RNA (ribonucleic acid, リボ核酸) も核酸 (nucleic acid) という化学物質でできている。ここで本題に入る前に、ごく簡単に化学の基礎を復習しておこう。原子 (atom) は、正の電荷を帯びた原子核 (atomic nucleus) と、それを取り囲む電子殻 (electron shell) の軌道に存在する、負の電荷を帯びた電子 (electron) でできている。図 1a に水素原子 (H) の例を示そう。原子核はさらに陽子 (proton) と電氣的に中性な中性子 (neutron) から構成される。

原子やイオン (ion) が集まって分子や結晶を作る際にできる原子やイオンの結びつきは化学結合 (chemical bond) と呼ばれ、分子内の化学結合には、

1. イオン結合 (ionic bond)
2. 共有結合 (covalent bond)
3. 金属結合 (metallic bond)

の 3 種類がある。

核酸に含まれる多くの原子は共有結合でつながっている。共有結合は、原子間での電子対の共有をとる化学結合である。例えば、2つの水素原子は図 1b に示すように、1 対の電子対を共有する単結合 (single bond) で水素分子 ( $H_2$ ) となる。これを H-H という構造式 (structural formula) で表す。酸素分子中の 2つの酸素原子 (O) は 2 対の電子対を共有する二重結合 (double bond) でつながって酸素分子 ( $O_2$ ) を形成している (図 1c)。これを  $O=O$  で表す。化合物の骨格を表す構造式では、図 1d のように、骨格中の炭素 (C) およびそれに結合している水素 (H) がしばしば省略されることに注意されたい。多くの場合、実際の化学構造は平面上に収まる立体構造にはなっていない。例えば、リボース (ribose) の立体構造の例を図 2(左) に示す。この図に示された向きで、例えば、五炭糖構造の上部に突き出ているヒドロキシ基 (-OH, hydroxy group, 水酸基) と、下部より結合しているものがあることが観測できる。これを表現するために、図 2(右) のハース投影式 (Haworth projection) 等が使用される。

さて、遺伝子は DNA にコードされており、合成されるタンパク質は DNA の塩基配列に由来する。RNA は DNA を鋳型にして合成され (transcription, 転写)、リボソームにて RNA を鋳型にしてタンパク質が合成される (translation, 翻訳)。DNA は通常、反平行 (antiparallel) 二本鎖の状態が存在し、二重螺旋構造をとる。一方、主に一本鎖の RNA は DNA と比較して構造が不安定である。水素結合 (hydrogen bond) は DNA の二本鎖を繋ぐ際に塩基同士の結合に使われている (図 4)。水素結合は酸素原子と水素原子の極性が大きく異なることを利用し、水素分子の正電荷に酸素原子の負電荷が近づくことによって電氣的な引力が働くことのできる弱い結合である。しかし、これらが複数存在すると大きな結合力を持つようになる。

核酸はヌクレオチド (nucleotide) と呼ばれる化学物質が連なった構造をしており、図 3 に示すように、ヌクレオチドは塩基、糖 (この 2つが結合した構造はヌクレオシド (nucleoside) と呼ばれる) およびリン酸から構成される。2つのヌクレオチドは、一方のヌクレオチドの 3' のヒドロキシ基 (-OH) ともう一方のヌクレオチドのリン酸がホスホジエステル結合 (phosphodiester bond, リン酸ジエス

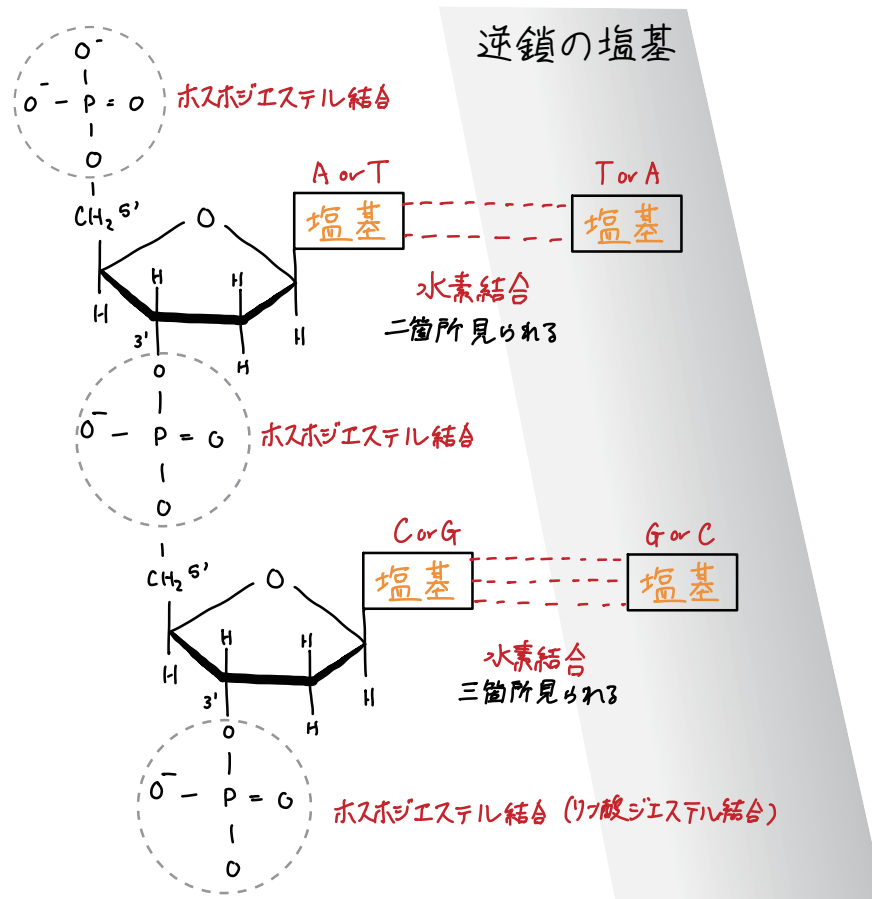


図 4: DNA 分子中の化学結合

一本の DNA 鎖中のヌクレオチドはリン酸を介して炭素同士がホスホジエステル結合で繋がっている。DNA の 2 つの鎖を繋ぐのはそれぞれの鎖のアデニン (A) とチミン (T)、または、シトシン (C) とグアニン (G) の間の水素結合である。アデニン (A) とチミングアニン (T) の水素結合箇所は 2 つだが、シトシン (C) とグアニン (G) のそれは 3 つあり、後者の塩基対の多い DNA 配列は壊れにくい。

テル結合)を介して繋がっている(図4)。ホスホジエステル結合は炭素原子間をリン酸が介した2つのエステル結合<sup>2</sup>によって強く共有結合している化学結合である。

RNAは糖の部分にリボースを使用しており、糖の2'の部分はヒドロキシ基(-OH)になっている。一方DNAでは、糖の部分にデオキシリボースを使用しており、糖の2'の部分は-Hになっている(酸素が取れているという意味でデオキシ(deoxy)と呼ばれる)。DNAの塩基はA(adenine, アデニン), C(cytosine, シトシン), G(guanine, グアニン), T(thymine, チミン)だが、RNAでは、A, C, G, U(uracil)の4種類である。

ヌクレオチドを構成するリン酸基の数は、1個、2個、そして3個の場合がある。1個をM(monoの略)、2個をD(diの略)、3をT(triの略)と表すことがある。例えば、リボヌクレオチドはしばしば塩基の種類 + リン酸基の数 + Pで表されるが、アデニン + 糖(この2つでアデノシン, adenosine) + 3つのリン酸のヌクレオチドは、ATPと表される(ATPは生体のエネルギー通貨。2つの高エネルギーリン酸結合を含む。)同様にCTPも存在しており、CDPもしくはCMPに加水分解されることでエネルギーを放出する。デオキシヌクレオチドの場合、リン酸基の数によって、“dCMP”, “dCDP”, “dCTP”(それぞれ1~3個の場合)のように表される。

### 3 DNAの複製メカニズム

DNAの複製は、細胞分裂の引き金となっていて、細胞の分裂を調節している。分子レベルの視点で見れば、図5に示すように、これは鋳型DNA配列に従って、正しいヌクレオチドを繋げてDNAを伸長していく反応である。一方、ゲノムレベルで考えると、この複製の仕組みは

1. 半保存的複製
2. 複製フォークでの反応過程
3. 染色体への分子構築

の3つに分けることができる。DNAの複製では、二重らせん構造のうちの片方のDNA鎖が鋳型となり、AにはT、CにはGなど塩基対がなされて複製される(図6)。この複製のことを半保存的複製を言う。DNA複製が進行中の状態は、形状がフォークのように2つに分かれるため、複製フォークとよばれる。この複製フォークでは、新しいDNA鎖を合成するDNAポリメラーゼという酵素とこれを働かせるプライマーが合成し、複製が行われる。DNAポリメラーゼはDNA鎖の5'-端から3'-端の方向に向けて鎖を伸長させるが、複製フォークの進行方向に対して、DNAは二重鎖であるため5' → 3'のものと3' → 5'になるものができてしまう。5' → 3'は連続的に複製できるが3' → 5'は連続して伸長できなくなる。そのため、連続的複製と不連続的複製の2つに分けられる。連続的複製でできた申請書をリーディング鎖(先行鎖)といい、不連続的複製でできた新生鎖をラギング鎖(遅延鎖)という。複製されたDNAは染色体に凝縮されたのちに2つの娘細胞に分配され圧縮される。DNAはタンパク質と結合したクロマチンとなっており、そこにヒストンというたんぱく質が巻き付いている。ヒストンが結合した単位をヌクレオソームといい、これが数珠上に連なって

<sup>2</sup>アルコール(炭化水素のHが-OHに置き換わっている)とカルボン酸(少なくとも一つのカルボキシ基(-COOH)を有する有機酸(酸性を示す有機化合物))の脱水(H<sub>2</sub>O)縮合で生じる結合。

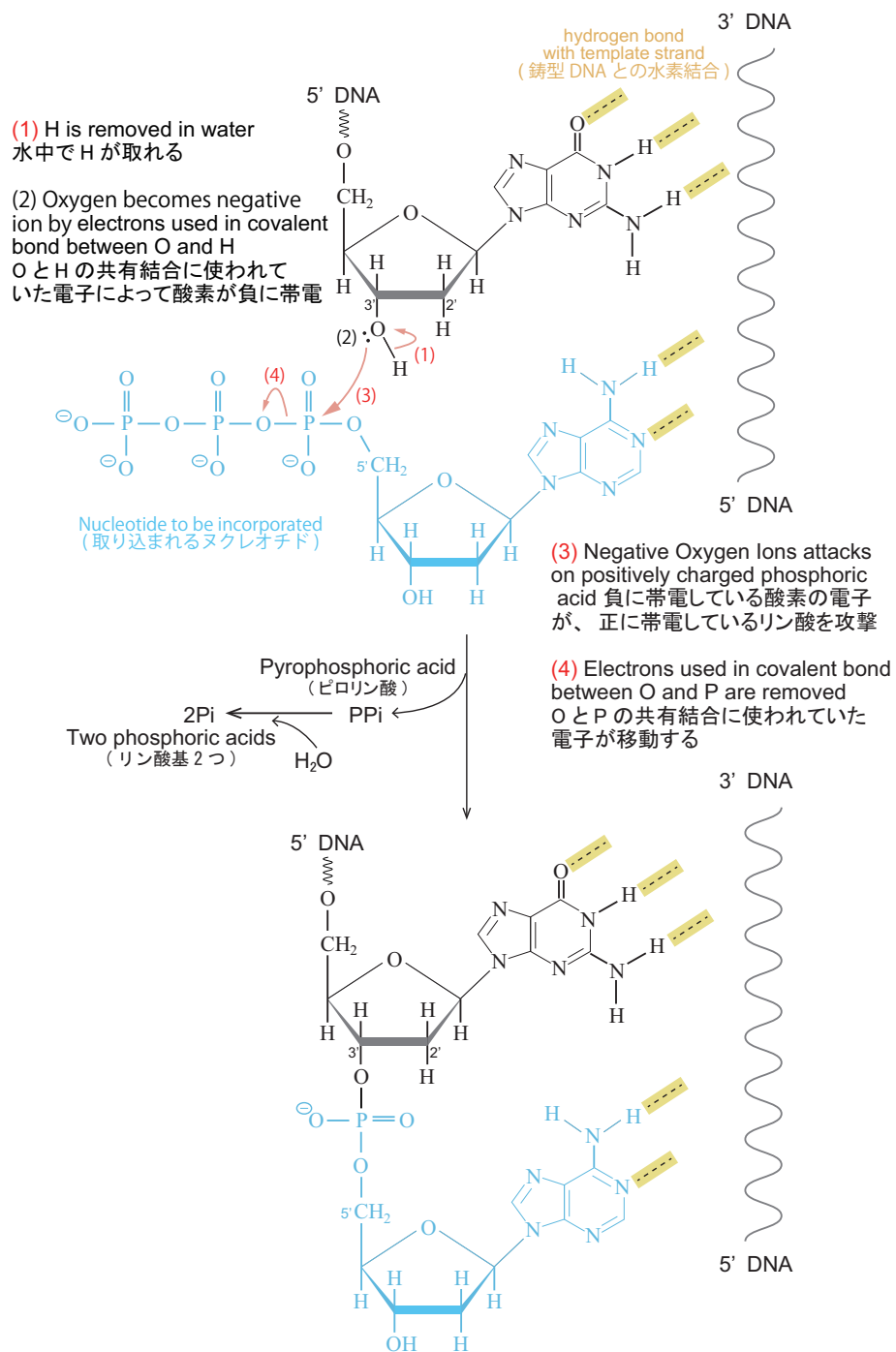


図 5: DNA 複製時に起きる DNA 伸長反応

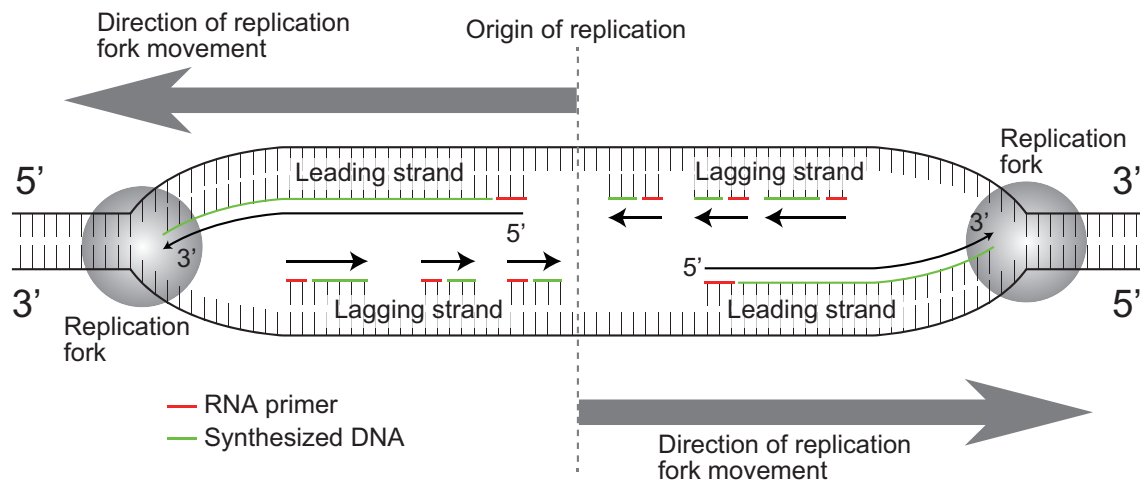


図 6: DNA 複製の仕組み

いる。この細かい折りたたまれ方がどのようにして染色体になっているかはわかっていないが、染色体の形成によって遺伝情報は伝達されている。

DNA の複製時には、らせん構造の二本鎖 DNA を離す必要があるが、図 7 の絡まった 2 本のコードに例えたように、単純に両側に引き離そうとしても、他の部分に張力 (tension) がかかり、一定以上に引き離すことができず、またその張力によって、DNA 分子全体に歪みが生じ、超らせん (DNA supercoil) を形成してしまう。DNA トポイソメラーゼ (DNA topoisomerase) は、複製時や転写時に鎖を切断・再結合することによって DNA のトポロジーを制御し、DNA 分子全体の歪みを解消する。DNA トポイソメラーゼには、I 型と II 型があり、I 型は片方の鎖だけ、II 型は両方の鎖を切断する。抗がん剤には、トポイソメラーゼを阻害することによって効果を発揮するものもある。

真核生物の染色体の末端部にはテロメア (telomere)<sup>3</sup> という特定の反復塩基配列と、それを保護するタンパク質 Shelterin からなる構造があり、DNA の複製時に DNA が損傷したりするのを防ぐ働きがある。哺乳類ではテロメアに TTAGGG という塩基配列が繰り返し並んでおり、ヒトの場合 1 万回以上繰り返されている。テロメアは細胞分裂のたびに短くなる。1 回の細胞分裂でこれが 25 ~ 200 塩基ずつ程減り、計約 5000 塩基減ると細胞が分裂できなくなる。また老人に比べて若者の方が 1.5 倍ほどテロメアの長さが長いことから生命の寿命にも関与しているのではないかと考えられている。

ただ、テロメアが短くなってもテロメラーゼという酵素が活性化していれば塩基配列を足してテロメアを長くすることができる。一部のがん細胞では、活性を持ったテロメラーゼが大量に存在しているため、無限に増殖を繰り返す。このため現在はテロメラーゼを不活性化する癌治療薬の開発が進められている。

<sup>3</sup>語源はギリシア語の"telos" (終わり) と"meros" (部分)



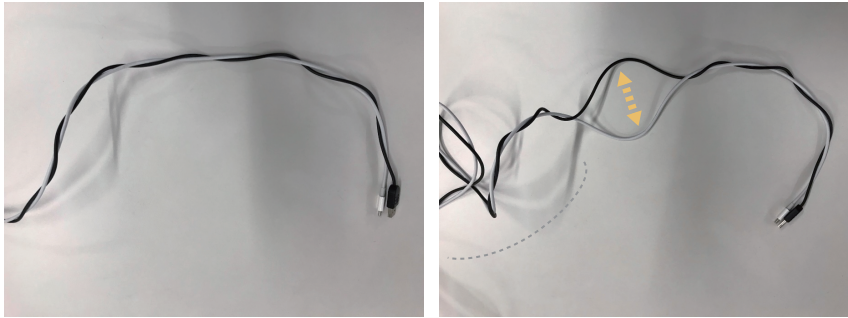


図 7: 絡まった 2 本のコードを引き離すときの問題

単純に両側に引っ張るだけでは、右図のように、引き離せる範囲に限界があり、引き離した箇所両側に大きな力がかかってしまう (右図で左側が引っ張ったときの力で曲がってしまっていることに注意)。

## 4 遺伝子地図の作成法

染色体上で各遺伝子がどのように配置されているかを示す情報を遺伝子地図 (Gene map) という。遺伝子地図には、大きく分けて、遺伝子の組換え率から作成される連鎖地図 (Linkage map, Genetic map, 遺伝地図) と、DNA 分子を直接実験的に調べて作成される物理地図 (Physical map) がある。ここでは前者について詳しく触れよう。

精子、卵子のような配偶子 (Gamete) は減数分裂 (Meiosis) と呼ばれる細胞分裂を経て誕生するが、このとき、第 1 減数分裂前期 (Prophase I) に、2 つ 1 組の相同染色体間で領域の一部の交換がしばしば観測される (図 8)[7]。これを相同組換え (Homologous recombination)、あるいは略して、組換え (Recombination) という [2, 6]。

染色体上のどこの位置でも等しい確率で、尚且つランダムに組換えが起こると仮定した場合、1 本の染色体上の 2 つの遺伝子座の位置が離れているほど組換えによって別々の染色体に分離する確率は高くなる。つまり、同一染色体上の 2 つの遺伝子の組換え頻度 (Recombination frequency) は、それらの遺伝子の染色体上の距離と比例関係にあると予想できる (図 8d)。この考え方を利用すると、多数の交配実験を行い、染色体上の様々な遺伝子座の組み合わせに対して組換え頻度を算出することで、染色体上の遺伝子間の相対的な位置を示した地図を作成する事ができる (図 9)。2 つの遺伝子間の組換え頻度 1% を 1 地図単位 (map unit), もしくは 1 cM (centimorgan) と表記する。

さて、染色体上のどこの位置でも等しい確率で組換えが起こるという仮定は、実は必ずしも成立しない。例えば、組換えホットスポットと呼ばれる領域は他の領域と比べ組換えの起こる頻度が高いことが分かっている [8]。このことは、2 つのマーカー間の遺伝地図上の距離は物理地図上のそれとは必ずしも一致しないことを意味している。

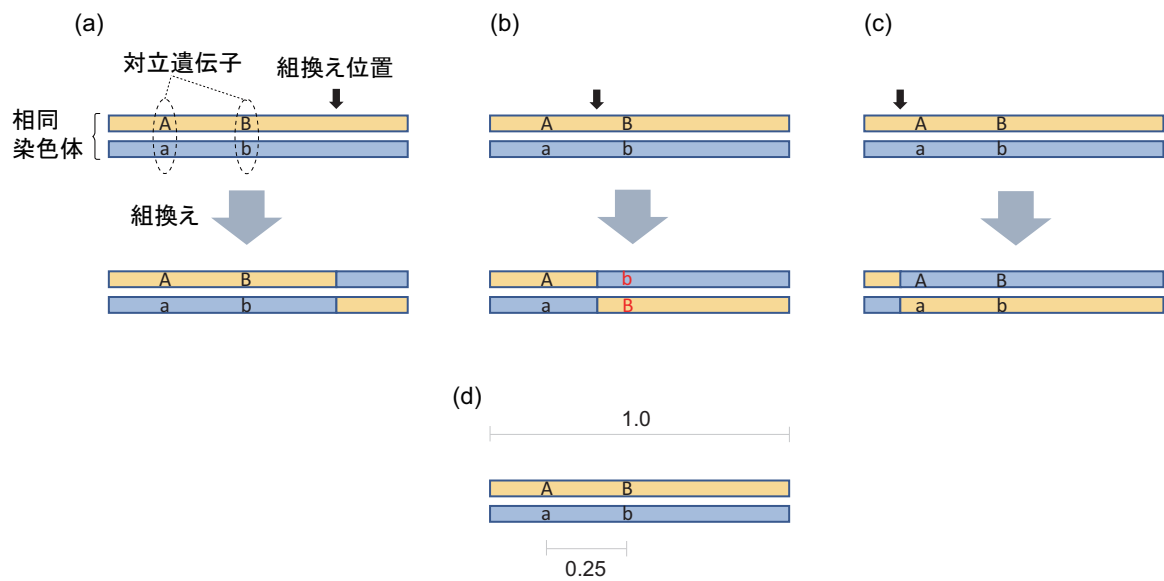


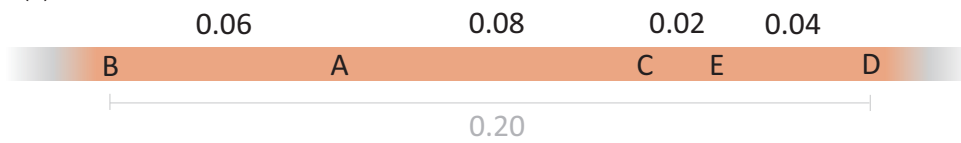
図 8: 染色体上の組換え位置と遺伝子型の組み合わせとの関係

(a), (c) ともに、相同染色体上に載っている遺伝子型はそれぞれ AB と ab だが、(b) のように遺伝子 A と B の間に組換えが起こると、Ab, aB という新しい組み合わせの遺伝子型が出現する。(d) 染色体の全長を 1、A-B 間の距離をその  $1/4=0.25$  とすると、仮に染色体のどこかで一箇所組換えが起こり、その場所は染色体上で完全にランダムだとすると、A-B の間で組換えが起こる確率は 0.25 となる。

(a)

	B	C	D	E
A	0.06	0.08	0.14	0.10
B		0.14	0.20	0.16
C			0.06	0.02
D				0.04

(b)



(c)

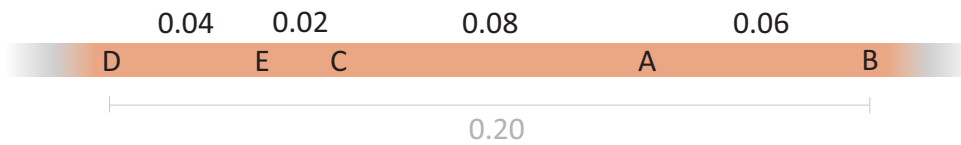


図 9: 組換え率からの連鎖地図の作成例

染色体上の遺伝子間距離と、組換え率が比例していることを前提として連鎖地図を作成する。(a) 遺伝子 A~E の間の組換え率の仮想例。(b) (a) の表から推定される連鎖地図。(c) (b) の地図を逆向きにしたもの。

## 5 ゲノム配列を読み取る技術

### 5.1 技術の変遷

全ゲノム配列のシーケンシングは、ゲノム配列中に存在する全ての塩基の順番を決定することと言い換えることができる。全ゲノム RNA 配列が読み取られた最初の例は、1976 年のバクテリオファージ MS2 (RNA ウィルス, 3,569 塩基) である [58]。全ゲノム DNA 配列が読み取られた最初の例は、1977 年のバクテリオファージ Phi X 174 (一本鎖 DNA ウィルス, 5,375 塩基) であり、これはフレデリック・サンガー (Fredrick Sanger) らによって達成された [59]。この時に使用された方法は、chain-termination 法 (サンガー法, Sanger sequencing, 節 5.2 参照) である。同時期にウォルター・ギルバート (Walter Gilbert) も独自のシーケンシング法を開発している [60]。これらのシーケンシング法はしばしば第一世代シーケンシング法 (first-generation sequencing) と呼ばれる。その後改良が進められ、様々な生物の全ゲノム配列の決定に貢献した [62]。1995 年にはバクテリア (真正細菌) では初となるインフルエンザ菌 (*Haemophilus influenzae*) の全ゲノム配列の読み取りが終了し [63]、1996 年には真核生物で初となる酵母菌 (*Saccharomyces cerevisiae*) の全ゲノム配列の読み取りが終了した。2001 年にはヒトゲノム配列のドラフト版が得られた [65, 66]。

ちなみに 2019 年末より、新型コロナウイルス (SARS-CoV-2, Severe acute respiratory syndrome coronavirus 2) の爆発的感染 (pandemic) により新型コロナウイルス感染症 (COVID-19, Coronavirus disease 2019) が世界的に流行したが、その全ゲノム配列 (コラム「新型コロナウイルスのゲノム」参照) は数ヶ月以内に決定・解析・公開がなされている [28]。

一方、1994 年頃より DNA 配列を合成しながら読み取る技術やスライド上で DNA を増幅する技術等、後に次世代シーケンシング法 (next-generation sequencing)、あるいは第二世代シーケンシング法 (second generation sequencing) と呼ばれる方法の技術基盤が開発され、発表された。次世代シーケンシング法は、数千、数百万の DNA 断片の配列を増幅して同一配列クラスターを形成させ、単一の手順 (run) でそれらを並行して読み取る方法である [70] (節 5.3 参照)。2005 年には 454 Life Sciences 社より次世代シーケンサーのプラットフォーム第一号 GS-20 がリリースされた [61]。その後、第三世代、第四世代のシーケンサーが開発されている。

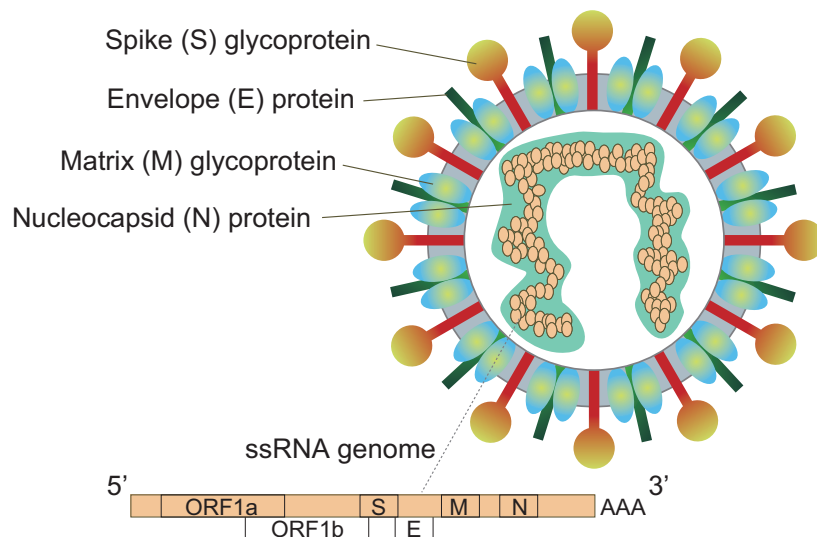
### 5.2 Chain-termination sequencing

通常 DNA が 1 塩基伸長するときは、図 5 に示すように、DNA の 3'-端のヒドロキシ基 (hydroxy group, -OH) が関わる反応を利用して、新しい塩基となる dNTP の 5'-端のリン酸基が結合する。従って、dNTP の 3'-端が脱酸素化した分子である ddNTP が DNA ポリマーの末端に結合すると、それ以上の伸長は阻害される (図 10)。ddNTP として ddATP, ddCTP, ddGTP, ddTTP の 4 種類を用意し、それぞれを異なる蛍光色素で標識し、ACGT の 4 種の塩基の区別をつけた状態で、鋳型 DNA と DNA ポリメラーゼ<sup>4</sup>、そして dNTP、ddNTP、プライマーを用いて Thermal cycle sequencing

<sup>4</sup> *Taq* DNA ポリメラーゼがよく用いられる。このポリメラーゼが活性化する最適の温度は、72 °C と言われているが、94 °C まで引き上げても活性がしばらく失われず、熱耐性を持つ。また、DNA の伸長反応を長く続けることができ、DNA 末端を分解してしまうようなことも少ないという特徴を持つ [1]。この酵素は *Thermus aquaticus* と呼ばれる極めて高温の環境で生息する細菌から得たものである。

### 新型コロナウイルスのゲノム

新型コロナウイルス (SARS-CoV-2) は約 3 万塩基の一本鎖 RNA をゲノムとし、エンベロープ (envelope, 脂質二重膜になっている) を持つウイルスであり [29]、従来のコロナウイルスである SARS-CoV と 79%、MERS-CoV と 50% のゲノム配列の相同性を持つ [32]。ゲノム配列中には 15 箇所 の ORF (節 6.1.1 参照) が 見つかっており、それらは合わせて 29 種類 のタンパク質をコードしている [31]。5' 端側最上流には、ウイルスの複製やタンパク質の分断等に関わるウイルス非構造タンパク質 (viral nonstructural protein, ウィルス粒子の一部ではないタンパク質) をコードする ORF (ORF1a/ORF1b) があり、ORF1a は非構造タンパク質 NSP1~NSP11 をコードし、ORF1ab は NSP12~16 をコードする。その下流には 5' から 3' の順に、スパイク (S)、エンベロープ (E)、膜 (M)、ヌクレオカプシド (N) の構造タンパク質 (viral structural protein, ウィルス粒子の一部となるタンパク質) をコードする ORF が並ぶ。また、構造タンパク質の ORF の間には、SARS-Cov-2 に特徴的であるとされている ORF8[30] を含むアクセサリタンパク質 (accessory proteins, ウィルスの遺伝子発現制御や宿主細胞の機能制御等、ウィルスの生存・増殖にとって間接的な役割を持つタンパク質) をコードする 9 つの ORF が点在している [31]。SARS-Cov-2 はしばしば変異が起こっているが、現在世界中で 1 万以上の塩基置換が認識および報告されており、予想される変異率は細胞感染 1 回につき、1 ヌクレオチド部位あたり 0.0001~0.01 置換の範囲であることが分かっている [31]。これらの変異は、ウィルスの様々なゲノム構造の個所に分布している。



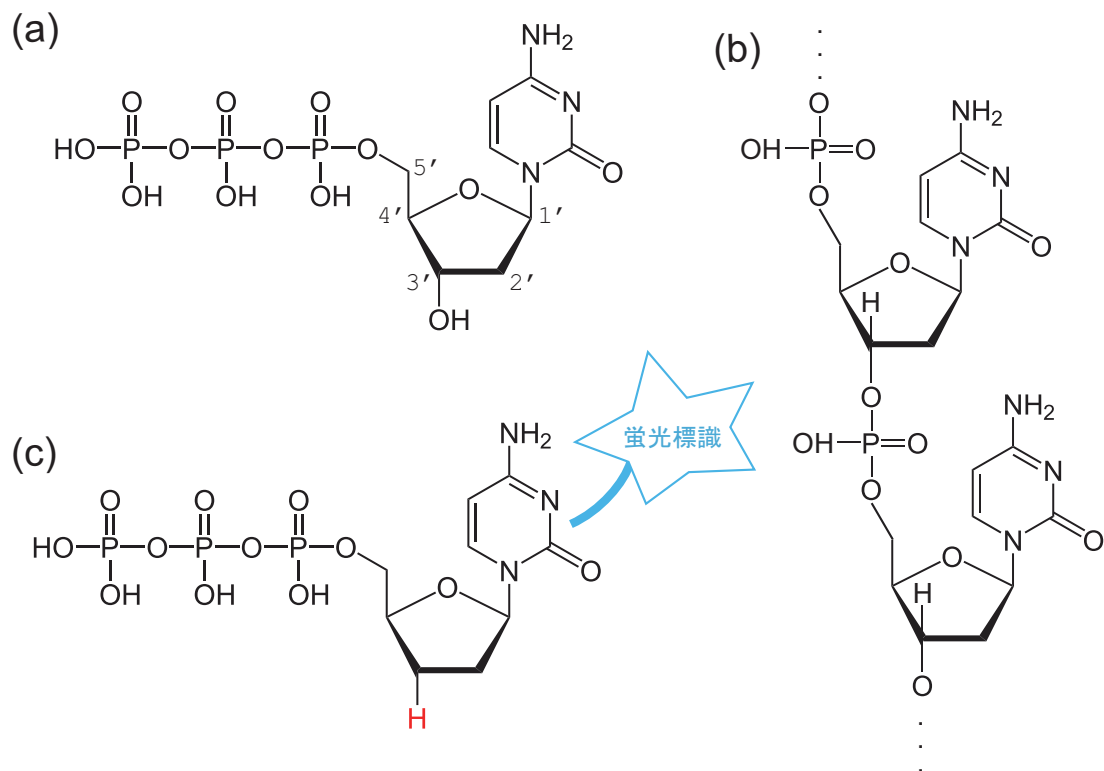


図 10: dNTP と ddNTP の化学構造の概略

(a) 例として、dCTP (Deoxycytidine triphosphate, デオキシシチジン三リン酸) を示す。糖の炭素に番号を付与した。(b) dCMP (Deoxycytidine monophosphate, デオキシシチジン一リン酸) の 3'-端の-OH(hydroxy group, ヒドロキシ基) を利用して起こる反応を介して次の dCMP が繋がったもの。DNA の一つの鎖の一部となる。なお、ヌクレオチドがこのような重合するときは、dNTP の外側の 2 つのリン酸が外れて、dNMP となる。(c) ddCTP (Dideoxycytidine triphosphate)。ddCTP では 3'-端が H に置き換わっており (なお、通常は骨格となる炭化水素部分の水素原子 (H) や炭素原子 (C) は省略形で作図される場合が多いことに注意)、これ以上 DNA 合成が伸長しない。さらに、ddCTP にのみ水色の蛍光標識が付加されているため、DNA の伸長が停止して水色の蛍光が観測された場合、末端の塩基はシトシン (Cytosine, C) であると分かる。

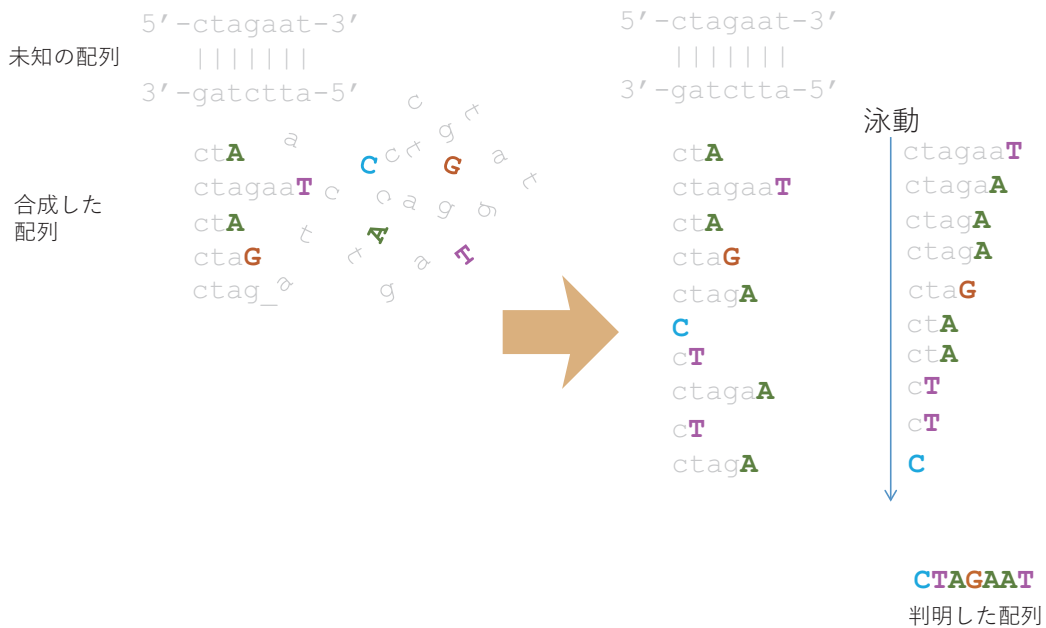


図 11: Chain-termination sequencing の概略

により鋳型 DNA に相補的な DNA 配列の合成を行った後、キャピラリーゲル中で電気泳動を行って塩基配列の決定を行うのが Chain-termination DNA sequencing である。

DNA は負に帯電しているため、キャピラリーゲル中では正の電荷がある方に泳動される。この時、DNA の分子量が大きいとキャピラリーゲル中で移動が遅くなるので、DNA 分子は分子量が小さい順に速く泳動が進む。ここで蛍光標識された末端の ddNTP が A, C, G, T のどれに該当するかを泳動中の DNA の蛍光検出器到着順に次々に記録していくことによって、塩基配列パターンを決定することが出来るのである (図 11)。

Chain-termination DNA sequencing で一度に決定できる配列の長さには限界があり、その原因には以下のようなものが考えられる。

**DNA ポリメラーゼの DNA 合成能力の問題** DNA 合成酵素が合成できる DNA の長さに限界がある。

**ddNTP が取り込まれ続ける確率の低下の問題** ddNTP が一度結合した DNA はそれ以上の伸長が起かない。塩基長が長ければ長いほど、この ddNTP の付加が生じて末端の塩基に辿り着くより前に伸長が止まる可能性が高まる。

Chain-termination DNA sequencing でも通常の PCR と同様、DNA の増幅が行われる。PCR ではプライマー 2 本鎖両方のプライマーを用いるため、1 回の反応で複製されて新しく増えた DNA 配列全てが、その次のサイクルで複製の対象となる。複製の鋳型となった分の DNA と、複製された DNA を合わせると数が 2 倍になり、これがサイクル毎に行われることで DNA の数が 2 倍ずつ指数関数的に増加することになる。一方、Thermal cycle sequencing ではプライマーを 1 種類しか加

えない。このプライマーは新しく複製された DNA には結合しないため、毎サイクルで鋳型となる DNA 量は一定であり、したがって増加する DNA 量はサイクルが進んでも一定で、直線的な増加となる。

## 5.3 次世代シーケンサー

### 5.3.1 次世代シーケンサー用のライブラリ調整

次世代シーケンサー用のライブラリ調整の目的は、読みたい DNA 分子を 100 ~ 500bp の断片に切り分け、それらの配列を別々に増幅させることである。これにより、DNA 配列を読み取るための反応を断片配列ごとに並列で行うことが可能になる。はじめに、超音波処理などによって DNA 分子をランダムに切断し、その後それらの DNA 断片を固定化する。固定化の方は二種類ある。一つ目の方法では DNA 断片にアダプター配列を結合させ、スライドガラスにコートしたオリゴヌクレオチド鎖とアダプター配列の相補的な塩基対形成によって固定化する。二つ目の方法では DNA 断片にビオチンを結合させ、金属ビーズにストレプトアビジンがついたビーズに結合させることで DNA 断片を固定化する。固定化された DNA 断片は、アダプター配列中に含まれる配列を元にプライマーを作ることが可能であるため、PCR によってそれぞれ増幅される。

### 5.3.2 イルミナーシーケンシング

Illumina Sequencing では、まず、標的 DNA をフローセル (flow cell) と呼ばれるガラス基板上で増やす。ここで、ブリッジ PCR という 1 本鎖 DNA を増幅する手法が使われる (図 12)。この手法では、あらかじめ配列の異なるオリゴヌクレオチド 1, 2 をそれぞれ用意し、フローセル上にその 5' 末端を固定する。次に、超音波などによりランダムに断片化された目的 DNA 断片の両末端にも先ほどのオリゴヌクレオチド 1, 2 の配列と相補的になるように、アダプター 1, 2 を付加し、これをフローセルに乗せる。すると、目的 DNA 断片とオリゴヌクレチドがアダプターを介して結合する。そして dNTP と DNA ポリメラーゼを加え、DNA 鎖を伸長させる。その後、伸長した DNA は 3' 末端側のアダプター配列を介して近くのオリゴヌクレオチドとブリッジ状 (架橋状) の結合を形成する。

架橋構造は伸長反応の後にアダプター間の結合が解け、それぞれ別のあらかじめフローセル上にあらかじめ固定してあったアダプターと結合をすることで、目的配列が増幅されてゆき、セル上で目的配列のクラスターを得ることが可能になる。

次にリボースの 3'-末端に保護基と呼ばれる化学構造を付加し、A,C,G,T の見分けがつくように蛍光標識を付けたヌクレオチドを用意する (図 13a)[16]。これによって、DNA 合成が 1 塩基ずつ一時停止しながら行われる。そして、一時停止中に蛍光の色からどの塩基かを識別する (図 13b)。それが終わると、酵素によって保護基が除去され (図 13c)、ヒドロキシ基 (-OH) を介して再び DNA の伸長合成が可能になる (図 13d)。これを繰り返すことで、塩基配列を決定できる。

### 5.3.3 SOLiD

SOLiD は Sequencing by Oligonucleotide Ligation and Detection の略称であり、大まかな流れは



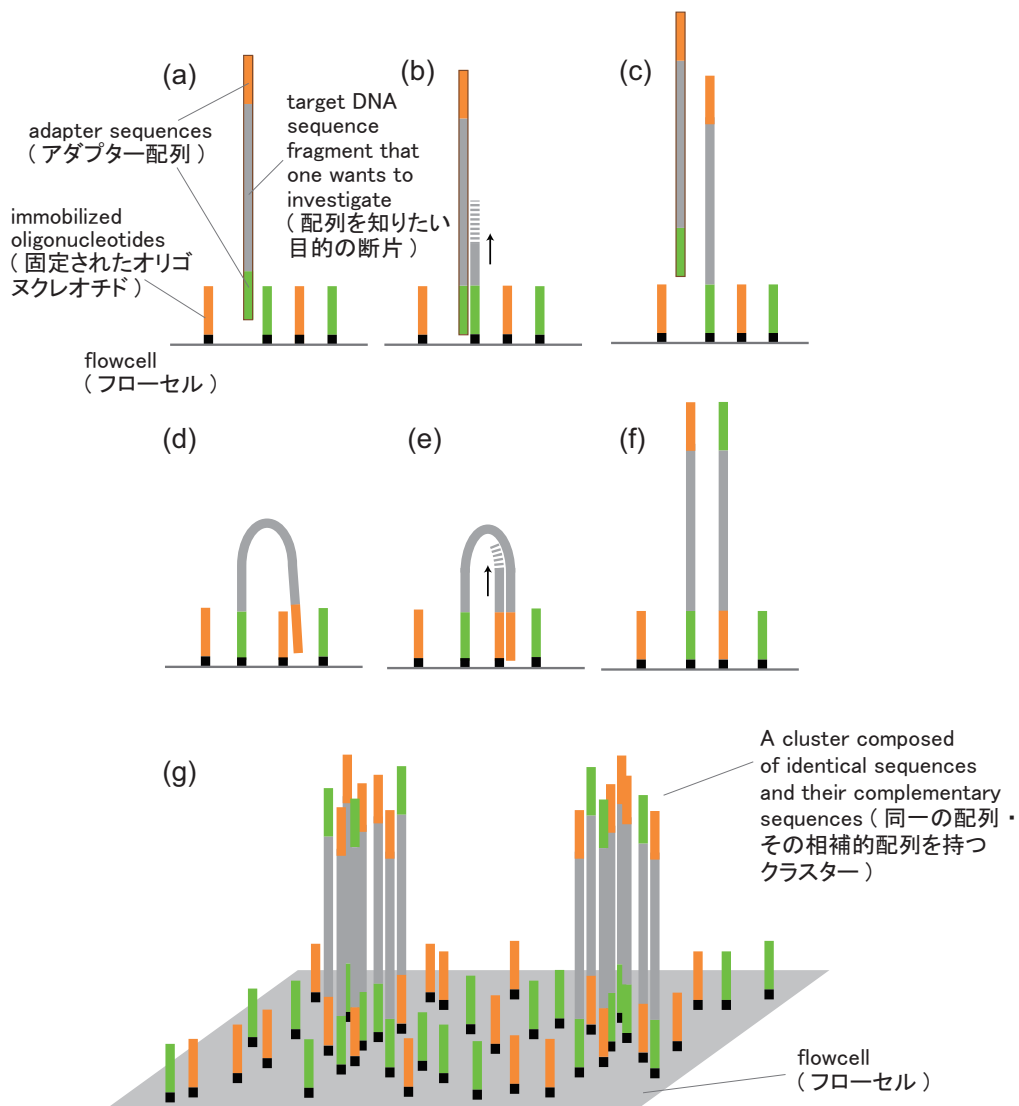


図 12: Bridge PCR の概要

(a) ランダムにバラバラにした DNA 断片の両端にアダプター配列を付加し、フローセル上に乗せる。(b) アダプター配列が、フローセル上に固定されたオリゴヌクレオチド配列と対合。DNA ポリメラーゼにより伸長反応が起こる。(c) 元の DNA 断片配列が離れる。(d) 伸長し対象の DNA 断片配列と相補的になった配列はその長さから近隣のオリゴヌクレオチド配列と結合し、橋渡しのような状態、いわゆるブリッジを形成する。(e) 同様に伸長反応を起こす。(f) 二本鎖が一本鎖に変性する。(g) このブリッジ形成、伸長反応、変性が繰り返されることでフローセル上に同一配列およびその相補的配列を持つ配列クラスターが形成される。最終的に、相補的配列は切断・洗浄され、同一配列だけが残る。

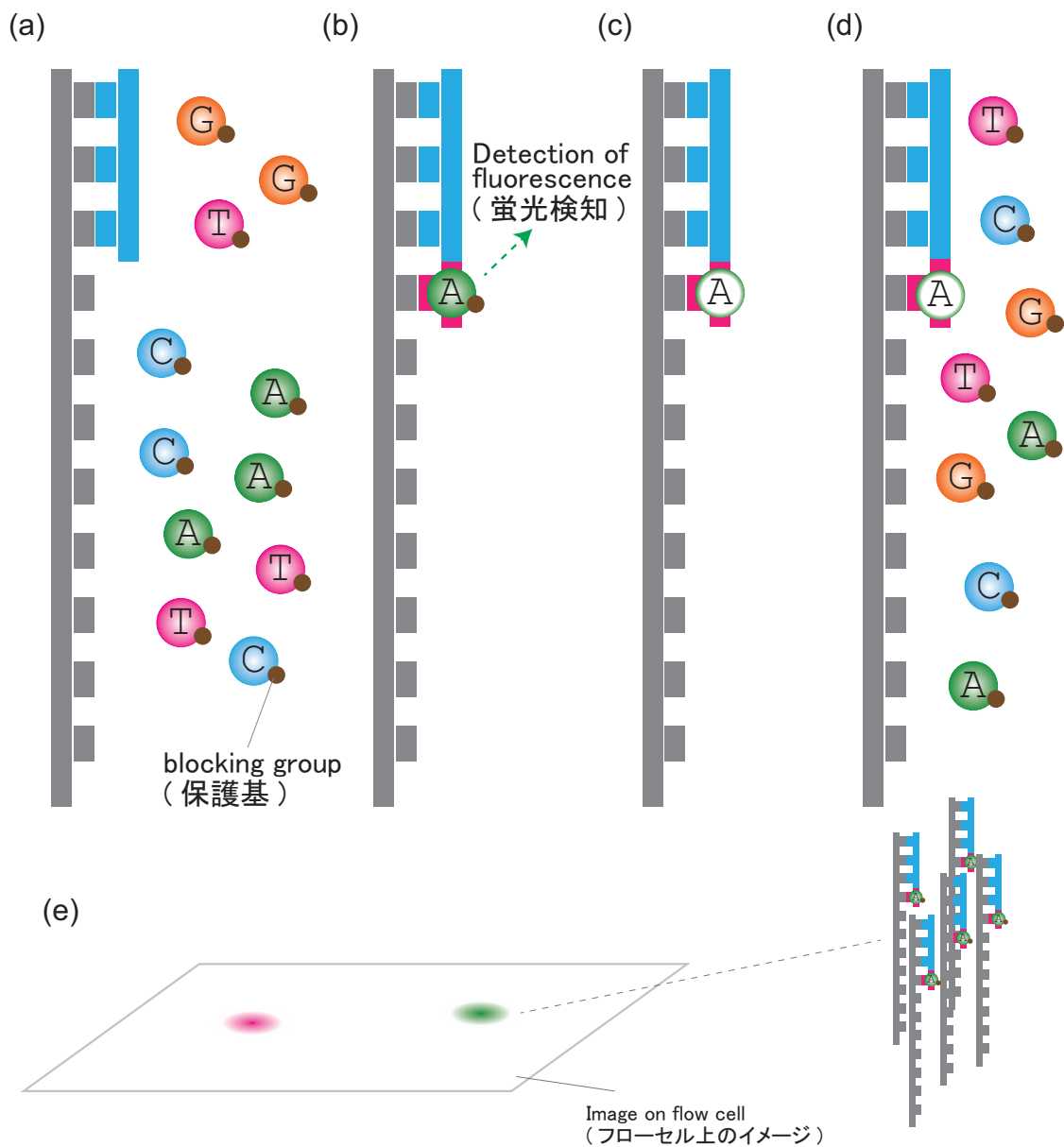


図 13: イルミネーションの概要

(a) 保護基と蛍光標識を付加されたヌクレオチドを DNA 伸長用に用意。(b) ヌクレオチドを取り込んだら、そこで伸長を一時停止。蛍光色によって、塩基の種類を判別。(c) 保護基および蛍光色素を除去。(d) 次のサイクルを開始する。(e) フローセル上で観測される蛍光の概略の例。A の合成で停止しているクラスター (同一の配列を持ち、同一の蛍光を発することに注意) からは、緑色の蛍光が観測されている。

1. ライブラリ (サンプル DNA) の作製
2. エマルジョン PCR (Emulsion PCR)
3. シーケンシング

である [17, 18, 19]。ライブラリの作成では、DNA を 100~150 bp くらいの大きさにランダムに断片化し、断片群の両端にアダプターを付加する。これをフラグメントライブラリ (Fragment library) と呼ぶ<sup>5</sup>。

DNA の増幅にはエマルジョン PCR を使用する (図 14)。エマルジョンとは、水と油が混ざり合ったような状態である<sup>6</sup>。通常水と油は混じり合うことはないが、超音波やせん断力を受けると、水または油が、微細な液滴となって分散して混ざり合い、乳化状になる。このような状態のことをエマルジョンと呼ぶ。エマルジョン PCR とは、オイル中に存在する各水滴内に

- 断片化した 1 分子の 1 本鎖 DNA
- 1 個のビーズ (Bead)

が含まれる状態で PCR 反応を行うことにより、ビーズ上で鋳型 DNA を増幅する方法で、次世代シーケンシングにおける DNA を増幅させる技術の一つである。

この PCR では、ビオチン (biotin) が 5' 末端に付加された DNA プライマーを使用する (図 14b) [23, 24]。ビオチンはストレプトアビジン (Streptavidin) というタンパク質と強固に結合する小さな有機分子であるため、ストレプトアビジンタンパク質に覆われた小さな金属ビーズを支持体として用いることで、DNA プライマーがこの金属ビーズに結合する。

そして、解析対象となる DNA 断片と金属ビーズがともに 1 つずつ水滴中に入るように混合して油と水の混合液の中で振盪し、増幅試薬とともに油中水滴エマルジョンに内包させ (図 14a)、DNA 断片を PCR で増幅する (図 14c)。こうして、各 DNA 断片は各水滴中のビーズ上で数百万コピーにまで増幅される。そしてエマルジョンを破壊してビーズを濃縮し、ビーズひとつが収まる穴が無数に開いたピコタイタープレート上に載せてシーケンシングを行う。

シーケンシングの際には、蛍光色素の付加された 1024 種類の 8 塩基プローブを使用する。このプローブには 4 色規定に基づいた 2 塩基 (図 15a の先頭の 2 塩基)、ATGC のいずれかの 3 塩基 (図 15a の NNN で表される塩基)、どの塩基とも対となることのできるユニバーサル塩基 (universal base) が 3 塩基 (図 15a からは省略)、そして蛍光色素が付加されている (図 15a のプローブの色で表現)。また、ユニバーサル塩基といずれかの 3 塩基の間は開裂部位 (cleavage site) になっており、ライゲーション (ligation) 時にユニバーサル塩基と蛍光色素は遊離する。シーケンシングの大まかな手順として、まず作成した DNA 断片 (テンプレート) に第 1 プライマーを付加し、テンプレート配列と合うプローブをライゲーションする。蛍光を読み取った後は、ユニバーサル塩基と蛍光色素を切断し、

<sup>5</sup>ライブラリ作成において、SOLiD で使用できるライブラリにはフラグメントライブラリとメイトペアライブラリ (Mate pair library) の 2 種類が存在する。フラグメントライブラリは使用する DNA が少量でよく、ライブラリ作製時間が短い、バーコード化が可能等のメリットがある。メイトペアライブラリの作製には、DNA サンプルを数 kbp ごとに断片化し、断片の両端に後に「内部アダプター」となるアダプターを付加する。そして両端のアダプターをつなげて DNA を環状にし、アダプターの両脇 50 bp 地点を制限酵素で切断し、その前後の 50bp 領域を解析対象 DNA の中でシーケンシングする部分 (タグ) とする。メイトライブラリは構造解析や反復領域の特定などに有効である。

<sup>6</sup>石油業界などで良く使われている用語

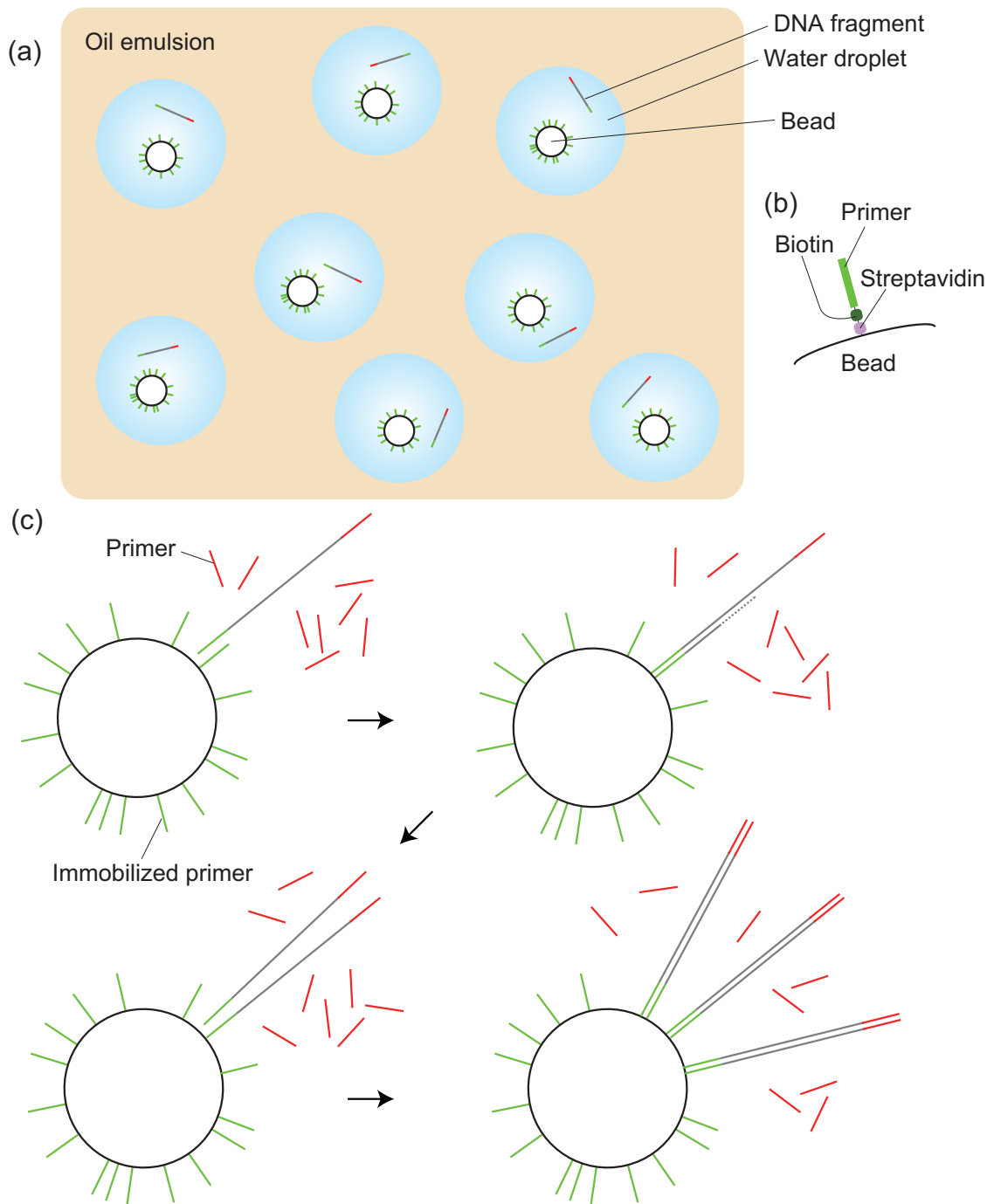


図 14: エマルジョン PCR の概要

(a) エマルジョンの様子。各水滴の中にはビーズと核酸断片配列が1つだけ入るように調節する。(b) ビーズ (Bead) の表面の拡大図。(b) 水滴の中のビーズ上で PCR がかかる様子。

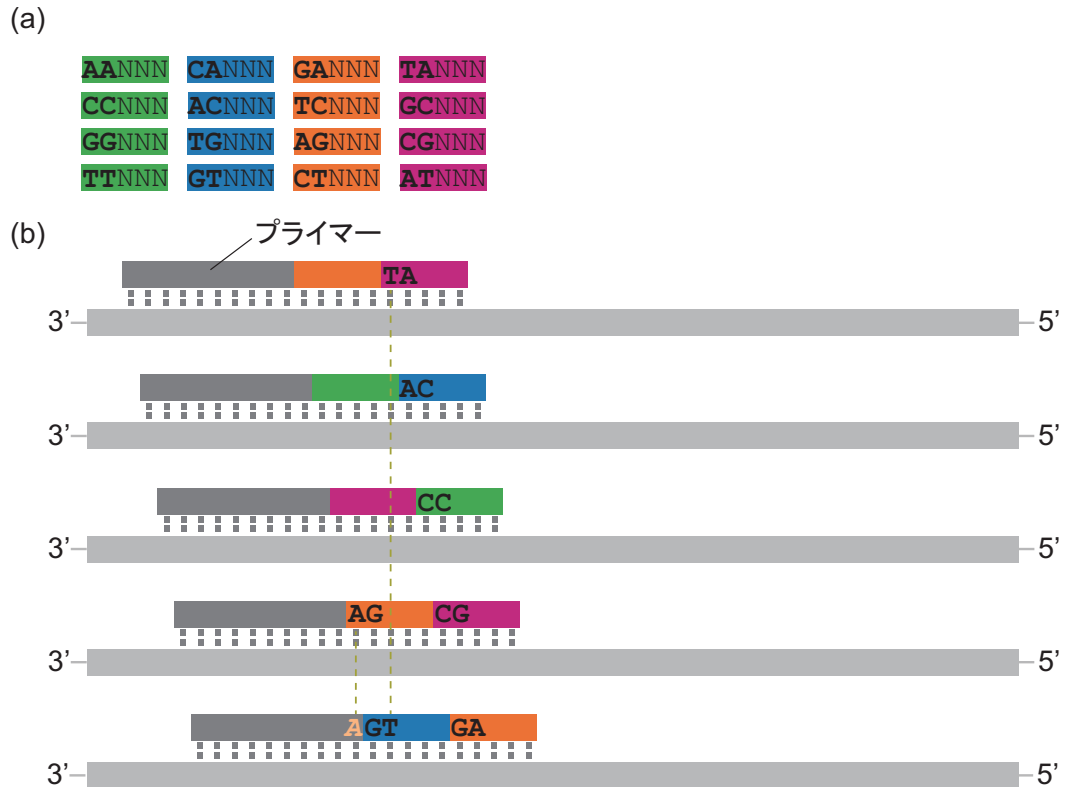


図 15: SOLiD によるシーケンシング

(a) 5 塩基全てのパターンをカバーするような 1024 通りの 5 塩基 DNA(プローブ, probe) を用意する (図を簡略化するため、ユニバーサル塩基は省略)。先頭の 2 塩基のパターンに基づいて、4 種類の蛍光標識を付与する。(b) プライマーを 1 塩基ずつずらして第 1 プライマーから第 5 プライマーまで 5 種類設計する。本図では、同一 DNA 断片の第 1 プライマー使用時から第 5 プライマー使用時までの様子を示している。プライマーのハイブリダイゼーション後、プローブのハイブリダイゼーション・ライゲーションと蛍光の読み取り (イメージング) を 3' 末端に向かって 1 つずつ行っていく。テンプレート配列末端までのハイブリダイゼーションが終了し、先頭から 1 塩基ずらして再びハイブリダイゼーションを行うべく、次のプライマーに切り替える時は、テンプレート配列に結合しているプライマーおよびプローブを一旦全て外すことに注意。第 5 プライマーおよびプローブのハイブリダイゼーションが終わった後は、第 1~5 プライマーで得られた蛍光パターンを統合して配列を読み取る。まず、プライマー配列末端の塩基の種類 (本例では "A" とする) からプローブの先頭の塩基 (本例では "A" とする) を読み取り、以降、蛍光色に基づいて、プローブの 2 塩基目の配列と、次のプローブの 1 塩基目の配列を割り出しながら、塩基配列を決めていく。

次のプローブのライゲーションを行う。これを何度も繰り返すことによって、蛍光を次々と読み取っていく。それが終了したら、プライマーおよびプローブを外し、第2プライマーで同様の作業をする。これを第5プライマーまで繰り返す。そして、図15bに示すように、各プライマー使用時の蛍光パターンを調べることで、配列を読み取ることができる。

SOLiDは配列がほとんど未知のゲノムを最初からシーケンスする *de novo* 配列決定には不向きであり、既に配列の大部分が分かっているゲノムの SNP 検出等に適用される。例えば、ゲノム上のある箇所  $p$  に SNP があることを確かめたい場合 (すなわち、 $p-1, p+1$  の位置の塩基は既知のものとして確定していることを前提とする)、 $p-1, p$  の位置に5' 端の2塩基が対合したプローブと、 $p, p+1$  の位置に5' 端の2塩基が対合したプローブの2色によって  $p$  の位置の塩基の読み取りが2回行われることになる。すなわち、 $p$  の前後の既知の塩基と SNP が疑われる位置の塩基、そして2つのプローブの蛍光色との整合性が確認されるため、より正確に塩基配列を決定できる [20]。

### 5.3.4 PacBio シーケンシング

PacBioとはPacBio社から発表された第3世代シーケンサーとよばれるシーケンサーである<sup>7</sup>[21, 22]。これは、新規のゲノム配列解析を行うに適した非常に長いリード長を得られる。PCRなしでもDNA1分子だけでシーケンスが可能となった技術である。その原理は、プライマーとポリメラーゼを結合させた鋳型DNAをzero-mode waveguide (ZMW) とよばれる無数の孔が開いたsingle molecule real time (SMRT) cellに入れることで、ポリメラーゼ複合体ができ、4種類の異なる蛍光がついた塩基がZMWの穴で光ることで検出され、配列がわかるようになっている (図16)[72, 73, 74, 75]。

### 5.3.5 ナノポアシーケンシング

第4世代シーケンサーであるオックスフォード・ナノポアテクノロジーズのMinIONの原理について概説しよう (図17)。ナノポアは生体分子もしくは人工材料でつくられた微細な穴を持つ構造物であり、シーケンサー内の電流を通さない二重膜中に配置されている。ナノポアの膜中での構造を保つ部位にはセンサーが取り付けられており、電流の変化を感知できるようになっている。ナノポアの微細な穴の付近もしくは内部をDNAやRNA等の分子が通過する際に、イオン電流に変化が生じるため、これをセンサーで感知し情報処理を行うことで通過した配列を決定することが可能となっている [26]。第4世代シーケンサーはそれ以前のシーケンサーと異なり、シーケンシング過程に酵素反応を含まず、1分子レベルに近い解像度で配列を読むことができる。そのため、DNAのエピジェネティックな修飾状態等、酵素反応を介すると揺らいでしまう情報までも得られることが期待されている。

Chain-termination sequencing を用いた塩基配列の決定には塩基の色素修飾を伴う核酸の伸長反応が必要であり、また、一般に次世代シーケンサーと呼ばれる第2世代、第3世代シーケンサーもDNAの伸長反応を蛍光標識で観察して塩基配列を決定するため、解析結果を得るまでに多くの時間を要するが、ナノポアシーケンサーは、そのようなシーケンス反応が不要な方法であるため、解析時間の大幅短縮が期待できる [79]。しかし塩基の読み取り精度に関しては、第1~3世代シーケンサーが > 99%であったのに対し、ナノポアシーケンサーでは 87~98% と現状では低い。

<sup>7</sup>[25]に掲載されている動画が分かりやすい。

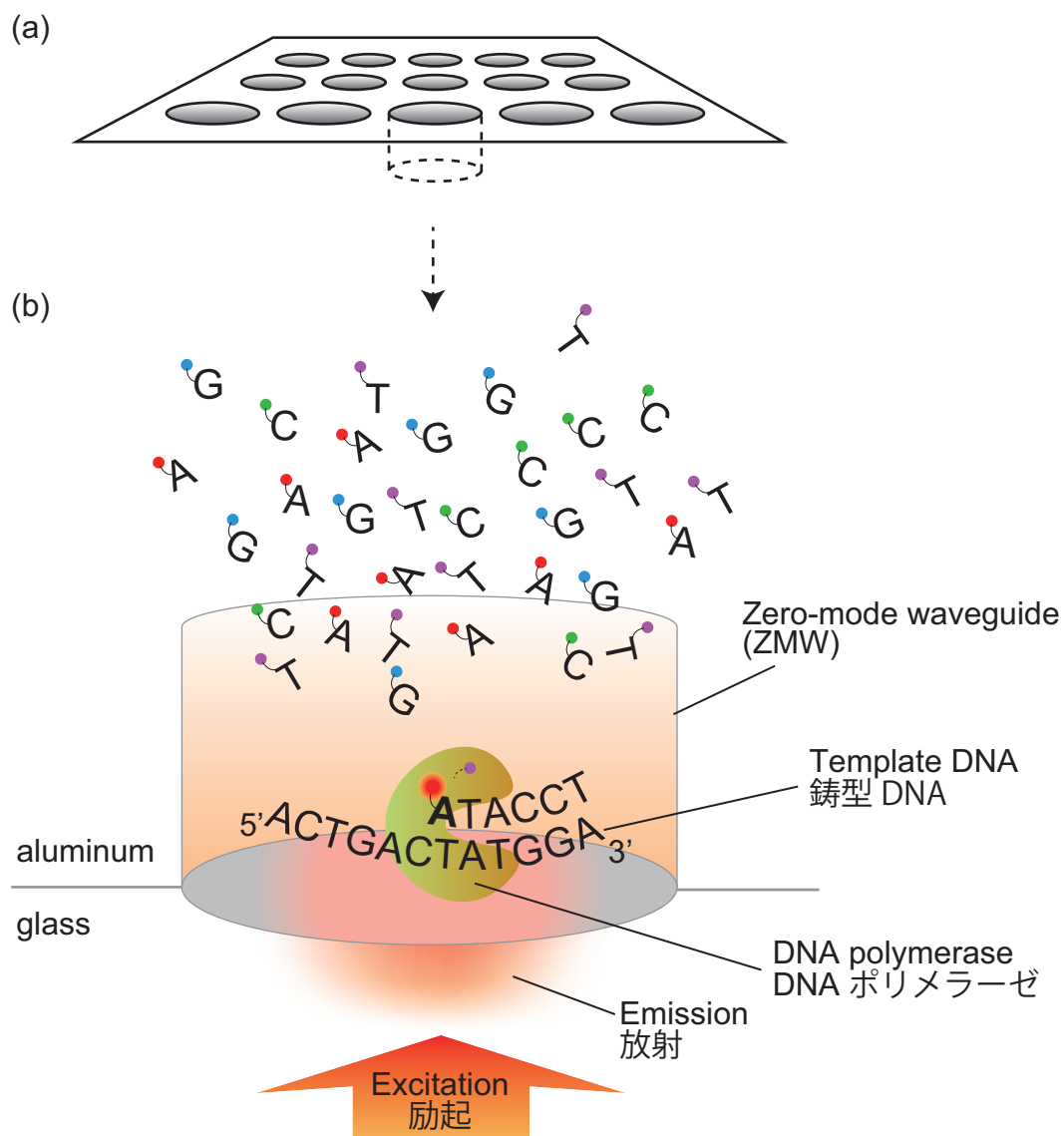


図 16: PacBio シーケンシングの概要

(a) single molecule real time (SMRT) cell の概略図。ここに、10 万個以上の zero-mode waveguide (ZMW) が配置されている。(b) zero-mode waveguide (ZMW)。底面に DNA ポリメラーゼが固定されており、蛍光標識された dNTP を取り込みつつ、DNA を伸長する。毎回 dNTP を取り込んだ時の蛍光色で、塩基の種類を判別する。ZMW の大きさは光の波長 (wave length) より小さくなっており、光は底面の DNA ポリメラーゼから離れたところまでは伝播しない。これによって、遊離している蛍光標識された dNTP まで照らしてしまうのを抑制している [76, 77]。

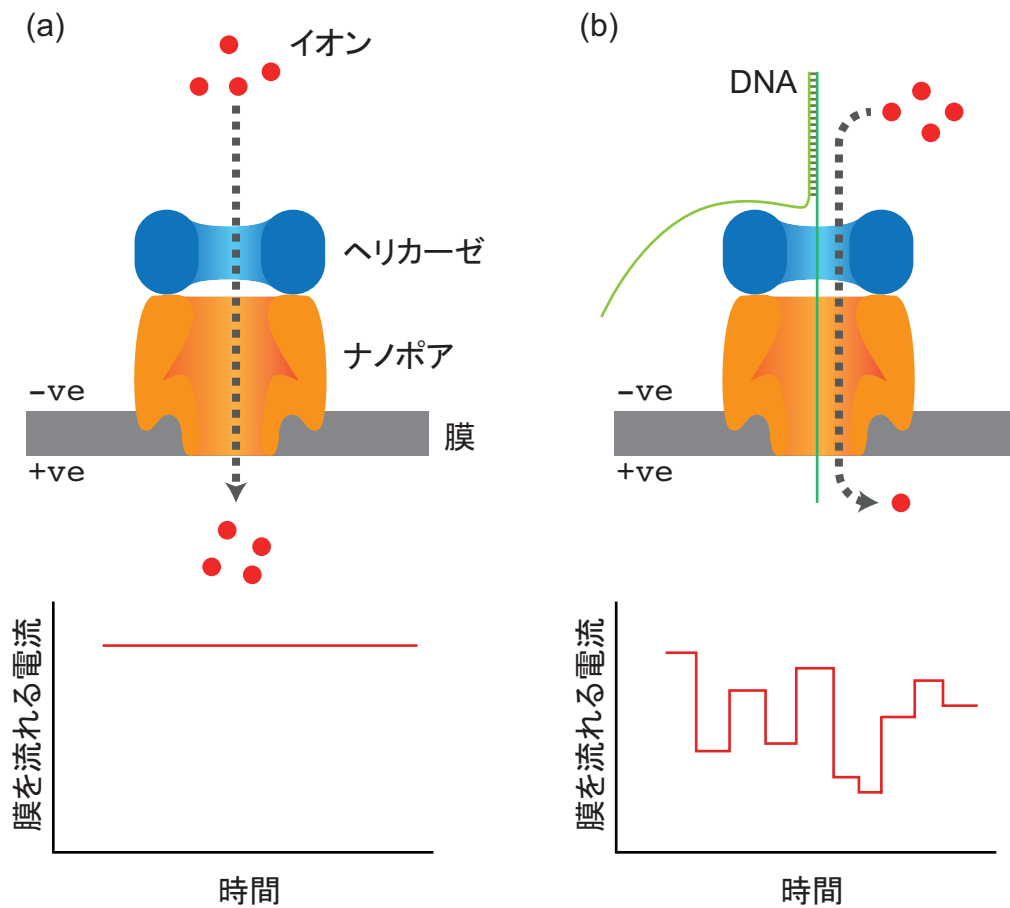


図 17: ナノポアシーケンシングの概要

(a) DNA が膜を通過していないときは、イオンが膜を通過する流れが乱されず、膜を流れる電流が一定となる。(b) DNA がヘリカーゼによって一本鎖となり、膜を通過するときは、イオンの流れが乱され、膜を流れる電流も変化する。この変化には塩基毎の特徴があるため、それによって塩基配列を読み取る。



## 5.4 反復配列の読み取り

### 5.4.1 ゲノムに存在する反復配列

反復配列 (繰り返し配列、repeated sequences, repetitive elements) は様々な種のゲノム中で観測される、繰り返し出現する配列パターンである。例えば、クモの糸を構成するフィブロイン (fibroin) という遺伝子は、90 個もの反復単位 (repeat unit) を持つものがあることが分かっている [12]。より身近なヒトゲノム配列に目を移すと、実に半分近くを反復配列が占めている [11]。特に反復配列パターンが隣り合って連続して並んでいるものをタンデムリピート (tandem repeat, 縦列反復配列)、隣り合わずにゲノム中に散在しているものを散在反復配列 (interspersed repeats) と呼んでいる。

マイクロサテライト (microsatellite) はタンデムリピートの一種で、通常 1~6 塩基対 (bp) 程度の大きさの塩基配列パターンの単位を複数回繰り返す反復配列であり [65]、ショートタンデムリピート (STR, Short Tandem Repeat) と呼ばれる。マイクロサテライトはヒトゲノム全体の 3% 以上を占め [67]、特に非翻訳領域に多く存在する。

マイクロサテライト領域では同じ配列パターンが反復するため、DNA 複製時に鋳型 DNA と複製鎖が反復単位分”ズレた”対合をしやすく (Slipped strand mispairing)、これによって反復単位分の配列の挿入や欠損が起こるため、マイクロサテライトによっては繰り返し回数が個体間で大きく異なる (多型性が高い) ことがある。このようなマイクロサテライトの特定の種別は、個人を識別するためのマーカーとして使用され得る。マイクロサテライトは 50 塩基対より長い構造変異の大部分を担っており [69]、また、これらのマイクロサテライト領域の顕著な変異性は、多くの真核生物における表現型に重要な役割を果たしている。特にマイクロサテライトはヒトの多くの遺伝性疾患に関与していることが示されているため、生物医学研究の重要なテーマとなっている [68]。

### 5.4.2 反復配列を読み取る際の問題

これまで学んだように、シーケンサーで読める配列の長さには限界がある。従って、長い配列を読む時は、多量の断片配列を読んだ後に、これらをつなぎ合わせて全長配列を推定する必要がある (配列のアセンブリ、Sequence assembly)。断片配列のパターンが解析対象全長配列の 1 箇所だけに存在する、すなわち、得られた断片配列パターンが全長配列の中で特異的である、と仮定できるならば、断片配列は全長配列の特定の 1 箇所由来であると仮定でき、可能な全長配列の候補を大幅に絞り込むことができる。

しかしながら、断片配列のパターンが全長配列の複数の箇所で繰り返し観測される可能性がある場合、その断片配列が全長配列のどの部位由来か、一意には決まらなくなることを意味する。この場合、可能な全長配列の候補が場合によっては、爆発的に増えてしまう。

この問題を理解するために、単純化して考えて、3 塩基までしか読めないシーケンサーを使って一本鎖の全長配列を決めるとしよう。ここでは、逆鎖の配列が読まれる可能性は考慮しない。そして、配列を何度読んでも AAC, AAA という 2 つの配列パターンしか出てこないケースを考えよう。仮にこれら 2 つの断片配列パターンが全長配列の中で特異的だと仮定すると、全長配列は AAAC と一意に決まる。しかし、この仮定が成り立たない場合、AAAC の他に、AAAAC, AAAAAC, ... 等、全長配列の候補がいくらかでも考えられてしまう。

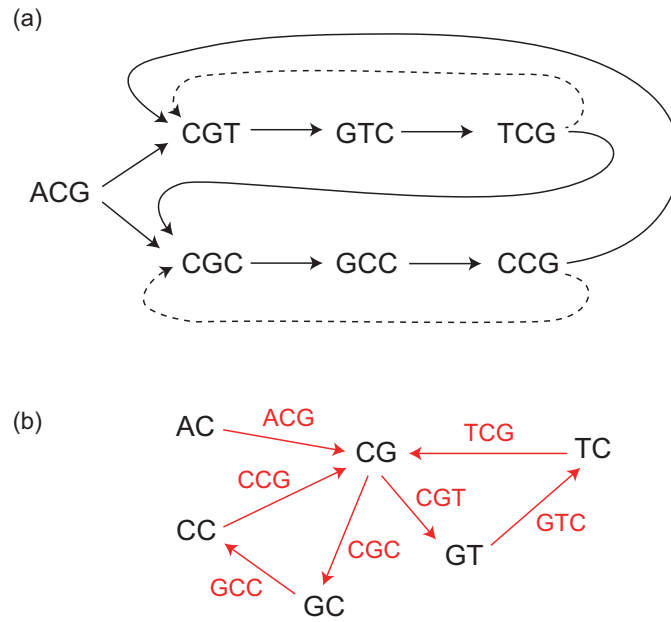


図 18: 3 塩基配列断片のアセンブルの模式図

(a) ある配列パターン尾部と、別の配列パターン先頭部が同一の場合、前者から後者へ向かって矢印が引かれている。矢印で結ばれたパターンは全長配列上で隣接している可能性を示しており、矢印の元は上流配列パターン、先は下流配列パターンを表している。例えば、ACG の最後の 2 塩基と CGT の最初の 2 塩基はともに CG であり、ACGT という配列の一部である可能性がある。各塩基断片が全長配列の中で特異的な場合、点線で示した経路は使われない。(b) (a) のグラフを de Bruijn graph に直したもの。エッジ (edge) の部分がシーケンサーから得られた配列パターンに対応することに注意。

もう少し複雑な例として、ACG, CGT, GTC, TCG, CGC, GCC, CCG という 7 種類の配列パターンが得られるケースを考える。これらの配列の前後隣接関係を有向グラフ (directed graph) と呼ばれる表現方法で示すと、図 18a のようになる。シーケンサーで得られた各配列はノード (node) と呼ばれる要素として扱われ、各ノードは前後の隣接関係を表すエッジ (edge, 図中の矢印) で結ばれる。各配列断片が全長配列の中で特異的な場合、配列のアセンブリは、図 18a の中から、全ての配列パターン、すなわち**全てのノードを一度だけ通る経路**を探すことに該当する。このような経路をハミルトン路 (Hamiltonian path) と呼ぶ。この場合は、全長配列の候補は ACGTCGCCG または、ACGCCGTCG の 2 種類に絞られる。しかしながら、特異的でない場合は、経路を何週もすることが許容されることになり、ACGTCGCCG, ACGTCGCCGT, ACGTCGCCGTC, ... 等、理屈の上では候補がいくらでも考えられてしまう。

なお、ハミルトン路の探索によって、全長配列を求めるのは、一般的に計算時間がかかるため、オイラー路 (Eulerian path, **全てのエッジを一度だけ通る経路**, 一筆書き) を探索する問題に定式化・帰着して解かれることが多い。この場合、図 18b に示すように、ノードではなく、エッジに配列パターンを対応させた、de Bruijn グラフと呼ばれるグラフを作成する [13]。詳しくは、Genomes 4 Chapter 4 p.102-3[1] を参照されたい。

## 6 ゲノム配列の機能解析法

### 6.1 ゲノム配列中の遺伝子領域の探索

#### 6.1.1 Open Reading Frame (ORF) の探索

生命現象において多くの化学反応や細胞・組織の維持を担っているのがタンパク質である。そこでゲノム配列や cDNA 配列 (mRNA に対応) が得られた後は、まずはタンパク質をコードする遺伝子領域 (タンパク質コード領域、protein coding region) はどこか、を探索するのが自然な流れとなる。

原核生物のゲノム配列や、真核生物より得られた cDNA 配列 (イントロンが取れている) が解析対象の場合、コード領域探索の最も簡単な方法は ORF (open reading frame) の探索である。これは開始コドンと同じ 3 塩基パターン + 間の 3 の倍数の長さの領域 + 終止コドンと同じ 3 塩基パターンで構成される領域のことである (図 19)。この定義では、「間の 3 の倍数の長さの領域」には同一フレームに終止コドンと同じ 3 塩基パターンが入ってはいけない。実はここがポイントで、逆に言えば、同一フレームに終止コドンと同じ 3 塩基パターンがずっと入らない長い領域が存在する場合、その領域はタンパク質をコードしている可能性が高くなる。

原核生物のゲノム配列や、真核生物より得られた cDNA 配列 (イントロンが除去されている) が解析対象の場合、コード領域探索の最も簡単な方法は ORF (open reading frame) の探索である。これは開始コドンと同じ 3 塩基パターン + 間の 3 の倍数の長さの領域 + 終止コドンと同じ 3 塩基パターンで構成される領域のことである (図 19)。さらに条件として、「間の 3 の倍数の長さの領域」には開始コドンと同じフレームに終止コドンと同じ 3 塩基パターンが入ってはいけない。実はここがポイントで、逆に言えば、同一フレームに終止コドンと同じ 3 塩基パターンがずっと入らない長い領域が存在する場合、その領域はタンパク質をコードしている可能性が高くなる。

```

1  acatctcccg gcggcgggccc gcggaagcag tgcagacgcg gctcctagcg gatgggtgct
61 attgtgagggc ggttqtaqaa gttaataaag gtatccatgg agaacctga aaactcagtg
121 gattcaaaat ccattaaaaa tttggaacca aagatcatac atggaagcga atcaatggac
181 tctggaatat ccctggacaa cagttataaa atggattatc ctgagatggg tttatgtata
241 ataattaata ataagaatuuu tcataaaagc actggaatga catctcggtc tggtagagat
301 gtcgatgcag caaacctcag ggaaacattc agaaacttga aatatgaagt caggaataaa
361 aatgatctta cacgtgaaga aattgtggaa ttgatgcgtg atgtttctaa agaagatcac
421 agcaaaagga gcagttttgt ttgtgtgctt ctgagccatg gtgaagaagg aataatuuuu
481 ggaacaaatg gacctgttga cctgaaaaaa ataacaaact ttttcagagg ggatcgtttg
541 agaagtctaa ctggaaaacc caaactuuuu attattcagg cctgccgtgg tacagaactg
601 gactgtggca ttgagacaga cagtgggtgt gatgatgaca tggcgtgtca taaaatacca
661 gtggaggccg acttcttcta tgcatactcc acagcacctg gttattattc ttggcgaaat
721 tcaaaggatg gctcctgggt catccagtcg ctttgtgcca tgctgaaaca gtatgccgac
781 aagcttgaat ttatgcacat tcttaccggg gttaacgcaa aggtggcaac agaatttgag
841 tcctuuuuuu ttgacgctac ttttcattga aagaaacaga ttccatgtat tgtttccatg
901 ctcaaaaaag aactctatuuu ttatcactaa agaaatgggt ggttgggtgg ttttttttagt
961 ttgtatgcca agtgagaaga tggatatatt ggtactgtat ttccctctca ttttgaccta

```

図 19: cDNA 配列中の ORF (Open Reading Frame) の例

GenBank ID: NM\_032991 (CASP3, caspase 3) の部分配列。97 塩基目から始まる色違いの塩基配列が最も長い ORF である。その上流にある短い ORF も下線と別の色で示した。

真核生物のゲノム配列にはイントロンが含まれており、かつ、遺伝子の密度も低いいため、ORF の探索のみでコード領域を予測するのは一般的には困難である。

### 6.1.2 同源性検索を用いた既知の転写産物情報・タンパク質情報の活用

ゲノム配列中の特定の領域の塩基配列情報をもとにタンパク質が合成されるためには、まずその領域が転写されることが必須である。このように考えると、注目しているゲノム配列領域が既知の転写産物と一致していれば、その領域が特定の条件で転写されている可能性が示唆される<sup>8</sup>。そこで、cDNA 配列やその短い断片である EST (expressed sequence tag) 等の既知の転写産物情報を集めたデータベース (dbEST 等) が構築されてきた [83]。

また、そのようなデータベースの中からゲノム配列の特定の領域と一致している配列を高速に抜き出す情報科学的手法の開発も進められた。一般的に、データベース中の塩基配列・アミノ酸配列から、与えられた塩基配列・アミノ酸配列に一致する、もしくは類似するものを抽出する情報科学的手法を同源性検索 (homology search) と呼ぶ。BLAST は初期に開発された著名な同源性検索用のソフトウェアである [84]。

同源性検索用ソフトウェアの中にはゲノム配列をアミノ酸配列に翻訳しつつ、データベース中の既知のタンパク質の中にそのアミノ酸配列と一致したもの、あるいは類似したものを検索できるものもある。これによって、注目しているゲノム領域が潜在的にタンパク質配列をコードし得るかを確かめることができる。

<sup>8</sup> これを実験的に検証するためには、RT-PCR やノーザンブロットング、転写産物のシーケンシング等が使われる。

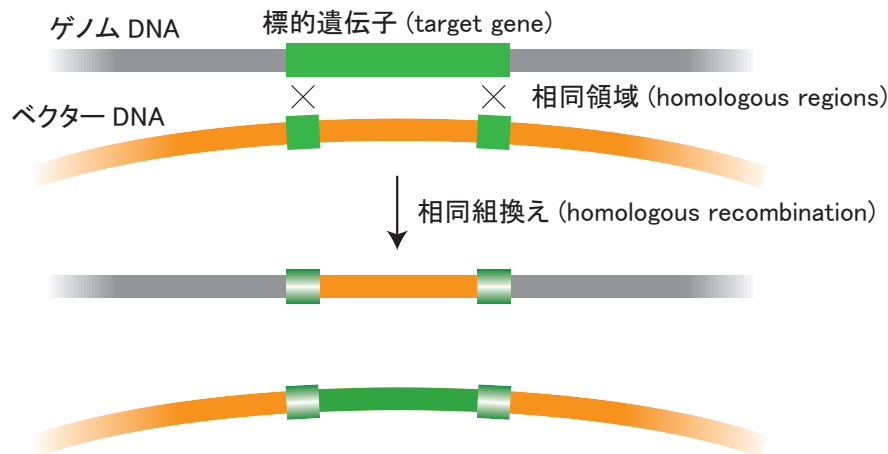


図 20: 相同組換えによる遺伝子のノックアウトの概要

注目しているゲノム配列領域が既知の配列と類似しない新規タンパク質をコードしている場合、相同性検索でその領域をタンパク質コード領域の候補として検出するのは難しくなる。

### 6.1.3 確率モデルの活用

タンパク質をコードするゲノム配列パターンには一定の傾向がある。この傾向を確率モデルを使って表現し、これをもとに、与えられたゲノム配列パターンの中でどの部分がどれくらいの確率でタンパク質をコードしているかを計算・予測する方法論の開発も進められた。特に隠れマルコフモデル (HMM, Hidden Markov Model)[85, 86] を用いた遺伝子領域探索は、分子進化的に起こる塩基配列やアミノ酸の置換・挿入・欠損をよく表現できることもあって、広く研究が進められた。例えば、AUGUSTUS[87] は、隠れマルコフモデルを利用した遺伝子予測プログラムであり、既知の遺伝子配列データを用いて、エクソン、イントロン、プロモーターなどの遺伝子領域の各部分の特徴を利用して遺伝子領域を予測する<sup>9</sup>。

## 6.2 遺伝子の機能阻害による機能解析

### 6.2.1 相同組換えによる遺伝子のノックアウト

遺伝子のノックアウト方法として基本的な、相同組換えによる方法を解説しよう (図 20)。まず機能を調べたい遺伝子の配列を読み、組み換え対象となる領域を定める。次に、組み換え対象領域と同じ 5' 末端および 3' 末端配列を持ち、かつ抗生物質耐性遺伝子等のスクリーニングに用いられる遺伝子を含むベクターを作成する。このベクターを細胞へと導入することで一定の確率で相同組み換えが生じ、ベクターによって導入された遺伝子によって対象遺伝子は中断され発現しなくなる。さらに、ベクターの導入が成功している細胞はスクリーニングが可能な状態となっているため容易

<sup>9</sup>イントロンの長さをモデル化することによって長い配列も解析できることが特徴の 1 つ。

に選別される。あとは、スクリーニングによって分別された細胞群間での形質の違いを見ることで、遺伝子機能の推定を行うことができる。

### 6.2.2 RNAi を用いた遺伝子のノックダウン

RNAi は、2 本鎖 RNA が標的となる RNA と対合してこれを分解し、その RNA が細胞内で機能を発揮するのを阻害する現象である [34]。標的が mRNA であれば、その mRNA からのタンパク質の翻訳が抑制されることになる。従って、特定の mRNA に対合する 2 本鎖 RNA を細胞内に導入し、RNAi の機構を働かせることができれば、その mRNA に対応する遺伝子の機能を阻害することができる。RNAi は本来は生命現象の 1 つであるが (ウイルスやトランスポゾンに対する防御機構等)[35]、このように、特定の遺伝子機能を阻害する実験手法の 1 つとして言及されることが多い。

具体的な作用機構としては、図 21 に示すように、2 本鎖 RNA が Dicer という酵素によって分解されて siRNA (small interfering RNA あるいは short interfering RNA) が生成された後、RNA-induced silencing complex (RISC) と呼ばれるタンパク質複合体によって取り込まれる。そして siRNA は 1 本鎖となり、これが標的配列を見つけるガイドとして機能し、標的 mRNA を切断する。

### 6.2.3 制御可能な Cas9 エンドヌクレアーゼを用いた遺伝子の改変

Cas9 (CRISPR associated protein 9) というエンドヌクレアーゼを用いると、ゲノム上の狙った領域を改変することができる。その機構の模式図を図 22 に示した。Cas9 は

- 標的塩基配列と相補的なガイド RNA (guide RNA) を含む crRNA (CRISPR RNA)
- crRNA と一部相補的に結合する tracrRNA (trans-activating CRISPR RNA)

と複合体を形成し、標的遺伝子を認識して切断する [36, 38, 39]。細胞内では非相同末端結合 (non-homologous end joining, NHEJ) と呼ばれる修復プロセスが働くが、このプロセスは、エラーを起こしやすく、それによって、切断部位の塩基の挿入や欠損が起こる。このようにして、エラーによって改変された DNA が転写・翻訳されても、多くの場合、合成されたタンパク質は本来の機能を失っている。これによる表現型の変化を調べることにより、遺伝子の本来の機能を推定することが可能である。

切断部位と相補的な配列を持つベクターを用い、相同組み換え (homologous recombination, HR) による修復経路 (homology directed repair, HDR) を働かせれば、切断部位にベクターが持つ配列を挿入することが可能である (図 23)[37, 41]。

この実験手法は、ゲノム配列そのものを改変するので、効果が一過性の RNAi と違って、遺伝子の不活性化が恒久的に続く。

近年はこの技術に関する報道が多くなってきており、社会の関心も高まっている。例えば、2018 年に中国の賀建奎・南方科技大学副教授は CRISPR/Cas9 を用いて CCR5 遺伝子を改変した双子の女兒を誕生させたと主張し、これがニュースとなって世界に衝撃を与えた。ここで、履修者による報道の紹介・論説の一部を紹介しよう。

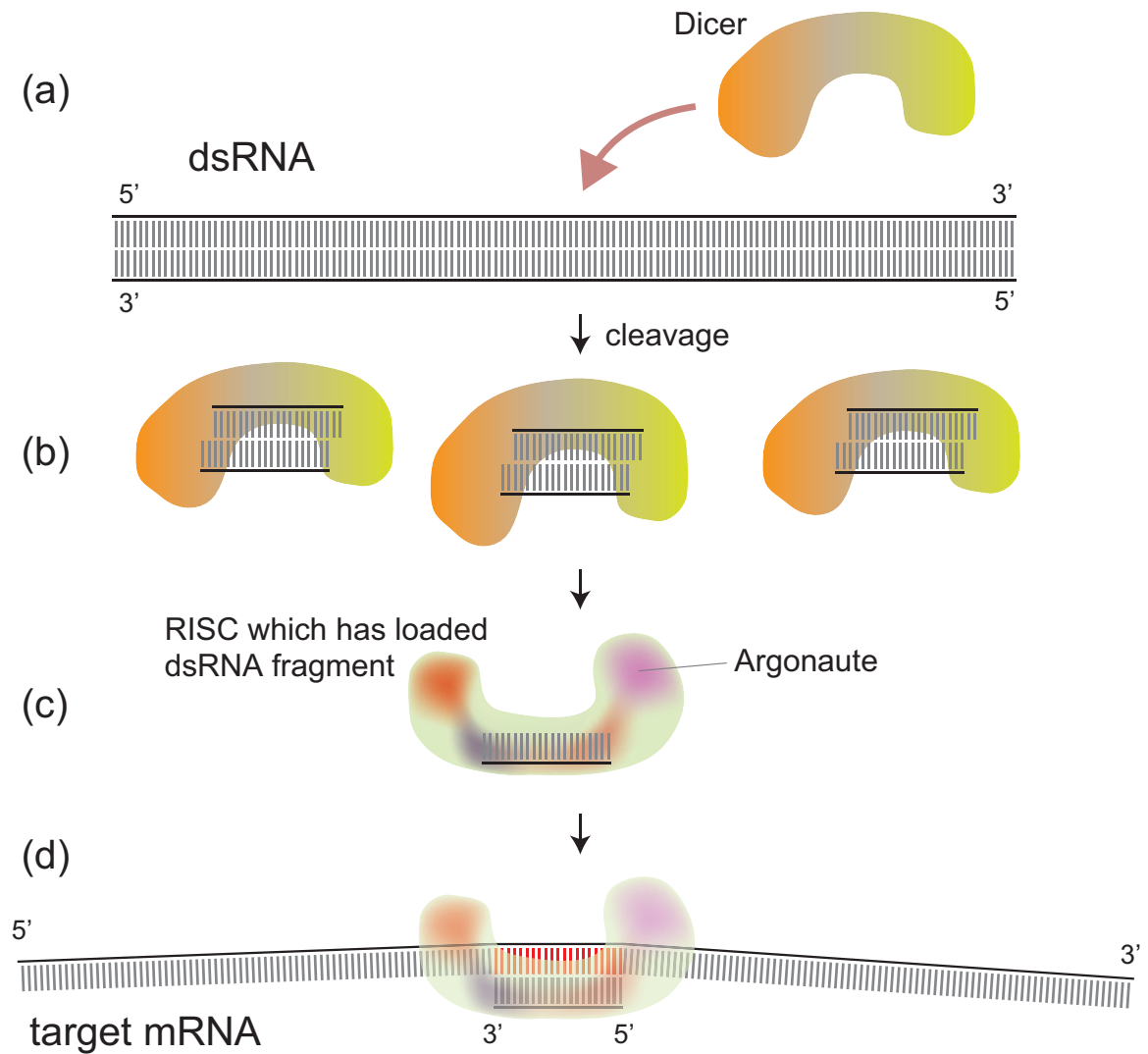


図 21: RNAi が働くメカニズム

(a-b) dsRNA を特異的に切断する酵素 Dicer によって、dsRNA が分解され、両鎖で 3' 末端が 2 塩基突き出た短い dsRNA (21-25 塩基くらい) が生成される。これは siRNA (small interfering RNA) と呼ばれる。(c) siRNA が Argonaute というタンパク質を含む、RISC (RNA-induced silencing complex) という複合体に取り込まれ、一本鎖となる。(d) 一本鎖となった RNA が、標的となる mRNA の一部と相補対を形成し、標的 mRNA は分解され、翻訳が阻害される。

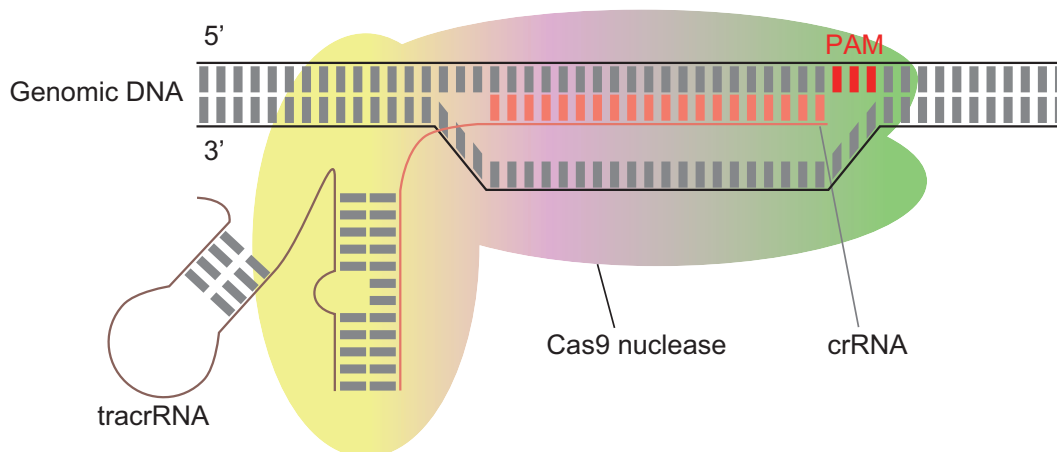


図 22: Cas9 エンドヌクレアーゼが働くメカニズム

Cas9 と tracrRNA (trans activating RNA), crRNA (CRISPR RNA) が複合体を形成し、crRNA の持つガイド RNA (guide RNA) 配列が DNA の標的部位と対合し、PAM (Protospacer adjacent motif) の 5' 上流側の DNA 配列を切断する。PAM は元来は 5'-NGG-3' という配列パターンである。

**CRISPR/Cas9 技術の改良 [40]** CRISPR/Cas9 の技術は革新的な技術で、食品、化学物質、薬剤や治療の分野において現代の遺伝子研究に大いに貢献してきた。ところが、Cas9 酵素は DNA の 2 本鎖を切断するため、細胞が修復しようとするとは完全に元に戻せるとは限らなくなる。そのため、より安全性の高い遺伝子編集技術が求められていた。そこで、Cas9 酵素を変異させて 2 本鎖を切断せずに DNA に結合する技術 (2017 年 12 月ソーク研究所) や、DNA 鎖を切断せずに 1 塩基だけ挿入や削除する技術 (ハーバード大学デイヴィッド・リュウ)、CRISPR が誤って標的と似ている遺伝子を切断しないように、活性を厳密に制御するツールの開発 (カリフォルニア大学バークレー校、カリフォルニア大学サンフランシスコ校、ハーヴァード大学、ブロード研究所、およびトロント大学の研究チーム) などが進められているさまざまな用途において、それぞれ適している新しい技術を組み合わせることが重要であろう。

**2019 年 3 月からゲノム編集を施した食品の流通が日本国内で解禁 [42]** CRISPR/Cas9 を用いた食品の販売は欧米では行われておらず、日本が最初とのこと。CRISPR/Cas9 を使用することで、普通のマダイよりも食べられる部位が 1.2 倍のマダイや、モミを大きくした稲などの食品が手に入るようになる。こうした消費性に優れた食品に関して、7 割の人が好意的な態度を向ける一方で、半数がまだ慣れない技術を使用して開発された食品への不安感も感じているという。ゲノム編集を使用して作られた食品がいかに安全なのかを伝えていくことが、今後の課題になるだろう。

**リラックスや血圧上昇抑制の効果が強化されたトマトの開発 [43]** CRISPR/Cas9 (ゲノム編集) が日本では農業のトマトに利用が進んでおり、リラックスや血圧上昇抑制の効果のある「γ-アミノ酪酸 (ギャバ)」に注目し、ギャバ生成酵素の活性中心を覆うフタをゲノム編集で除き、



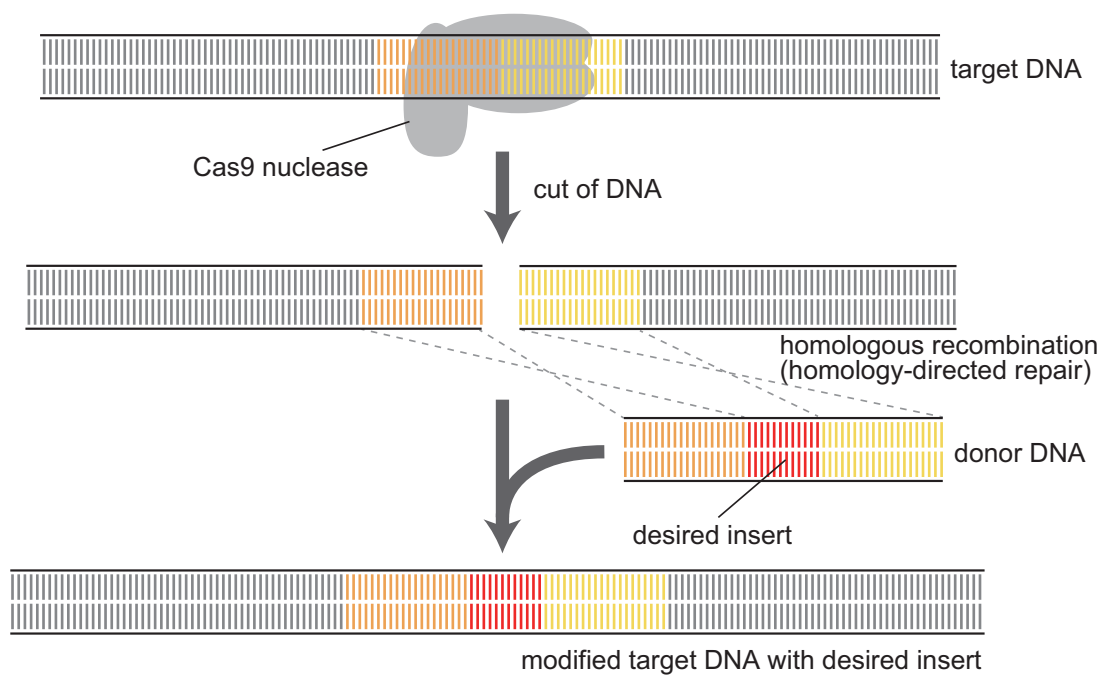


図 23: 相同組み換え修復 (HDR) によるゲノムへの DNA 配列の挿入

挿入したい DNA 配列の両側の配列を切断・挿入位置のゲノム配列と同一にしてベクターにする。Cas9 が DNA 二本鎖の狙った位置を切断した後、ベクター配列との間で相同組み換え修復 (HDR) が起こり、DNA 配列が挿入される。

「元々の作る力を強める」ことで、ギャバを野生型の15倍含むトマトを作った。

**CRISPR/Cas9による自閉症の霊長類モデルの作成** [44] 2019年6月13日のGENによるニュースで、中国科学大学とMITの合同チームがCRISPR/Cas9による自閉症の霊長類モデルの作成に成功したことが報告された。自閉症にはShank3という遺伝子が関わっていて、そのミューテーションにより自閉症をはじめとする様々な神経発達障害を引き起こす。今後は他の神経発達障害の生物モデルを作成し、研究の幅が広がることが期待されている。

## 7 転写産物の解析法

### 7.1 遺伝子発現解析の意義

細胞の形質や細胞中で起こる反応の多くは、遺伝子の発現量・発現パターンによって制御されている。従って、多くの発現量・発現パターンは細胞腫や組織によって異なる。また、薬剤などによる細胞への刺激やその後の時間の経過でもパターンは変化する。細胞は環境・状況に応じて、必要な遺伝子を適度に発現させているのである。従って、様々な遺伝子の発現パターンを解析することによって、遺伝子の機能を推定することが可能である。

これは医療への応用につながる。病気によって遺伝子の発現パターンが変わる場合はその病気の診断やメカニズム解明につながる。例えば、癌細胞特異的に発現が増加している遺伝子があれば、その遺伝子は細胞の癌化に関連している可能性があり、それをコードするタンパク質を標的にした薬剤を開発する等、創薬の分野に応用できる可能性がある。さらに新薬の開発の上で、投薬時の遺伝子の発現を調べることで薬の効きを判断し、また副作用についても評価できる。例えば、薬の効果が人によってバラつきがある場合があるが、患者の遺伝子を解析することで薬の効果の是非が測定できるものである。SLFN11と呼ばれる哺乳類特有の遺伝子があるがこれが抗がん剤の効果予測のバイオマーカーとしての使用が期待できる [121]。

遺伝子発現解析によって、ある生命現象に関わる遺伝子を推定できた場合、その遺伝子をRNAiを用いてノックダウンしたり、CRISPR-Cas9を用いてノックアウトしたりすることによって、その生命現象に影響があるか、検証することができる。

実験より得られる遺伝子発現データには、相対発現量と絶対発現量がある。

マイクロアレイやqPCRで遺伝子発現を調べる際には、サンプル中の全細胞の平均的な発現状況を見ており、例えば癌の微小環境において、癌細胞やそこに浸潤している免疫細胞がおり、さらに免疫細胞の中でも疲弊しているものや細胞障害性を有する細胞があるなど非常に複雑であるが、細胞数が少ないサブセットが発現させている遺伝子は結果に反映されにくい。しかしながら細胞数が少なくても重要な役割を果たしている細胞集団がいると考えられており、一細胞単位でどのような遺伝子発現をしているかを調べ、今までのサンプル中の平均的な遺伝子発現量という検出方法では調べることができなかったサブセットとその機能についての研究が進んでいる。

## 7.2 プロモーターによる転写開始の仕組み

プロモーターは遺伝子の転写開始に関与する DNA の塩基配列のことを指し、転写開始部位と RNA ポリメラーゼとそれに付随するタンパク質が結合する配列を含んでいる。通常は遺伝子の上流 (5' 側) の領域にあり、配列は 100~1000bp ほどの長さを持っている [47]。RNA ポリメラーゼは DNA を mRNA に転写し、mRNA は最終的にタンパク質へと翻訳される。つまり、プロモーターはゲノム上において、遺伝子がどのように発現するかを制御している。プロモーターには、コアプロモーター (core promoter)、近位プロモーター (proximal promoter)、遠位プロモーター (distal promoter) の 3 種類が存在する [46]。コアプロモーターは、転写開始部位や RNA ポリメラーゼの結合部位がある領域のことを指す。近位プロモーターは、転写開始部位から 250bp ほど上流の領域であり、転写因子が結合する領域である。遠位プロモーターは近位プロモーターのさらに上流に位置するプロモーターであり、制御因子を多く含む。

プロモーターによる転写制御の仕組みは、細菌と真核生物で異なり、真核生物のほうが複雑なプロセスを実行する必要がある。細菌の転写の流れは以下の通りである。細菌の DNA には、転写開始部位から -10bp、-35bp の部分にプロモーターを認識させる配列が存在し、大腸菌の場合は 5'-TATATT-3' と 5'-TTGACA-3' である。また、細菌の RNA ポリメラーゼは、プロモーターの配列を認識する  $\sigma$  因子というサブユニットをもっている。まず、転写開始前の RNA ポリメラーゼは DNA に結合し、プロモーター領域に向かって移動する。次に RNA ポリメラーゼがプロモーター領域に出会うと、 $\sigma$  因子が -10bp、-35bp の配列を認識し、 $\sigma$  因子が RNA ポリメラーゼから遊離し転写が開始される。最後に、RNA ポリメラーゼがターミネーター配列を認識することで、RNA ポリメラーゼは DNA から離散して再び  $\sigma$  因子が結合する。このように、細菌のプロモーターは、-10bp、-35bp の配列を  $\sigma$  因子によって識別させるという単純な構造になっている。

一方で、真核生物では RNA ポリメラーゼに加えて、転写基本因子がプロモーターに結合することで転写が開始される。ここでは、RNA ポリメラーゼ II の転写について説明しよう。真核生物のプロモーターには TATA ボックスという塩基配列が存在する。まず、この配列を転写基本因子の一つである TFIID が認識することで、プロモーターとほかの転写基本因子である TFIIB や TFIIF などが結合する。次に、TFIIF が DNA の二重らせんを解くと同時に、RNA ポリメラーゼをリン酸化する。これにより、RNA ポリメラーゼは転写基本因子から遊離して転写を開始する。このように、真核生物のプロモーターによる転写制御は、転写基本因子によって複雑に制御されている。

## 7.3 エンハンサーによる転写制御

エンハンサー (enhancer) とは、特定の転写因子と結合して遺伝子の転写を促進するシス制御エレメント<sup>10</sup>であり、遺伝子発現の制御において重要な役割を果たす [48]。エンハンサーは、個体発生や細胞分化、各種細胞内における遺伝子発現の特異的な調整を可能にする。エンハンサーはプロモーターから数千塩基対離れた位置であっても機能し、DNA の折りたたみ構造により、エンハンサーとプロモーターが物理的に接触して転写を促進する [49]。したがって、エンハンサーは遺伝子の上流・下流、イントロン内部など、多様な場所に存在する可能性がある。さらに、エンハンサーは配列の

<sup>10</sup>シスエレメント (*cis*-element) またはシス制御エレメント (*cis*-regulatory element) は DNA や mRNA 等に存在する領域で、その領域と同じ分子上の遺伝子発現を調節する因子となるものである。

向き(順方向または逆方向)にかかわらず作用することができ、柔軟な調節因子として機能する。特定の領域がエンハンサーであるかどうかの推定は、当初はレポーター遺伝子の発現を用いて行われていたが、最近ではエンハンサー候補の特定は、DNase 高感受性<sup>11</sup>、トランスポザーゼ<sup>12</sup>のアクセシビリティ(accessibility, 接近可能性)<sup>13</sup>、双方向エンハンサー RNA(eRNA)<sup>14</sup>の転写、CpG の低メチル化<sup>15</sup>、高頻度のヒストン修飾 H3K27ac および H3K4me1<sup>16</sup>、配列特異的な転写因子の結合、共役因子の誘導等、クロマチンの特異的な特徴に基づくようになってきている [48]。

近年は次世代シーケンシング技術の進歩により、ゲノム全体でのエンハンサー配列の同定とその機能の解明が進み、エンハンサーがどのようにして遺伝子発現プログラムを時間的・空間的に制御しているのかに関する研究が進んでいる [49]。さらに、エンハンサーは異常が生じた場合に疾患の原因となりうるということが明らかになっており、特にがんに関連するエンハンサーの変異や転写因子との結合異常が注目されている [48]。

エンハンサーとプロモーターは、遺伝子発現の制御に関与する DNA 領域ではあるものの、異なる役割を持っている。プロモーターは遺伝子の転写開始点の比較的すぐ上流に位置し、RNA ポリメラーゼと基本的な転写因子が結合して転写を開始する部位のことである。一方、エンハンサーは比較的遠位に位置し、様々な転写因子と結合することで遺伝子発現を調整し、プロモーターから遠い位置でも遺伝子の発現を活性化させる。

## 7.4 マイクロアレイ

DNA マイクロアレイは特定の細胞や組織から抽出した RNA サンプルをもとに、数万もの遺伝子の発現状態を同時に調べることが可能である(図 24)[50, 51]。遺伝子発現を検出する原理としては、まず細胞や組織から RNA を抽出し(図 24(a))、その配列に相補的な、蛍光標識された核酸配列を合成する(図 24(b))。そして、標識された核酸配列をマイクロアレイ上のオリゴヌクレオチド(DNA プロブ、標識された核酸配列と相補的)とハイブリダイズさせる(図 24(c))。ハイブリダイゼーションの結果、サンプル中に多く存在する RNA に対応する標識済核酸配列は対応するプロブと大量に結合し、強い蛍光を発することになる。逆に少ない場合は、蛍光が弱くなる。蛍光の強さを数値化することによって、各遺伝子の発現量を数値化することができる。但し、この数値は、遺伝子の発現量だけでなく、核酸の標識の効率や、ハイブリダイゼーションのしやすさ等、多数の要因の影響を受ける。しかも、これら要因の多くは遺伝子によってその影響の大きさが異なる。従って、各遺伝子の発現量が、複数の異なる条件において、どのように変動しているかを調べるのには有用であるが、遺伝子間の発現量の網羅的比較や、細胞内 RNA の絶対量の網羅的解析には、普通不向きである。

(以下、内容確認・校正中) マイクロアレイの一種として、タイリングアレイという実験法がある [52, 53]。この実験は、解読済みのゲノムデータから塩基配列を等間隔で抜き出し、検出用プロブ

<sup>11</sup>DNase は DNA を切断する酵素。本酵素によって非常に切断されやすい領域ではクロマチン(真核生物の細胞核にある DNA とタンパク質(ヒストン)の複合体)が濃縮されておらず、DNA が露出している可能性が高い。

<sup>12</sup>DNA を別の場所に挿入する酵素

<sup>13</sup>アクセシビリティが高ければ DNA が露出している可能性が高い。

<sup>14</sup>エンハンサーで双方向に転写されることのある RNA。

<sup>15</sup>プロモーター領域の CpG 配列がメチル化されると遺伝子発現が抑制される傾向がある。

<sup>16</sup>転写活性があることを示すマーカーとなりうる。



として DNA チップをタイル状に並べることで、未知の RNA についても塩基配列の一部を知ることが可能になる実験手法である。このチップに、生体内で遺伝子として転写発現している RNA を鋳型とした標識サンプルをハイブリダイズすることで、シグナル情報が観測できる。抗体処理したサンプルと抗体未処理のサンプルを比べることによって、ゲノムの DNA 結合領域が割り出せるというものになっている。そして、どこの染色体に多くタンパク質が DNA に結合しているかがわかったら、染色体上のどの部分にどの遺伝子が存在するのか、というアノテーション情報を取得し、解析する必要がある。シグナルにはノイズも大きく、そしてプローブ間隔も広いといった問題点がある。

## 7.5 RNA-seq(内容確認・校正中)

転写産物である mRNA がどの程度存在しているかを調べる手法の一つとして、RNA-seq が挙げられる。RNA-seq では、

1. mRNA の精製
2. 逆転写酵素を用いての mRNA から相補的な cDNA (complementary DNA) の合成
3. アダプターの連結と PCR による cDNA の増幅
4. ハイスループットシーケンサーによる配列決定
5. リファレンスゲノムもしくは転写産物モデルへのアライメント

という過程を経て、元々の mRNA の存在量を求める [55]。

mRNA の精製の際、真核生物の場合には mRNA がポリ A 鎖による修飾を受けるため、これに相補的なポリ T オリゴヌクレオチドを用いて mRNA を単離する。一方、原核生物の場合にはポリ A 鎖の修飾がないため、rRNA に相補的な配列を用いて rRNA を除去することで精製を行う [56]。RNA の逆転写、PCR による増幅の過程では酵素を用いているために対象となる配列間で増幅バイアスが存在しており、正確性等を検証する研究も行われている [57, 71]。また、酵素による逆転写および増幅を行わず、ナノポア等を利用した RNA の直接シーケンシングも試みられている [78]。RNA-seq において読まれた配列は、一部のモデル生物を始めとした全長ゲノムが既知の生物の場合にはゲノムへとマッピングされ、シーケンサーが読んだリード数によって元々の mRNA の存在量が計算される。一方リファレンスゲノムが存在しない場合には、読まれた cDNA 配列同士をアセンブルしてつなぎ合わせることで転写産物モデルを構築し、その上で転写産物モデルにアライメントすることで転写産物量の推定を行う。

## 7.6 CAGE 法

CAGE (Cap Analysis of Gene Expression) 法とは真核細胞中の遺伝子発現パターンを解析する手法の一つで [81, 82]、理化学研究所にて開発された [80]。CAGE 法では、転写産物を捕らえるために、mRNA の 5' 末端に存在するキャップ構造を利用する。まず、解析対象の細胞中の mRNA に対してプライマー (ランダムプライマー) を使って cDNA を合成し、DNA-RNA ハイブリッド鎖を

形成させる。次に、mRNA のキャップのジオールを酸化してビオチン (biotin, ビタミンの一種) を付加して標識する。そして、結合活性の高いアビジン (avidin, タンパク質の一種) で被覆した磁気ビーズを使ってビオチンおよびハイブリッド mRNA を捕らえる。ここで RNA 鎖の方を分解し、2 本目の cDNA 鎖を合成し、これを配列決定 (シーケンシング) する。

RNA-seq 法の利用に合わせて改良された CAGE 法では、2 本目の cDNA 鎖合成前に制限酵素 EcoP151 の認識配列を含むオリゴヌクレオチドを 5' 端に連結しておく。これにより、EcoP151 によって制限部位から 27 塩基離れた位置で切断されるため、最初の mRNA の 5' 末端を含む 27 塩基長のタグができる (CAGE tag)。

すなわち、CAGE から得られるのは、短い断片配列 (CAGE tag) および各断片配列が観測された回数になる。これら断片配列を参照ゲノム配列上にマッピングすることにより、ゲノム上の転写開始部位を突き止めることができ、その下流に存在する遺伝子を調べることにより、解析対象の細胞中でその遺伝子がどれくらい発現していたかを推定することができる。

そして転写開始部位周辺の配列を調べることにより、特定の生物学的機能に関わる制御因子 (プロモータ等) を発見できる可能性がある。

## 8 タンパク質解析法

### 8.1 細胞内に存在するタンパク質を解析する意義

特定の細胞内にはどのようなタンパク質が存在するかを調べる意義は、節 7.1 で述べたような遺伝子発現 (どのような遺伝子が転写されているか) を解析する意義と重なる部分が多い。細胞は環境・状況に応じて、必要な遺伝子を適度に発現させているため、様々な遺伝子の発現パターンを解析することによって、遺伝子の機能を推定することが可能になる、というのが、遺伝子発現解析の意義であった。ただ、遺伝子の本来の機能はその遺伝子がタンパク質に翻訳されてから発揮されるという立場に立つならば、転写のパターンを解析するのでは実は必ずしも十分とは言えない側面がある。

まず、転写によって合成された mRNA が全て翻訳されてタンパク質になるわけではなく、細胞内の mRNA の量が本当に翻訳量を反映しているか、疑問の余地がある [88, 89]。また、mRNA のどの部分がタンパク質に翻訳されているのか、自明ではないケースも少なくない。さらに重要なことに、タンパク質は翻訳後にリン酸化やユビキチン化等の翻訳後修飾を受ける。このような修飾は mRNA を解析するだけでは分からない。

そこで、細胞内にはどのようなタンパク質が存在するかを解析する技術が開発されてきた。細胞内には雑多なタンパク質が存在するため、細胞内に存在するタンパク質を調べるにあたって、まずそれらのタンパク質を種類毎に分離する (separation) 技術が開発されてきた (図 25)。

### 8.2 SDS ポリアクリルアミドゲル電気泳動 (SDS-PAGE)

電気泳動は電荷 (+か-) を持つ物質に電気を流し移動させその移動距離により物質を分離させる実験手法である。おもにゲル等の網目状の物質中を移動させる。分子量や電荷の違いにより移動距離に違いが出るため試料の分離が可能となるものである。

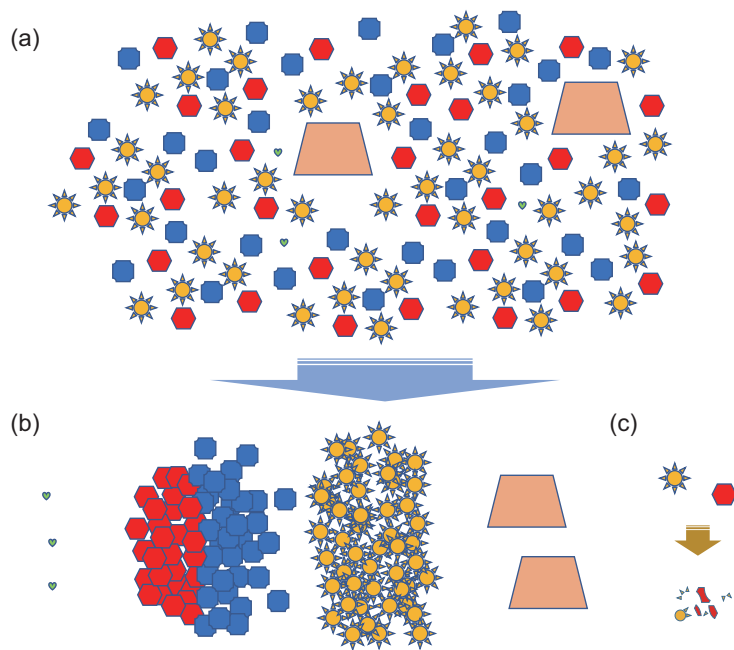


図 25: 細胞内タンパク質の実験的分離の模式図

一般的には細胞内には雑多なタンパク質が存在しているが (a)、それを分離 (b) することで、解析しやすくなる。(c) bottom-up アプローチでは、細胞から抽出した雑多なタンパク質集団を予め酵素によって断片化することに注意。



SDS とは Sodium dodecyl sulfate : ドデシル硫酸ナトリウムもしくは Sodium lauryl sulfate : ラウリル硫酸ナトリウムであり、この SDS とポリアクリルアミドゲルを利用した電気泳動を SDS-PAGE と呼ぶ [90]。SDS は負の電荷をもったタンパク質変成作用がある界面活性剤であり、タンパク質が高次構造やジスルフィド結合などをなるべく排除した状態で SDS と共に存在すると、主ペプチドに一定の割合で結合し一直鎖状の構造をとる。つまり負に電荷した一直線のペプチドができることになる。その状態でポリアクリルアミドゲルを用い電気泳動することによって分子量に応じてタンパク質を分離することが可能となる。

SDS-PAGE を行った後目的のタンパク質を疎水成膜 (メンブレン) に写し取り、抗体を用いてタンパク質を検出する方法がウェスタンブロッティングである [91]。SDS-PAGE 後のゲルに存在するタンパク質は負の電荷を持つ。そのゲルとメンブレンを重ね合わせ電圧をかけてやることでタンパク質をメンブレンに移動させる。これを「転写」と呼ぶ。そのメンブレンに対して一次抗体、二次抗体を使用し、検出する。一次抗体は目的のたんぱく質に特異的に反応し、二次抗体は一次抗体に特異的に反応する。二次抗体は酵素標識があるものでこの酵素反応による発光もしくは発色を検出して電気泳動結果のタンパク質のバンド (ゲルに現れる線) が出現する。

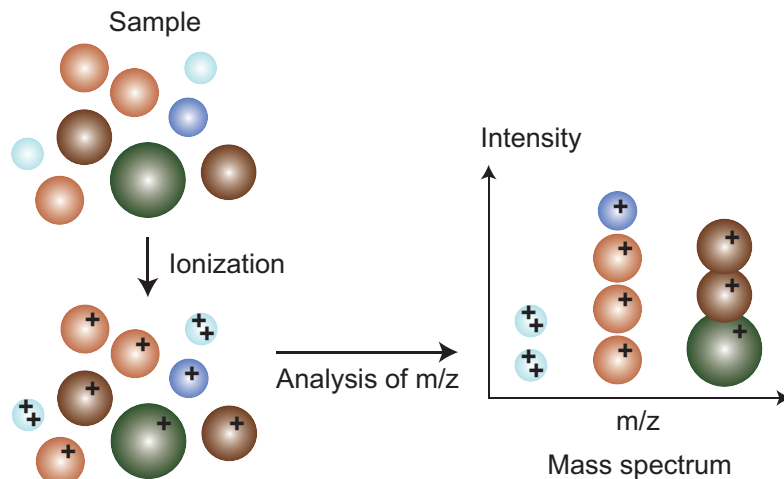
### 8.3 クロマトグラフィー

クロマトグラフィーでは、カラム内にある固定相の内部をキャリアーガスなどの移動相によってタンパク質が通るときに分離できる [92, 93, 94]。イオンクロマトグラフィーでは、タンパク質の表面電荷と pH が固有であることを利用している。イオン交換クロマトグラフィーには陰イオン交換体と陽イオン交換体があり、タンパク質が等電点と等しい pH になると電荷がゼロとなり、可溶化しなくなる。等電点  $pI$  より塩基性の pH に可溶化したタンパク質は負に荷電し、陰イオン交換体と結合する。逆に、 $pI$  より酸性の pH に似可溶化した他の悪は生に価電子、陽イオン交換体と結合する。この交換体 (担体) には様々な種類があるため、使い分けることで純度の高いタンパク質を得ることができるようになる。他にも、ゲル濾過カラムクロマトグラフィーでは、カラムに充填された担体が分子を分子量に従って分離することができる。サンプルをカラムに通すと、担体に開いている小さい孔に小さい分子ほど入り込み、高分子は早く溶出され低分子は遅くに溶出される。ゲル濾過クロマトグラフィーではタンパク質だけでなく、多糖類や核酸などの生体分子にも使われる。

高速液体クロマトグラフとは、試料を液体の流れに乗せ、カラムの通過速度により各成分に分離し、それらを検出するシステムである [95, 96]。送液ポンプ→試料導入装置→カラム (固定相) →検出器の順番で移動相 (液体) が流れる。移動相には水や有機溶媒が用いられ、送液ポンプにより一定の流量で送られる。試料導入装置でこの液体の流れに試料をのせ、カラムで分離する。カラムには、ステンレス製クロマトグラフィー管に用途に応じた大きさの充填剤を詰めたものを用いる。カラムで分離された各成分はカラムから溶出するが、試料の成分が固定相・移動相のどちらに親和性を持っているか (ここには各成分の固定相・移動相に対する吸着、分配、イオン交換、サイズなどの差が絡む) によって、溶出スピードが異なる。溶出した成分は、吸光光度検出器や蛍光検出器などで検出され、電気信号に変わる。検出されたデータからその濃度が測定され、分析結果が表示される。

## マススペクトル

質量分析は試料をイオン化し、イオン化された化合物群の  $m/z$  (質量・電荷比、イオンの質量をイオンの電荷で割った値)[106] を分析する方法である。質量分析装置からは、マススペクトルと呼ばれるデータが出力される。マススペクトルは、横軸に  $m/z$ 、縦軸に検出強度をとったグラフである。高い値を示す特定の  $m/z$  があれば、その  $m/z$  の値を持つ物質が存在することが推定される。



## 8.4 質量分析

### 8.4.1 概要

細胞内タンパク質を分析する実験的手法として質量分析法がある<sup>17</sup>[97, 98]。この方法では、MALDI(マトリックス支援レーザー脱離イオン化)法(節 8.4.2 参照)やエレクトロスプレーイオン化法などによりサンプル中のタンパク質あるいはタンパク質を消化して得られるペプチド断片をイオン化し、飛行時間型質量分析法(TOF-MS)や四重極型質量分析法(Q-MS)を利用して各イオンの質量電荷比( $m/z$ )を分析する(マススペクトル、コラム参照)。検出される  $m/z$  の値はタンパク質のアミノ酸配列に依存するため、試料中のタンパク質を同定するための情報が得られる。

質量分析の流れは大きく分けて、

1. 酵素消化：タンパク質を消化酵素を使ってペプチド断片にする(図 25c)。
2. MS/MS 分析：質量分析内に導入されたタンパクまたはペプチド断片のイオンを分子レベルで断片化する。

となる。このようにペプチド断片にしたり、分子レベルで断片化したりすることで、なるべく高感度に検出して、精密質量、アミノ酸配列、翻訳後修飾の種類・位置を解析することが容易になる。

後述するトップダウン法では 2 のみ、ボトムアップ法では 1 と 2 の両方を実施する(図 26)。

<sup>17</sup>質量分析法＝方法論。装置のことを表現する場合は「質量分析計」。

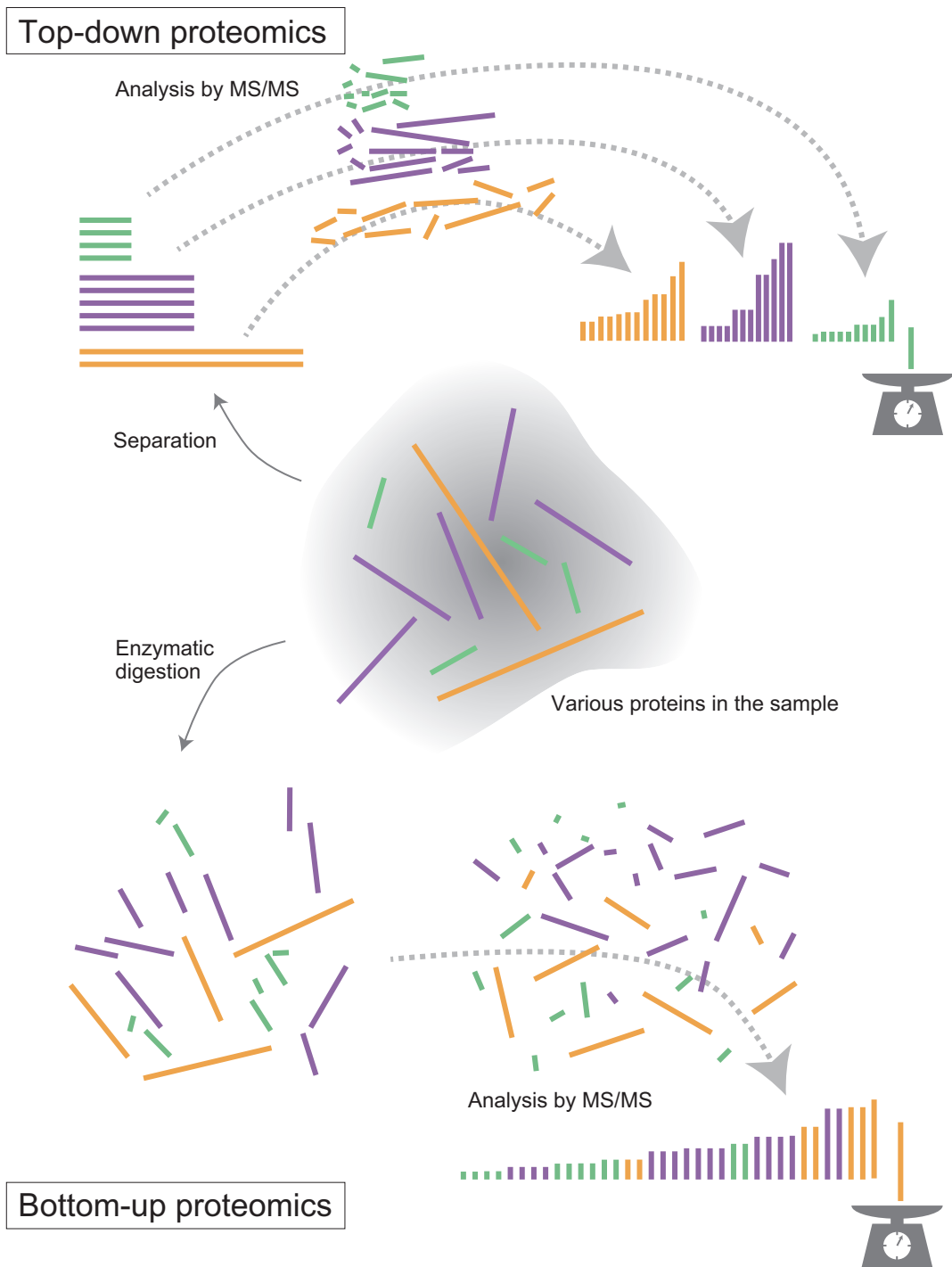


図 26: プロテオーム分析のトップダウン法とボトムアップ法

ボトムアップ法では、酵素によるタンパク質の消化 (digestion) を行う。ボトムアップ法においても、分離 (separation) が広く行われているが、図では省略。

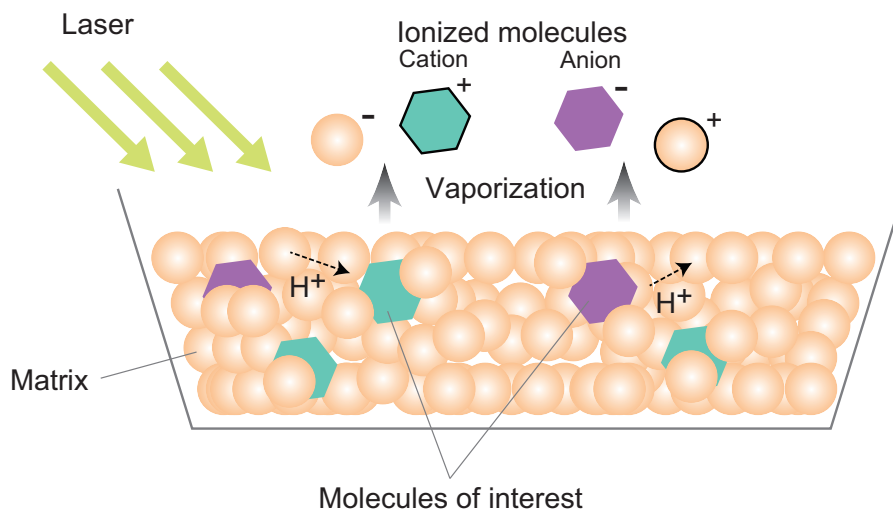


図 27: MALDI(マトリックス支援レーザー脱離イオン化)法の概要

分析対象の試料をシナピン酸等のマトリックスと混ぜ合わせ、そこに紫外線レーザーを照射する。するとマトリックスは励起・イオン化され、マトリックスから試料中の分子へ、あるいは試料中の分子からマトリックスへ電子が移動し、試料中の分子がイオン化される。この時、マトリックスはレーザーによって試料中の分子が破壊されるのを防いでいることにも注目されたい。さらにイオン化された分子は蒸発・気化され、TOF-MS等の質量分析装置を用いて分析が可能となる。

#### 8.4.2 MALDI(マトリックス支援レーザー脱離イオン化)法

MALDI (Matrix Assisted Laser Desorption / Ionization, マトリックス支援レーザー脱離イオン化)法は、分子量が数十万に及ぶタンパク質やペプチドなどの生体高分子を非破壊的にイオン化できる方法(ソフトイオン化法の一つ)であり、1980年代後半に開発された[99]。本手法では、まず試料をマトリックスと混合し、レーザー光(波長が337～355nmの紫外光)を照射する(図27)[99]。するとマトリックスは試料中の分子が破壊されるのを防ぎつつ、レーザー光を吸収し、熱エネルギーに変換する。これによってマトリックスの一部は急速に加熱され、試料とともにイオン化・気化される[100]。試料中のイオン化された分子は、TOF-MS等の質量分析装置によって分析が可能となる。

微生物のコロニーをマトリックスと混合し、分析することも可能である。MALDI-TOFMSを用いて、各微生物が持つ固有のタンパク質に対応するスペクトルパターンを検出できれば、微生物種を識別できるため[101]、臨床サンプルから直接細菌や真菌を同定するような応用もある[102]。

#### 8.4.3 MS/MS分析

MS/MS分析では、質量分析装置の不活性ガスを満たした空間内で全長タンパク質イオンを振動させてガス分子と衝突させながら断片化する。最適化されたある一定の範囲内のエネルギー条件で

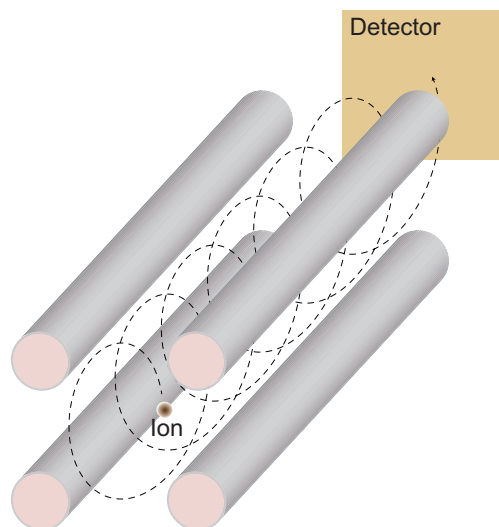


図 28: 四重極型質量分析計の原理の概略

は、ペプチド結合のどこか1箇所(時折2箇所)がランダムに切断される。各断片はN末端側に1級アミノ基、C末端側にKかRが1個ある状態になり、フラグメントイオンとして検出されてくる。

この断片化パターン(タンパク質のMS/MSスペクトル)のみからアミノ酸配列を読み取るのは無理である。そこで、タンパク質の三次元構造・アミノ酸配列と断片化のパターンにある程度の関係性があることを利用し、あらかじめタンパク質標品を分析して蓄積したスペクトルデータベースと実測データを照合して、定性解析する<sup>18</sup>。すなわち、検出されるフラグメントイオンの $m/z$ のパターンは、ペプチドのアミノ酸配列と翻訳後修飾の種類・位置に依存するので、測定で取得したMS/MSスペクトルのピークパターンを解析することで、元のペプチドの定性解析ができる<sup>19</sup>。

#### 8.4.4 四重極型質量分析計

四重極型質量分析器(quadrupole mass analyzer, QMS)は、4つの電極(四重極)によって電場を制御し、特定の $m/z$ の分子のみを通過させることで質量分析を行う装置である(図28)。それぞれの電極には、直流電圧と高周波交流電圧がかけられており、互いに向き合う電極には同じ極性の電圧、隣接する電極には異なる極性の電圧がかけられている[103]。このため、四重極の中では高速で位相が変化する電場が生じている。また、この時の電場の状態は直流と交流の電圧、交流の周波数によって決まる。

四重極の中をイオンが通過しようとするときイオンは上下左右に振動する。特定の $m/z$ をもつイオンは安定した振動を行い、四重極を通過して検出器に到達する。しかし、同じ条件でも、イオンの $m/z$ が電場に対して大きすぎる、小さすぎる場合には振幅が大きくなり電極に衝突してしまう

<sup>18</sup>この場合の定性解析の戦略はメタボロームのような低分子解析の場合と同様である。

<sup>19</sup>定性解析の詳細については、インターネットで、「ペプチドシーケンスタグ法」というキーワードで検索するとよい。

[104]。したがって、四重極の直流と交流の電圧、交流の周波数を調整することで、特定の  $m/z$  のみを検出器に送ることができる。

四重極型質量分析計は、他の質量分析計と比べて低い分解能と精度であるものの、小型でシンプル、低真空、安価、操作しやすい等利点が多く、最も使われている質量分析計の一つである [103, 105]。また、四重極型質量分析計を2つ用いるトリプル四重極質量分析計 (Triple quadrupole mass analyzers) も存在する。これにより、四重極型質量分析計の検出限界と精度を高めることができる [105]。

#### 8.4.5 トップダウン法とボトムアップ法

トップダウン法では、図 26 で示すように、LC(液体クロマトグラフィー) やゲル電気泳動で分離した全長タンパク質をイオン化して質量分析法で解析する。

ボトムアップ法では、タンパク質を酵素消化して得たペプチド断片を LC 分離後、イオン化して質量分析法で解析する。タンパク質と比較して、ペプチドは溶解しやすく、LC で比較的容易に分離可能で、かつ ESI でのイオン化効率も良いので、同じモル数ならトップダウン法より高感度に検出できる。ペプチドイオンであれば、MS/MS でランダムに切断されるペプチド結合を 1~2 箇所に抑えるようなエネルギー条件を設定でき、アミノ酸配列や翻訳後修飾の種類・位置の解析に適している<sup>20</sup>。

酵素消化で用いる酵素はいくつか種類があるが、K と R の C 末側のペプチド結合を選択的に切断するトリプシン (Trypsin) が汎用される。トリプシン消化で生じるペプチド断片は

- C 末端側に K か R が 1 個 (タンパク質の C 末除く)
- 配列の途中に K または R が数個程度
- N 末端側に 1 級アミノ基<sup>21</sup> が 1 個 (タンパク質の N 末除く)

という形になる。

質量分析でのイオン化では、N 末アミノ基と塩基性アミノ酸の側鎖に正電荷が乗るので、理論上は酵素消化で生じた全てのペプチド断片がイオン化されて MS で検出される。

#### 8.4.6 全長タンパク質のままでも質量分析は可能？

全長タンパク質のままでもイオン化は可能ではあるが、以下のようなデメリットもあることに注意しなければならない。

- 巨大分子になるほどイオン化の効率は低下するので、感度は稼ぎにくくなる<sup>22</sup>。

<sup>20</sup>ランダムにどこでも切れてしまうと MS/MS スペクトル (=ペプチド断片にエネルギーをかけて得られたフラグメントイオンのスペクトル) の解析が複雑になってしまい、元のペプチド断片のアミノ酸配列の予測が難しくなるため。シーケンスタグ法、*de novo* シーケンスタグ法の原理を理解したうえで考えるとわかりやすい。

<sup>21</sup>ペプチドの N 末端アミノ酸に由来するアミノ基。酵素消化により、タンパク質のペプチド結合 (アミド結合) の NH と CO の間で切断されるので、その結果生じるペプチド断片の N 末側はすべて一級アミノ基を 1 つ持つことになる。

<sup>22</sup>LC/MS で使用される ESI 法は低エネルギーのソフトイオン化法なので、タンパク内の共有結合を切断して分解する力はほぼない。ただし、非共有結合はもろいので、タンパク質の 2~4 次構造はイオン化の過程で崩れると考えられる。複数のサブユニットから成るタンパク質の場合、条件によっては各ユニットに分かれたものが検出される可能性もある。

- タンパクのような高分子は立体構造が一律でなく不安定なため、LCで再現性高くきれいに分離できないことが多い。
- LCを使う場合、サンプル溶媒やLCの移動相に界面活性剤を添加できない。そのため、LCシステム内で十分溶解した状態を維持できず、流路(カラム、配管)に吸着してサンプルが失われることもあり(サンプルロス)、感度低下につながる。
- ゲル電気泳動(SDS-PAGE、2D-PAGE)の場合は、自動化すれば比較的再現性よくタンパク質の分離まではできるが、細かいゲルの網目の中から分離したタンパク質を取り出す際の回収率は悪いので、感度低下につながる(そのため、ゲル電気泳動後はゲル内消化でペプチド断片化してから抽出し、ボトムアップ法で解析するのが実は一般的である)。
- TOF MSやOrbitrap MSのようなイオンを高分解能で検出できる装置を使えば、理論上はタンパク質の全長質量を検出可能ではある。但し、分子量が近接しているもの、翻訳後修飾のパターンだけが違うもの、が混ざってくるとスペクトルが重なって複雑になり、MSスペクトルから算出した全長質量の値は怪しくなってくるので、事前の分離精製レベルに大きく依存することになる。
- イオン化した際に乗る正電荷の数が不安定であり、また異なるタンパク質間での分子量差が大きすぎるため、MS/MSでランダムに切断されるペプチド結合を1~2箇所を抑えるようなエネルギー条件を設定できない。

## 8.5 免疫染色

細胞の免疫染色はタンパク質の細胞内局在を調べる際に有用である。質量分析計を使った解析のような網羅性の部分は弱点であるが、発現を見たいタンパク質の局在をピンポイントで確認することができる。まず、4%パラホルムアルデヒドやホルマリンなどを使用して細胞を固定する。これにより細胞の生化学反応を止め、サンプルの状態を保存する。続いて洗浄、界面活性剤を用いた膜透過処理をした後、ブロッキングを行う。ブロッキングには、抗体反応の特異性を高める効果がある。ブロッキングが終わったら、見たいタンパク質に特異的な抗体を添加し、1次抗体反応を行う。その後洗浄をし、1次抗体に特異的な、蛍光標識された2次抗体を添加する。最後に、DAPIなどの対比染色をするための物質を添加し、洗浄をして、蛍光顕微鏡で細胞を観察する。蛍光を放つ2次抗体は、1次抗体にのみ結合しており、1次抗体は目的のタンパク質にのみ結合しているため、蛍光があるところに目的のタンパク質が存在しているということになる。上記の手法は「間接法」を用いた場合であり、タンパク質特異的且つ蛍光標識されている抗体があれば、1次抗体を添加した時点で観察すれば良い。これを「直接法」という。

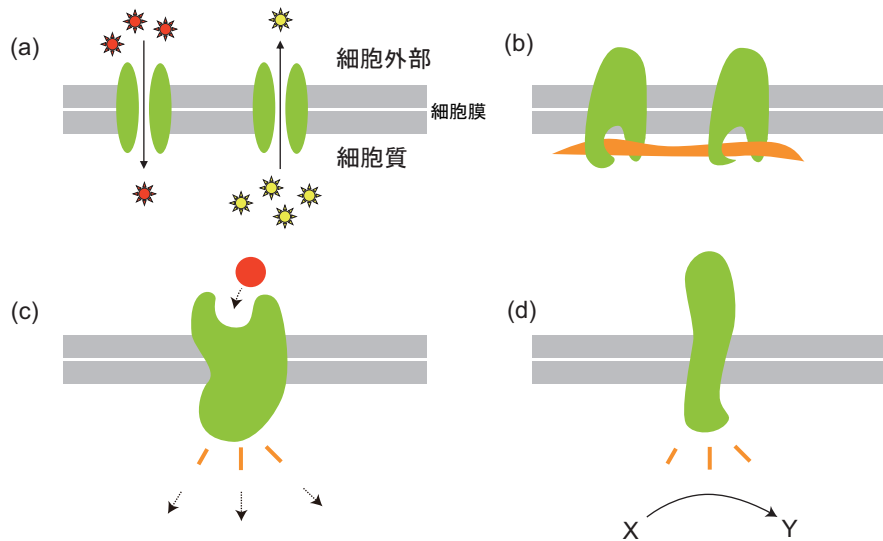


図 29: 膜タンパクの様々な機能

膜タンパク質が持つ機能のうち、主なものを4つ図示した。(a) 輸送体、(b) 連結体、(c) 受容体、(d) 酵素。細胞外から細胞内へのシグナル伝達の手段として、(c) は広く使われている。

## 9 細胞内分子間相互作用

細胞内には、核酸やタンパク質、脂質、その他代謝物質等、様々な分子が存在するが、多くの分子は細胞の中で常に単独で機能しているわけではなく、分子同士で様々な相互作用をする<sup>23</sup>。相互作用には、分子同士で集まって何か特定の機能を発揮するという「物的」な面と、ある分子から他の分子へシグナルを伝達するという「情報の」側面がある。

分子生物学の中でも、シグナル伝達は特に面白い。細胞は絶えず栄養状態や栄養源の有無など、その内外の環境を監視している。環境の変化を検知すれば、それに応じて遺伝子発現パターンを調整している。外界の変化に反応するという意味で、細胞が生きている証しである。

細胞外からのシグナルが細胞内に伝わる時は、何らかの分子あるいは情報が細胞膜を通過することになる。細胞膜には多様な種類の膜タンパク質が存在しており(図 29)、その一部は細胞外からのシグナルを細胞内に伝える役割を持つ。一般的には、細胞外から細胞内へのシグナル伝達は、シグナル因子となる分子が細胞膜上の受容体に結合することで起きる(図 30)。これによって、受容体の構造が変化し、細胞内に影響を及ぼすのである。細胞内では、典型的には主にタンパク質のリン酸化によってシグナルが伝わる。シグナル伝達経路上、上流のタンパク質が下流のタンパク質をリン酸化することが繰り返されることによって、シグナルが下流に次々と伝わっていくのである。

シグナル伝達経路の中には、JAK-STAT シグナル伝達経路のように、レセプターが転写因子を直接活性化する比較的単純な経路になっているものもある。JAK-STAT シグナル伝達経路は、胚の発

<sup>23</sup>多細胞生物の場合、細胞内だけでなく、細胞間でもシグナル伝達を行っている。その様式は大まかに内分泌型、傍分泌型(パラクリン型)、神経型、接触型の4つに分類される。



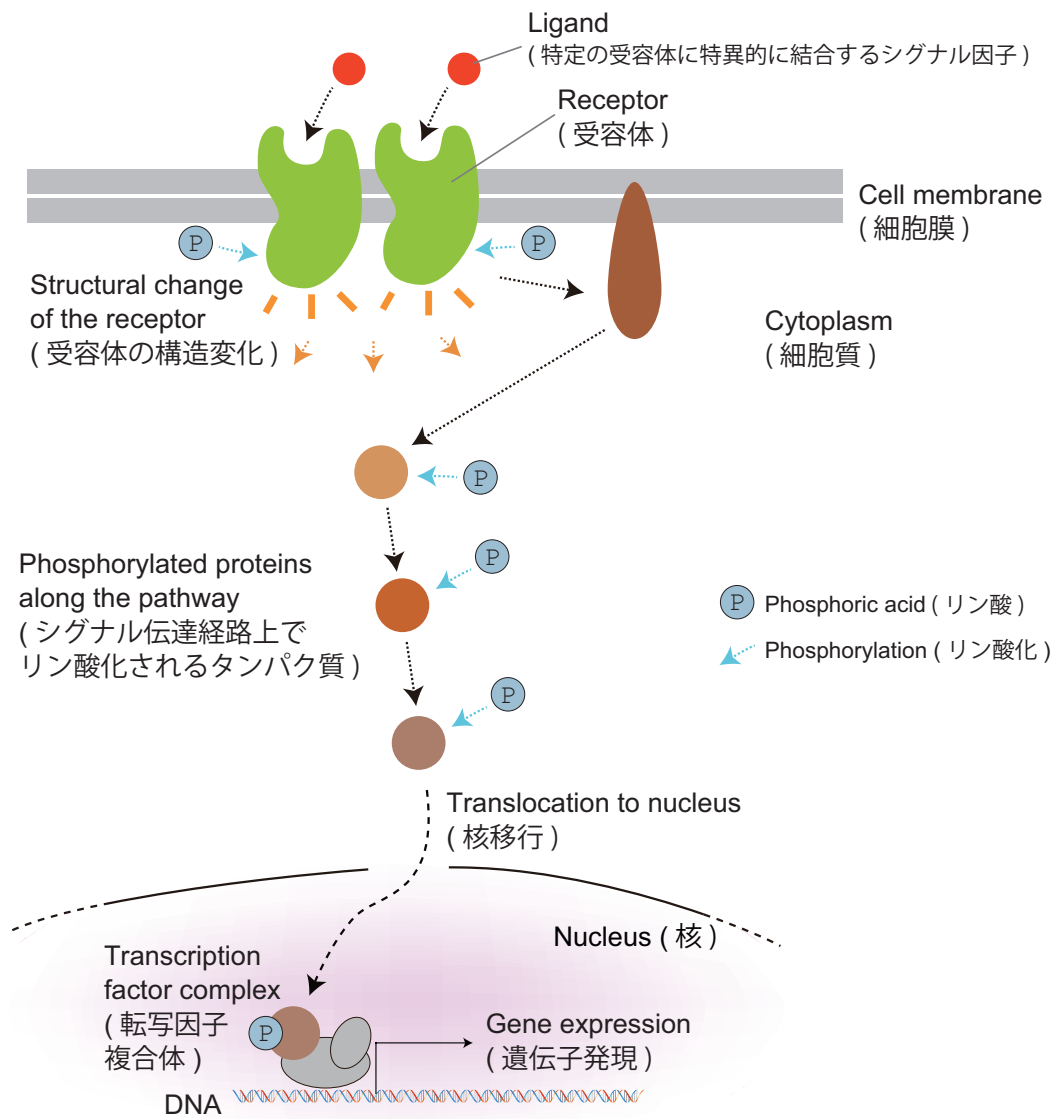


図 30: 細胞内シグナル伝達の概略の例

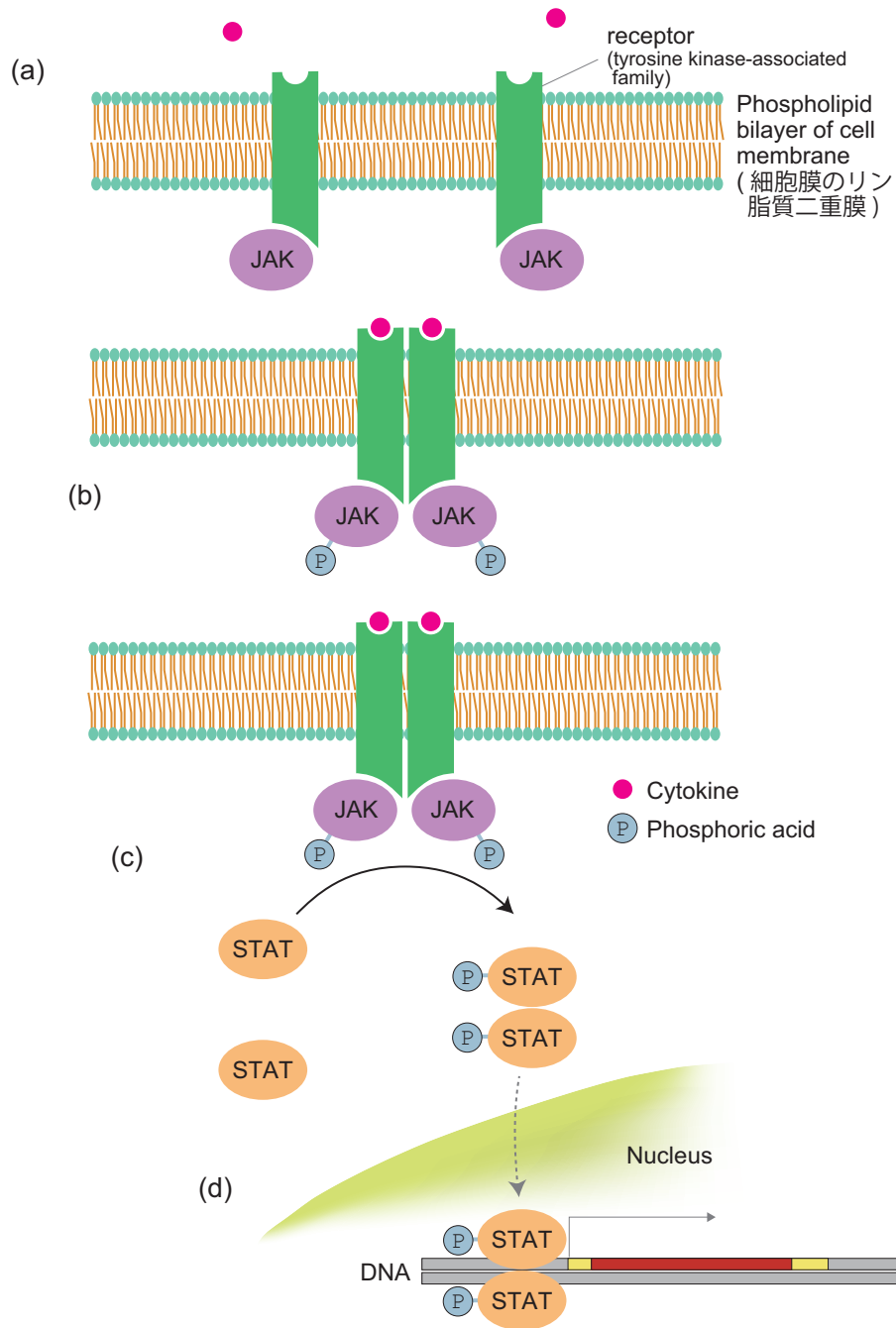


図 31: JAK/STAT パスウェイの概要

(a) レセプター (receptor) が JAK と結合している。(b) サイトカイン (cytokine) が結合すると、レセプター同士が二量体を形成し、JAK が互いをリン酸化し合う。(c) リン酸化された JAK は転写因子 STAT をリン酸化する。(d) リン酸化された STAT は核内に移行し、様々な遺伝子を発現させる。

生、組織の成長、造血器の発達と分化、自然免疫と適応免疫、血液がんなどを引き起こすこともある、細胞外からの情報を伝達する経路である(図 31)[122, 123, 124]。この経路は細胞表面受容体、Janus キナーゼ (JAK) および 2 つのシグナルトランスデューサーと転写活性化因子 (STAT: Signal Transducers and Activators of Transcription) タンパク質の主に 3 つの成分からなる。シグナル伝達には JAK チロシンキナーゼと 2 つ以上の受容体サブユニット細胞ドメインが結合している必要がある。膜貫通型受容体にサイトカイン (cytokine)<sup>24</sup> がリガンド (ligand) として結合すると、受容体のオリゴマー化によりそこに結合した JAK が、JAK キナーゼによる自己リン酸化やトランスリン酸化を可能にする。活性化された JAK は、受容体の細胞質領域にあるチロシン残基をリン酸化して、SH2 (Src homology 2) ドメインを有するタンパク質に結合部位を作製する。そして STAT は SH2 ドメインを介して受容体に結合し、STAT 自体が JAK の基質となる。SH2 ドメインと C 末端の特定部位の間にあるチロシン残基がリン酸化されると STAT のヘテロ二量体またはホモ二量体が形成され SH ドメインとの相互作用により安定化する。STAT 二量体はその後核へと移動し標的遺伝子のプロモーター領域内の特定配列に結合し、転写を開始する。

MAP キナーゼ経路はさまざまな生物種で多数見つかっているシグナル伝達である。パスウェイの上流にある Ras は細胞膜の内側に結合し、分裂促進因子受容体 (mitogen receptor) の近くに配置されている(図 32a)。GEF によって Ras に GTP が結合している状態になると、Ras は活性型となり、逆に GAP によって GDP が結合している状態になると、不活性型となる。このように活性型・不活性型のバランスは、GEF と GAP によって制御されている

分裂促進因子 (mitogen) が受容体に結合すると、受容体が二量体になり、自己リン酸化が起こる(図 32b-d)。次に SOS のような GEF が受容体に結合し、Ras の GDP を GTP に置き換える。活性化された Ras は Raf をリン酸化し、リン酸化された Raf は今度は Mek をリン酸化する(図 33)。すると、Mek は Map キナーゼ (MK) をリン酸化する。リン酸化により活性化された MAP キナーゼは核へ移動し、いくつもの転写因子などをリン酸化して活性化させる。つまり、この MAP キナーゼ伝達系の最後で MAP キナーゼは、下流にある転写調節因子などの様々なエフェクタータンパク質をリン酸化し、遺伝子の転写調節機能を変化させる。この遺伝子の発現パターンによって、細胞の増殖促進や分化の誘導などが起こる。また、リン酸化された MAP キナーゼは Rsk をリン酸化可能である。すると、Rsk は核内に移行し一連の転写因子を活性化する。

## 10 分子間相互作用解析法

### 10.1 実験的方法

#### 10.1.1 アフィニティークロマトグラフィー (内容確認・校正中)

アフィニティークロマトグラフィー (Affinity Chromatography) はタンパク質の持つ結合親和性を利用してタンパク質を分ける手法である [114, 115, 116]。各タンパク質は特定の物質に対して強い結合親和性を有しており、その物質をリガンドと呼ぶ。アフィニティークロマトグラフィーではリガ

<sup>24</sup>細胞から分泌される低分子のタンパク質で生理活性物質の総称。インターロイキン (interleukin) やインターフェロン (interferon, IFN) 等がある。インターロイキンは、ヘルパー T 細胞 (leukocyte から-leukin) によって分泌され、細胞間 (inter-) コミュニケーションの機能を果たすもの。インターフェロンとは動物体内で病原体 (特にウイルス) や腫瘍細胞などの異物の侵入に反応して細胞が分泌する蛋白質。

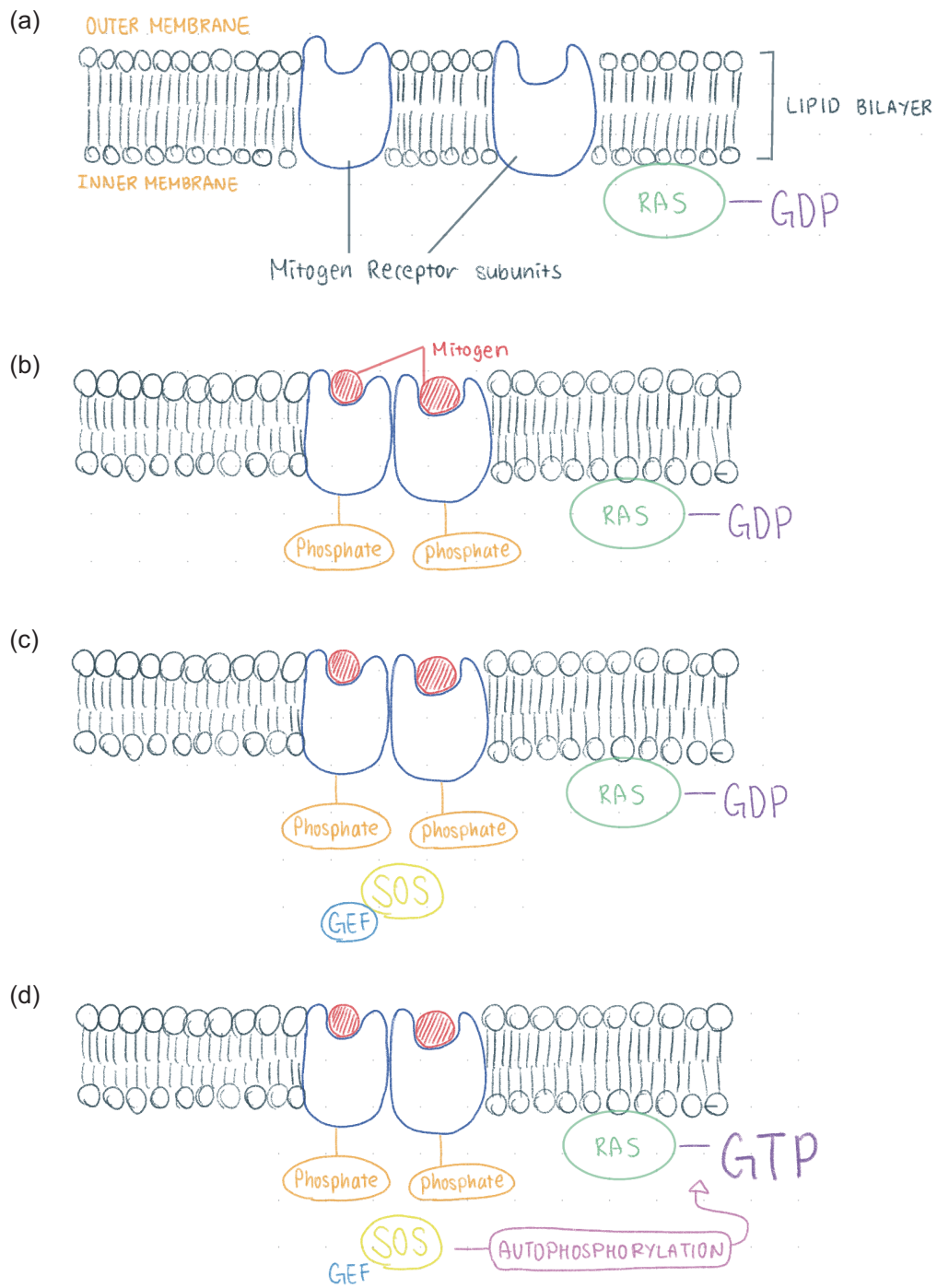


図 32: MAPK/ERK pathway の最初の段階

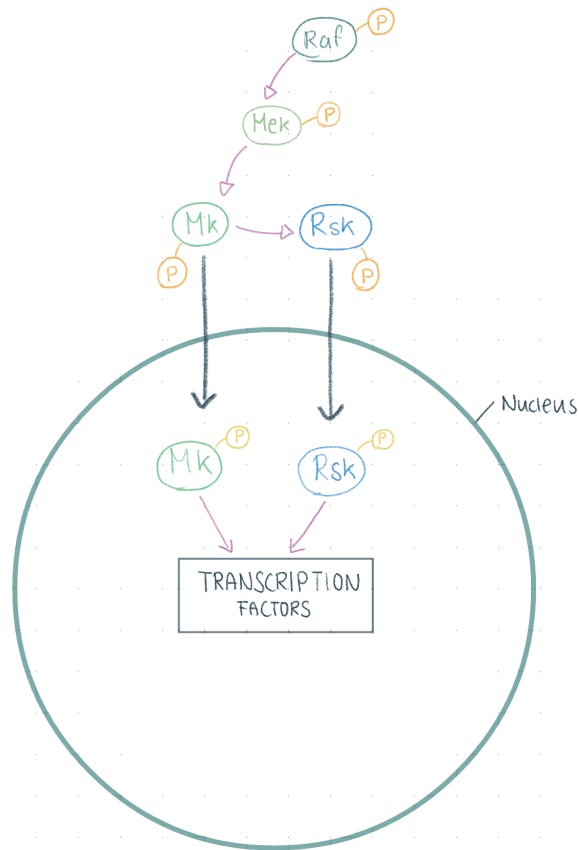


図 33: MAPK/ERK pathway - 遺伝子発現制御まで

ンドをカラム内に固定し、タンパク質溶液を通すことで、リガンドと相互的に結合するタンパク質のみを抽出することができる。なお、リガンドと結合したタンパク質は高濃度の塩により結合が解かれ、分離する。

### 10.1.2 共免疫沈降法 (内容確認・校正中)

目的物質と相互作用するタンパク質の調査に用いられる。共免疫沈降法では、目的物質および目的物質に結合することが知られる抗体を準備する。この抗体を磁気ビーズ等の担体で回収できるようにしておくことで、サンプル溶液中で目的物質と他のタンパク質とが相互作用して複合体を形成した場合に、抗体の回収と同時に複合体を丸ごとサンプル溶液中から分離することが可能である。利用する抗体や用いる担体の種類は様々存在する [107, 108]。

また、相互作用を調べたい目的物質がタンパク質でかつ複数存在する場合には、プルダウンアッセイと呼ばれる手法も存在する。この手法では共通のアミノ酸配列を持つポリペプチドタグを目的物質に付加する形で発現させることで、それぞれの目的物質に結合する抗体を準備することなく、ポリペプチドタグに結合する抗体のみで相互作用を調べることができる [109]。

### 10.1.3 酵母ツーハイブリッド法 (内容確認・校正中)

酵母 two-hybrid 法 (Y2H 法) は、酵母の GAL4 タンパク質の性質を利用し、酵母の生育・非生育により、2つの既知のタンパク質間の相互作用をスクリーニングする手法である [112, 113]。酵母の GAL4 タンパク質は、酵母が生育するために必要なガラクトース利用酵素をコードする遺伝子の発現に必要な転写活性化因子であり、DNA バインディングドメインと転写活性化ドメインから成る。GAL4 タンパク質はこれらのドメイン間に他のタンパク質が挟まれていても機能するが、これらのドメインが物理的に近い距離にない場合には機能しない。したがって、GAL4 の DNA バインディングドメインにタンパク質 X を、GAL4 の転写活性化ドメインにタンパク質 Y が結合したタンパク質を発現させると、タンパク質 X とタンパク質 Y との相互作用が存在する場合のみ、ガラクトース利用酵素をコードする遺伝子が転写され、酵母が生育する。

### 10.1.4 ファージディスプレイ法

ファージディスプレイ (phage display) とは、タンパク質同士の相互作用や DNA, RNA などの標的分子との相互作用を同定するために、使用される実験手法である。この手法は、特定の分子と結合するタンパク質を容易に精製できるため、医薬品や再生可能エネルギーの開発など、幅広い分野で活用されている。また、2018 年にはファージディスプレイの研究を行った George P Smith と Gregory P Winter がノーベル化学賞を受賞している [117]。

この方法には、バクテリオファージが用いられている。ファージはゲノム上に任意の遺伝子を組み込むと、そのペプチドやタンパク質を外殻タンパク質に融合させて提示する性質を持っている。ファージディスプレイでは、はじめにそれぞれのファージに異なる遺伝子を組み込み、複数のタンパク質を表面に提示したファージの集合体 (ファージライブラリー) を作成する [118]。ファージライブラリーの作成手順は以下の通りである [120]。

### 遺伝子の発現相関からネットワークへ (内容確認・校正中)

遺伝子の発現パターンから、単に遺伝子同士の機能の関連性を推定するだけでなく、ネットワークを推定する情報学的方法の開発も盛んに進められている。以下、ベイジアンネットワークを使った例を紹介しよう [54]。まず遺伝子 A, B の発現量の相関係数をとる。この時、正の相関や負の相関が現れたものは、グラフのエッジとして描画できる。これを、多くの遺伝子に対して行い、遺伝子の相関関係をグラフにする。しかし、これでは遺伝子 A と B の発現量に相関があった時、 $A \rightarrow B$  なのか  $B \rightarrow A$  の順番で反応が起きているのかがわからず、因果関係が掴めない。ここで、ベイジアンネットワークの考え方を適用する。つまり、重み付き有向グラフの重みが確率になったもので、遺伝子  $A \rightarrow B$  と  $A \rightarrow C$  という関係があった時に、A が発現したら B が発現する確率が何パーセントかという図式でネットワーク化する。しかし、この時の因果関係をネットワーク全体で推定する方法は完全には確立されていない。具体的には、実験データなどで機能を解明して補うか、DAG 探索問題に帰着させて既存のアルゴリズムで解いたり、Greedy Hill Climbing アルゴリズムと呼ばれる、貪欲法の一つの方法で解いたりする。つまり、描画したグラフの中で有向グラフがループしない条件下で尤もらしい形を探索するためのアルゴリズムである。そして、ネットワークを描画した後にネットワークのハブとなっている部分や、特定の遺伝子に相関を持った遺伝子群を検出することができるようになった。

まず、ランダムな DNA 配列や cDNA ライブラリーなどをファージ DNA に挿入する。次に、DNA を挿入したファージを大腸菌に感染させ、ファージを増殖させる。ファージは大腸菌を破壊することなく増殖し、ファージ粒子は細胞外へ放出される。ファージライブラリーを作成した後、ライブラリーの中から標的分子と結合するファージを選別する操作 (パニング) を行う [120]。まず、ファージライブラリーを標的分子が固定されているウェルに流し込み、標的分子とファージを結合させる。次に、Wash を行い、結合していないファージを洗い流し、選別を行う。最後に、標的分子と結合しているファージを溶出し、大腸菌を用いて再び増殖させる。この操作を複数回繰り返すことで、標的分子との結合があるタンパク質を高純度で精製することができる。また、ファージのゲノム配列を解析することで、このタンパク質のアミノ酸配列を同定することも可能である。

## 10.2 情報科学を用いた方法

### 10.2.1 遺伝子の発現相関 (内容確認・校正中)

類似の発現プロファイルを持つ遺伝子は相互作用するタンパク質をコードする可能性が高いことから、類似の発現プロファイルを持つ遺伝子をスクリーニングすることで、相互作用するタンパク質を情報科学的に予測することが可能である [119]。

### 10.2.2 系統プロファイル法

系統プロファイル法 (Phylogenetic profiling) は機能的に関連のあるタンパク質群を予測する方法である (図 34)。本手法では、まず、種ごとにゲノム配列中に各タンパク質に対応する遺伝子配列が

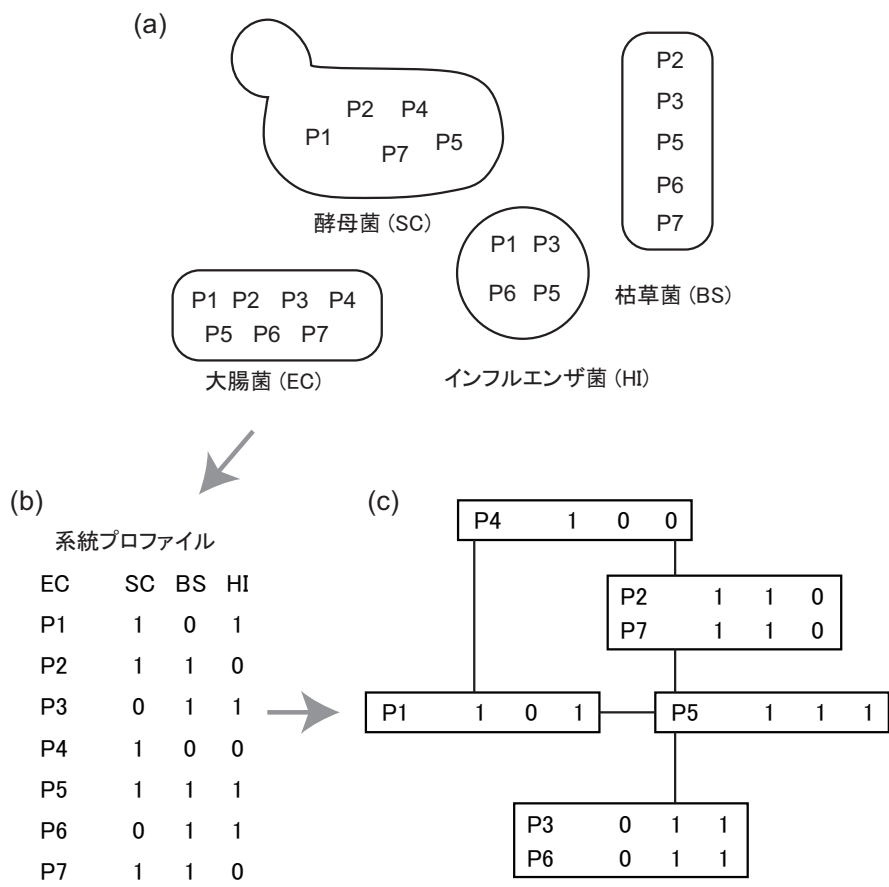


図 34: 系統プロファイル法の概要

(a) 4つの生物種 (SC, EC, HI and BS) が持つタンパク質の模式図。注目しているタンパク質は、P1, P2, ..., P7 の 7 種類である。例えば、種 HI は、注目している 7 種類のタンパク質のうち、P1, P3, P5 および P6 の 4 種類のタンパク質を持つ。(b) 各タンパク質を各生物種が持っているかどうかを表した表。これを系統プロファイル (phylogenetic profile) という。0 は持っていない、1 は持っているを表す。(c) 同じプロファイル (0 と 1 のパターン) を持つタンパク質群をグループ化し、1 つ違いのプロファイルを持つタンパク質群のペアを線でつないだ図。この例では、P2 と P7 および P3 と P6 の間に機能的なつながりがあると推測される。(文献 [110] の図を改変。)



コードされているか確認する。そして、各タンパク質に対して、各生物種における存在の有無、すなわちプロファイルを作成する(図 34(b))。そして、プロファイルの類似性が強いタンパク質群に機能的関連があると予測する。この手法には例えば次のような理論的根拠がある。ある時点で種 X に機能  $\alpha$  に必要な P1 と P2 というタンパク質が存在したとする。その後の分子進化で P1 というタンパク質が種 X から失われたとすると、機能  $\alpha$  は種 X からなくなる。P2 が機能  $\alpha$  のみに必要なタンパク質だとすれば、機能  $\alpha$  が消滅した後は、P2 は不要になり、自然淘汰されやすくなる。こうして、機能的に関連があるタンパク質群はある種には全て存在し、また別の種には全く存在しないというプロファイルができあがる。

### 10.2.3 オペロン解析 (内容確認・校正中)

共通のオペレーター配列によって同時に調節される遺伝子群であるオペロン (転写単位) の予測について概説する。オペロンは原核生物においてよく見られる遺伝子構造であり、1つのオペロンに含まれる遺伝子は転写のタイミングが同じため、ある生命現象に同時に関わっているケースが多い。よってオペロンの予測はすなわち機能的に関連のある遺伝子群の予測を意味する。オペロン予測は、ある生物種で過去の実験により 1 転写単位中に含まれることが既知の遺伝子群が別の種でも保存されているかを調べることで予測する手法の他、遺伝子間距離を指標とする手法、実験とは独立にオーソログ解析を行って生物種間での遺伝子の共通性からオペロン予測を行う手法等が存在する [111]。

## 11 ゲノム科学の進歩が生活・社会に与える影響

超高齢化社会になってゆくわが国において、医療およびその費用は重要な問題である。ゲノム科学はそのような問題の緩和に役立つことが期待される。例えば薬剤の使用に関しては、効果、副作用、費用等が問題となるが、人はそれぞれの体質 (アレルギー等)・生活習慣 (飲酒・喫煙等)・環境などの因子が異なるため同じ薬でも、効き方が異なる。さらにこれらの因子と遺伝子は複雑な相互作用をしている。

そこで、ゲノム情報から薬の効き方が予測できれば、その人に一番あった治療法を行い、また、無駄な薬の投与による無駄な治療費を抑えることが可能であろう。実際、ワルファリンという抗凝固薬には SNP によって薬の投与の適量が変わることがわかっている [45]。

すなわち理想的には、ゲノム科学が一人一人にあった治療法を見出し、医療費を抑えながらも国民の生活水準を維持できる。

現在シーケンシング技術の向上に伴い、個人単位で広い範囲の SNP (一塩基多型) を特定することが可能になりつつある。さらに特定の疾患や形質と SNP との対応を記録することで、各個人の DNA 配列を過去のデータと照合し、疾患リスクや適切な薬剤を推定することが可能になりつつある。これがより社会に広まることで、各個人に最適な医療の提示を行えるようになることが期待される。

ゲノム解析によって、がん、生活習慣病、老人性痴呆症、アレルギー性疾患などの複雑なヒト疾患の原因遺伝子群が解明され、遺伝子ネットワークや遺伝子システムの機能予測が可能になり、そ

れを基盤にした疾患の新しい診断、治療、予防法の開発が可能となる。さらに、個人差や人種差がゲノムの多型として明らかにされ、このゲノム特性に応じたヘルスケア等が現実になるであろう。また、病原性や病因に関するゲノム情報に基盤をおいた新しい創薬原理の開発も可能となる。

健康診断の一環としてゲノムを読みとる時代が来るかもしれない。

ゲノム科学がもたらしうる未来の対象はヒトの医療だけに限ったことではない。他の生物種の研究の成果が、農業、畜産、食料・新素材・環境問題などグローバルな課題の解決につながり、また、新産業が創出されることが期待される。

例えば、有用細菌、栽培植物、家畜のゲノム等の解析によって、従来の遺伝子組換え技術による品種改良のレベルをはるかに超えて、真に有用で、安全な遺伝子システムを用いた応用が農業はもとより、遺伝子資源の確保と利用、生物エネルギーの開発、環境保全・改良事業等、人類と地球規模の課題の解決に貢献すると共に、ゲノム産業ともいべき新産業が誕生するであろう。生物・医学関連のあらゆる科学と産業の基盤となる巨大なゲノム研究を短期および長期に見通して、その発展段階と、時代の社会的要請に応じて、多元的な戦略をたてることが重要である。

## 参考文献

- [1] Brown TA "Genomes 4" Garland Science 2017
- [2] Alberts et al. (著), 中村桂子 (翻訳), 松原謙一 (翻訳) 「細胞の分子生物学 第6版」 ニュートンプレス 2017
- [3] Chimpanzee Sequencing and Analysis Consortium (2005) **Initial sequence of the chimpanzee genome and comparison with the human genome.** *Nature* 437(7055):69-87
- [4] 「メダカからヒトゲノムの謎を学ぶ」 染色体遺伝子検査の分かりやすい説明ガイドライン III 平成 21 年 3 月 1 日 日本染色体遺伝子検査学会
- [5] 宮田隆 「分子進化学への招待」 講談社, 1994
- [6] Wolpert L (著), Tickle C(著), 武田洋幸 (翻訳), 田村宏治 (翻訳) 「ウォルパート発生生物学」 メディカルサイエンスインターナショナル 2012
- [7] Lascarez-Lagunas L, Martinez-Garcia M, Colaiacovo M (2020) SnapShot: Meiosis - Prophase I. *Cell* 181(6):1442.e1
- [8] [http://nature.cc.hirosaki-u.ac.jp/lab/3/animsci/text\\_id/Linkage.html](http://nature.cc.hirosaki-u.ac.jp/lab/3/animsci/text_id/Linkage.html) 「遺伝子の連鎖と遺伝子地図」
- [9] 八谷剛史, 入江直樹, 荒川和晴 「次世代シーケンサーの有効活用術」 *実験医学* 29(11): 1789-1796 2011
- [10] 荒川和晴 「ナノポアシーケンサーがもたらす未来」 *実験医学* 36(1): 2 -8 2018
- [11] Treangen TJ, Salzberg SL (2011) **Repetitive DNA and next-generation sequencing: computational challenges and solutions.** *Nat Rev Genet* 13(1):36-46
- [12] 河野暢明, 荒川和晴 「解読困難な長鎖リピート配列の解析に挑む クモ糸遺伝子を例に」 *実験医学* 36(1): 49-53, 2018
- [13] Compeau PE, Pevzner PA, Tesler G (2011) **How to apply de Bruijn graphs to genome assembly.** *Nat Biotechnol* 29(11):987-91
- [14] 板谷光泰, 柘植謙爾 「長鎖 DNA の合成と合成生物学での活用」 *生物工学会誌* 91 巻 6 号 p.319-321
- [15] <http://dna.brc.riken.jp/ja/manual/technicalnotes12> (2017/5/18 閲覧) 理化学研究所 遺伝子材料開発室 「PCR の話し」
- [16] [https://jp.illumina.com/content/dam/illumina-marketing/apac/japan/documents/pdf/2014\\_techsupport\\_session6.pdf](https://jp.illumina.com/content/dam/illumina-marketing/apac/japan/documents/pdf/2014_techsupport_session6.pdf) 米田瑞穂 「NGS をはじめよう！ これだけは知っておきたい MiSeq ～解析原理と必要な試薬キット、装置の使い方～」 イルミナ株式会社 (2014)

- [17] <http://infobio.co.jp/?portfolio=次世代 dna シーケンシングの原理>
- [18] [http://kenkyo.bureau.tohoku.ac.jp/terea-tsc/wp-content/uploads/2011/02/SOLiDsystem\\_Web.pdf](http://kenkyo.bureau.tohoku.ac.jp/terea-tsc/wp-content/uploads/2011/02/SOLiDsystem_Web.pdf) 「SOLiD システム紹介 ～ 原理・アプリケーション～」 ライフテクノロジーズジャパン株式会社 (2010)
- [19] <http://www.k.u-tokyo.ac.jp/omics/solid.html> ABI SOLiD
- [20] 上田哲也, 狩長亮二 「DNA 配列解析の前処理技術」 (超) 微量成分・不純物の同定・定量ノウハウ -試料前処理からデータ解釈まで- 技術情報協会出版, p816-818, 2014
- [21] [https://www.hssnet.co.jp/2/2\\_3\\_10\\_1.html](https://www.hssnet.co.jp/2/2_3_10_1.html) 「次世代シーケンス概要」 北海道システム・サイエンス株式会社
- [22] 大場利治 (2017) 「次世代を超えた DNA シーケンス技術」 生物工学 95(9):543-546
- [23] [http://catalog.takara-bio.co.jp/PDFFiles/52\\_32-33.pdf](http://catalog.takara-bio.co.jp/PDFFiles/52_32-33.pdf) 「世代シーケンサーによる高速シーケンス解析サービス」 タカラバイオ株式会社
- [24] <https://lne.st/2012/12/04/1737/> 「多様化する次世代シーケンスの現状」 株式会社リバネス
- [25] <https://youtu.be/Nj9EX0XI1Nk> PacBio シークエンシング原理
- [26] <https://nanoporetech.com/jp/how-it-works> Oxford NANOPORE Technologies シークエンシングの仕組み
- [27] 谷口正輝 「1 分子シーケンサーの現状と可能性」 日本農芸化学会 化学と生物 54: 6, 2016
- [28] Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML, Zhang YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC, Zhang YZ (2020) **A new coronavirus associated with human respiratory disease in China.** *Nature* 579(7798):265-269
- [29] <http://jsv.umin.jp/news/news200210.html> 日本ウイルス学会 「新型コロナウイルス感染症について」 2020 年 2 月 10 日
- [30] Imamura T, Isozumi N, Higashimura Y, Ohki S, Mori M (2021) Production of ORF8 protein from SARS-CoV-2 using an inducible virus-mediated expression system in suspension-cultured tobacco BY-2 cells. *Plant Cell Rep* 40(3):433-436
- [31] Al-Qaaneh AM, Alshammari T, Aldahhan R, Aldossary H, Alkhalifah ZA, Borgio JF (2021) Genome composition and genetic characterization of SARS-CoV-2. *Saudi J Biol Sci* 28(3):1978-1989
- [32] Hu B, Guo H, Zhou P, Shi ZL (2021) Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol* 19(3):141-154

- [33] Hamashima K, Fujishima K, Masuda T, Sugahara J, Tomita M, Kanai A (2012) **Nematode-specific tRNAs that decode an alternative genetic code for leucine.** *Nucleic Acids Res* 40(8):3653-62
- [34] <https://www.s.u-tokyo.ac.jp/ja/story/newsletter/keywords/04/03.html> 東京大学「RNA 干渉」
- [35] 安田純「RNAi」分かる実験医学シリーズ - RNA が分かる p.102-106 羊土社 (2003)
- [36] [https://www.cosmobio.co.jp/product/detail/crispr-cas.asp?entry\\_id=14354](https://www.cosmobio.co.jp/product/detail/crispr-cas.asp?entry_id=14354) (2018/6/17 閲覧) 特集：CRISPR-Cas9 とは - DNA 二本鎖を切断してゲノム配列の任意の場所を削除、置換、挿入することができる新しい遺伝子改変技術
- [37] <https://www.cosmobio.co.jp/support/technology/a/genome-editing-donor-vectors-gcp.asp> ドナーを利用した効果的なゲノム編集アプリケーション
- [38] <http://www8.cao.go.jp/cstp/tyousakai/life/haihu94/shiryo2-2.pdf> (2018/6/17 閲覧) CRISPR/Cas9 の作用原理
- [39] <https://www.funakoshi.co.jp/contents/8508> (2018/6/17 閲覧) guide RNA によるゲノム編集ツール CRISPR-Cas9 システム
- [40] <https://wired.jp/2018/02/04/whats-next-for-crispr/>
- [41] Committee on Human Gene Editing: Scientific, Medical, and Ethical Considerations **Human Genome Editing. Science Ethics, and Governance.** The National Academy Press, 2017
- [42] 解禁・ゲノム編集 出番待つマダイやトマト日本経済新聞 (2019/03/26)
- [43] <https://newswitch.jp/p/10651>
- [44] <https://www.genengnews.com/news/crispr-cas9-gene-editing-used-to-engineer-primate-model-of-autism/>
- [45] [http://www.riken.jp/pr/press/2013/20130624\\_2/](http://www.riken.jp/pr/press/2013/20130624_2/) アフリカ系米国人で抗凝固薬の効きやすさに関連する遺伝子多型を発見 - オーダーメイド医療の実現に向け抗凝固薬療法の個別化へ前進 - 理化学研究所
- [46] <https://www.addgene.org/mol-bio-reference/promoters/> Promoters
- [47] LaFleur TL, Hossain A, Salis HM (2022) **Automated model-predictive design of synthetic promoters to control transcriptional profiles in bacteria.** *Nat Commun* 13(1):5159
- [48] Arnold M, Stengel KR (2023) **Emerging insights into enhancer biology and function.** *Transcription* 14(1-2):68-87

- [49] Panigrahi A, O'Malley BW (2021) **Mechanisms of enhancer action: the known and the unknown.** *Genome Biol* 22(1):108
- [50] [https://www.chem-agilent.com/pdf/200810\\_WhatIsMicroarrayJ.pdf](https://www.chem-agilent.com/pdf/200810_WhatIsMicroarrayJ.pdf) マイクロアレイとは?
- [51] <https://www.learningatthebench.com/start-microarray.html> マイクロアレイとは? マイクロアレイを用いた発現解析の原理と解析方法
- [52] [https://www.molsis.co.jp/wp-content/themes/molsis/pdf/partek\\_200701.pdf](https://www.molsis.co.jp/wp-content/themes/molsis/pdf/partek_200701.pdf) Partek - 解析事例紹介: Affymetrix タイリングアレイによる発現解析
- [53] [http://www.jst.go.jp/nbdc/bird/jinzai/literacy/streaming/h22\\_2\\_1.pdf](http://www.jst.go.jp/nbdc/bird/jinzai/literacy/streaming/h22_2_1.pdf) 油谷幸代 「トランスクリプトーム解析・プロテオーム解析入門」
- [54] <http://www.scls.riken.jp/wp-content/uploads/2014/10/141028kobeuniv.pdf> 土井淳 「計算生命科学の基礎: 1.3 遺伝子ネットワーク解析」
- [55] [https://jp.illumina.com/content/.../2018-webinar-rna-seq%20design-session1\\_20180131.pdf](https://jp.illumina.com/content/.../2018-webinar-rna-seq%20design-session1_20180131.pdf) 寺倉 伸治 「実験デザイン編 これから RNA-Seq を始める方に」
- [56] <https://jp.illumina.com/content/.../2018-ilmn-webinar-Bacteria-RNA-Seq-Workflow-180829.pdf> バクテリア RNA-Seq をはじめよう
- [57] 城口克之 「RNA シークエンシング」 *生物物理* 53(6):290-94, 2013
- [58] Fiers W, Contreras R, Duerinck F, Haegeman G, Iserentant D, Merregaert J, Min Jou W, Molemans F, Raeymaekers A, Van den Berghe A, Volckaert G, Ysebaert M (1976) **Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene.** *Nature* 260(5551):500-7
- [59] Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, Hutchison CA, Slocombe PM, Smith M (1977) Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* 265(5596):687-95
- [60] Maxam AM, Gilbert W (1977) **A new method for sequencing DNA.** *Proc Natl Acad Sci U S A* 74(2):560-4
- [61] Gupta N and Verma VK (2019) **Next-Generation Sequencing and Its Application: Empowering in Public Health Beyond Reality** *Microbial Technology for the Welfare of Society* 17:313-341
- [62] Heather JM, Chain B (2016) **The sequence of sequencers: The history of sequencing DNA.** *Genomics* 107(1):1-8

- [63] Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM, et al. (1995) **Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd.** *Science* 269(5223):496-512
- [64] Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG (1996) **Life with 6000 genes.** *Science* 274(5287):546, 563-7
- [65] Lander ES et al(2001) **Initial sequencing and analysis of the human genome.** *Nature* 409(6822):860-921
- [66] Venter JC et al (2001) **The sequence of the human genome.** *Science* 291(5507):1304-51
- [67] Chaisson MJ, Huddleston J, Dennis MY, Sudmant PH, Malig M, Hormozdiari F, Antonacci F, Surti U, Sandstrom R, Boitano M, Landolin JM, Stamatoyannopoulos JA, Hunkapiller MW, Korlach J, Eichler EE (2015) **Resolving the complexity of the human genome using single-molecule sequencing.** *Nature* 517(7536):608-11
- [68] Cortese A, Simone R, Sullivan R, Vandrovcova J, Tariq H, Yau WY, Humphrey J, Jaunmuktane Z, Sivakumar P, Polke J, Ilyas M, Tribollet E, Tomaselli PJ, Devigili G, Callegari I, Versino M, Salpietro V, Efthymiou S, Kaski D, Wood NW, Andrade NS, Buglo E, Rebelo A, Rossor AM, Bronstein A, Fratta P, Marques WJ, Zuchner S, Reilly MM, Houlden H (2019) **Biallelic expansion of an intronic repeat in *RFC1* is a common cause of late-onset ataxia.** *Nat Genet* 51(4):649-658
- [69] Hunter JE, Berry-Kravis E, Hipp H, Todd PK (1993) **FMR1 Disorders.** In: Adam MP, Feldman J, Mirzaa GM, et al., eds. GeneReviews. University of Washington, Seattle 1993
- [70] 武田健一「DNA シーケンサ」特別 WEB コラム『新型コロナウイルス禍に学ぶ応用物理』公益社団法人 応用物理学会 [https://www.jsap.or.jp/columns-covid19/covid19\\_2-4-1](https://www.jsap.or.jp/columns-covid19/covid19_2-4-1)
- [71] SEQC/MAQC-III Consortium (2014) **A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium.** *Nat Biotechnol* 32(9):903-14
- [72] Levene MJ, Korlach J, Turner SW, Foquet M, Craighead HG, Webb WW (2003) **Zero-mode waveguides for single-molecule analysis at high concentrations.** *Science* 299(5607):682-6
- [73] Zhu P, Craighead HG (2012) **Zero-mode waveguides for single-molecule analysis.** *Annu Rev Biophys* 41:269-93
- [74] Eid J et al. (2009) **Real-time DNA sequencing from single polymerase molecules.** *Science* 323(5910):133-8

- [75] Rhoads A, Au KF (2015) **PacBio Sequencing and Its Applications**. *Genomics Proteomics Bioinformatics* 13(5):278-89
- [76] Iizuka R, Yamazaki H, Uemura S (2022) **Zero-mode waveguides and nanopore-based sequencing technologies accelerate single-molecule studies**. *Biophys Physicobiol* 19:e190032
- [77] Arnaud CH **DNA Sequencing: Zero-Mode Waveguides Turn 10** <https://cen.acs.org/articles/91/i51/DNA-Sequencing-Zero-Mode-Waveguides.html> Chemical & Engineering News, Dec 23, 2013
- [78] Garalde DR, Snell EA, Jachimowicz D, Sipos B, Lloyd JH, Bruce M et al (2018) **Highly parallel direct RNA sequencing on an array of nanopores**. *Nat Methods* 15(3):201-6
- [79] 宮澤光太郎, 松浦裕一, 池田圭吾, 岩丸祥史 (2022) 「ナノポアシーケンサーを用いた豚熱ウイルス国内流行株と経口生ワクチン株の迅速識別法」日本獣医師会雑誌 75(5):e83-90
- [80] Shiraki T, Kondo S, Katayama S, Waki K, Kasukawa T, Kawaji H, Kodzius R, Watahiki A, Nakamura M, Arakawa T, Fukuda S, Sasaki D, Podhajaska A, Harbers M, Kawai J, Carninci P, Hayashizaki Y (2003) **Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage**. *Proc Natl Acad Sci U S A* 100(26):15776-81
- [81] Takahashi H, Kato S, Murata M, Carninci P (2012) **CAGE (cap analysis of gene expression): a protocol for the detection of promoter and transcriptional networks**. *Methods Mol Biol* 786:181-200
- [82] <https://fantom.gsc.riken.jp/protocols/basic.html>
- [83] Boguski MS, Tolstoshev CM, Bassett DE Jr (1994) **Gene discovery in dbEST**. *Science* 265(5181):1993-4
- [84] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) **Basic local alignment search tool**. *J Mol Biol* 215(3):403-10
- [85] Eddy SR (1998) **Profile hidden Markov models**. *Bioinformatics* 14(9):755-63
- [86] 矢田哲士 「隠れマルコフモデルが明かす生物配列の情報」生物物理 40(1):25-30, 2000
- [87] Stanke M, Waack S (2003) **Gene prediction with a hidden Markov model and a new intron submodel**. *Bioinformatics* 19 Suppl 2:ii215-25
- [88] Vogel C, Marcotte EM (2012) **Insights into the regulation of protein abundance from proteomic and transcriptomic analyses**. *Nat Rev Genet* 13(4):227-32



- [89] Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Goodlett DR, Aebersold R, Hood L (2001) **Integrated genomic and proteomic analyses of a systematically perturbed metabolic network.** *Science* 292(5518):929-34
- [90] <https://ruo.mbl.co.jp/bio/support/method/sds-page.html> 株式会社医学生物学研究所「ポリアクリルアミド電気泳動 (SDS-PAGE) の原理と方法」
- [91] <https://ruo.mbl.co.jp/bio/support/method/westernblotting.html> 株式会社医学生物学研究所「ウエスタン・ブロッティング (WB) の原理と方法」
- [92] 鈴木祥夫「総タンパク質の定量法」ぶんせき 2018(1), 2-9, 2018-01 (日本分析化学会)
- [93] [http://www.pssj.jp/archives/protocol/purification/ion\\_01/ion\\_01.html](http://www.pssj.jp/archives/protocol/purification/ion_01/ion_01.html) イオン交換クロマトグラフィー～MonoQによる RecA の精製～ (日本蛋白質科学会)
- [94] [http://www.pssj.jp/archives/protocol/purification/GPC\\_01/GPC\\_01.html](http://www.pssj.jp/archives/protocol/purification/GPC_01/GPC_01.html) ゲル濾過クロマトグラフィーによるタンパク質の精製及び分子利用決定 (日本蛋白質科学会)
- [95] <https://www.hitachi-hightech.com/hhs/products/tech/ana/lc/basic/index.html> 液体クロマトグラフ (HPLC) 基礎講座 (株式会社日立ハイテクサイエンス)
- [96] <https://www.jaima.or.jp/jp/analytical/basic/chromatograph/lc/> 高速液体クロマトグラフの原理と応用 (JAMA, 日本分析機器工業会)
- [97] 平野久「プロテオーム解析概論」ぶんせき 367:348-353, 2005 (日本分析化学会)
- [98] [https://www.cosmobio.co.jp/product/detail/gnn-20130514.asp?entry\\_id=11030](https://www.cosmobio.co.jp/product/detail/gnn-20130514.asp?entry_id=11030) プロテオーム解析 質量分析受託サービス
- [99] <https://azscience.jp/column/category/top01-sub03/> 「MALDI-TOF/TOF-MS とは? 基礎や原理・用途などを解説」 アズサイエンス株式会社
- [100] <https://www.an.shimadzu.co.jp/service-support/technical-support/analysis-basics/maldi> 「MALDI-TOFMS の原理」 島津製作所
- [101] Clark AE, Kaleta EJ, Arora A, Wolk DM (2013) **Matrix-assisted laser desorption ionization-time of flight mass spectrometry: a fundamental shift in the routine practice of clinical microbiology.** *Clin Microbiol Rev* 26(3):547-603
- [102] Calderaro A, Chezzi C (2024) **MALDI-TOF MS: A Reliable Tool in the Real Life of the Clinical Microbiology Laboratory.** *Microorganisms* 12(2)
- [103] <https://www.an.shimadzu.co.jp/hplc/support/lib/lctalk/61/61intro.htm> 「四重極型, イオントラップ型, 飛行時間型 各質量分析計の分離機構と特徴」 島津製作所
- [104] <https://www.jaima.or.jp/jp/analytical/basic/mass/qms/> 「四重極形質量分析計の原理と応用」 日本分析機器工業会 (JAIMA)

- [105] Dunn WB (2011) **Mass spectrometry in systems biology an introduction.** *Methods Enzymol* 500:15-35
- [106] [https://www.jstage.jst.go.jp/article/massspec/54/5/54\\_5\\_217/\\_pdf](https://www.jstage.jst.go.jp/article/massspec/54/5/54_5_217/_pdf) 内藤康秀  
「m/z の定義とその使用方法について」
- [107] <https://www.thermofisher.com/jp/ja/.../co-immunoprecipitation-co-ip.html> Thermo Fisher Scientific 「共免疫沈降法 (Co-IP)」
- [108] <http://ruo.mbl.co.jp/bio/support/method/co-immunoprecipitation.html> MBL ライフサイエンス 「共免疫沈降 (Co-IP) の原理と方法」
- [109] <https://www.thermofisher.com/jp/ja/.../pull-down-assays.html> Thermo Fisher Scientific 「プルダウンアッセイ」
- [110] Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO (1999) **Assigning protein functions by comparative genome analysis: protein phylogenetic profiles.** *Proc Natl Acad Sci U S A* 96(8):4285-8
- [111] Bergman NH, Passalacqua KD, Hanna PC, Qin ZS (2007) **Operon prediction for sequenced bacterial genomes without experimental information.** *Appl Environ Microbiol* 73(3):846-54
- [112] Fields S, Song O (1989) A novel genetic system to detect protein-protein interactions. *Nature* 340(6230):245-6
- [113] Maple J, Møller SG (2007) **Yeast Two-Hybrid Screening.** In Rosato E. (eds) *Circadian Rhythms. Methods in Molecular Biology*, 362: 207-223 Humana Press
- [114] <http://ruo.mbl.co.jp/bio/support/method/chromatography.html> MBL ライフサイエンス 「クロマトグラフィーの原理と応用」
- [115] <https://www.gelifesciences.co.jp/technologies/affinity/index.html> GE ヘルスケア・ジャパン 株式会社 ライフサイエンス統括本部 アフィニティークロマトグラフィー
- [116] <https://www.thermofisher.com/jp/ja/.../overview-affinity-purification.html> ThermoFisher SCIENTIFIC 「アフィニティー精製に関する概要」
- [117] Barderas R, Benito-Pena E (2019) **The 2018 Nobel Prize in Chemistry: phage display of peptides and antibodies.** *Anal Bioanal Chem* 411(12):2475-2479
- [118] <https://www.biologica.co.jp/products-service/custom-antibody/phage-display-library/> 株式会社バイオロジカ ファージディスプレイ&抗体ライブラリーサービス
- [119] Ge H, Liu Z, Church GM, Vidal M (2001) **Correlation between transcriptome and interactome mapping data from *Saccharomyces cerevisiae*.** *Nat Genet* 29(4):482-6

- [120] <http://www.bionano-molec.org/backgrounds/phage-display/> Bionanomolecule Design Web  
ページ提示法
- [121] Murai J et al (2018) **SLFN11 Blocks Stressed Replication Forks Independently of ATR.** *Mol Cell* 69(3):371-384
- [122] Koppikar et al (2012) **Heterodimeric JAK-STAT activation as a mechanism of persistence to JAK2 inhibitor therapy.** *Nature* 489(7414):155-9
- [123] Kiu H, Nicholson SE (2012) **Biology and significance of the JAK/STAT signalling pathways.** *Growth Factors* 30(2):88-106
- [124] <https://jp.sinobiological.com/pathways/jak-stat-signaling-pathway> Sino biological

## 索引

ddNTP, 16

dNTP, 16

ナノポア, 26

## Contributors

---

履修者 三浦昌浩、武山尚生、栗原悠、鈴木亮佑、岡本理沙、鈴木昌樹、aa、チョコビ、中井瑞、瀧本ゆりあ、はる、望月智美、伊藤颯汰、Akitaro、Genesis、ボブ、asuuu、鶴岡亜里紗、岡部晴子、呉安真耶子、永田真結子、Amberjack、菅原颯人、岡崎貴裕、五十嵐巳玲一、YS、渡部素世香、櫻井恒暢、鈴木結香子、儀満光紀、苑原雄也、椎名万里子、星野直樹、成木美羽、山本フィリップ、犬、Shin JungMin、dorayaki、こめこ、Cagalingan Lianne、Clara、相場はるな、Jiwoo、西田有希、齋藤穰、smith、佐藤もも、岡田裕貴、鴛渕孝太、笹尾優羽、Sean、佐藤慶典、Masa、池田龍生、菅原佳澄、永樂友駿、山根沙和、岡本祥、ruka、相川茉奈、るか

(敬称略)

執筆協力 森大博士

本冊子を作成するにあたって、西原達哉博士、荒川和晴博士、河野暢明博士、佐藤清敏博士、齊藤康弘博士、伊藤道俊博士には情報提供・資料提供・ご意見ご批判等、多大なご支援を頂いた。厚く御礼申し上げたい。



慶應義塾大学先端生命科学研究所  
ゲノム分子生物学 (2018-)

---