

バイオインフォマティクスシリーズ 7

エピゲノム情報解析

浜田 道昭 監修

中戸 隆一郎 著

コロナ社

シリーズ刊行のことば

現在の生命科学においては、シークエンサーや質量分析器に代表される計測機器の急速な進歩により、ゲノム、トランスクリプトーム、エピゲノム、プロテオーム、インタラクトーム、メタボロームなどの多種多様・大規模な分子レベルの「情報」が蓄積しています。これらの情報は生物ビッグデータ（あるいはオミクスデータ）と呼ばれ、このようなデータからいかにして新しい生命科学の発見をしていくかが非常に重要となっています。

このような状況の中でその重要性を増しているのが、生命科学と情報科学を融合した学際分野である「バイオインフォマティクス」（生命情報科学，生物情報科学）です。バイオインフォマティクスは、DNA やタンパク質の配列などの、生物の配列情報をデジタル情報として捉え、コンピュータにより解析を行うことを目的として誕生しました。このような、生物の配列情報を解析するバイオインフォマティクスの一分野は「配列解析」と呼ばれます（これは本シリーズでも主要なテーマとなっています）。上述の計測機器の進歩とともに、バイオインフォマティクスはここ数十年で飛躍的に発展し、いまや配列解析にとどまらずに、トランスクリプトーム解析、メタボローム解析、プロテオーム解析、生物ネットワーク解析など多岐にわたってきています。また、必要な知識も、統計学、機械学習、物理学、化学、数学などの多くの分野にまたがっています。しかしながら、これらのバイオインフォマティクスの多岐にわたる分野を、教科書的・体系的に学ぶことができる成書シリーズは、国内外を見てもほとんどありません。

そこで、大学生、大学院生、技術者、研究者などに、バイオインフォマティクスの各分野を体系的に学習することを可能とするための教科書を提供することを目的として本シリーズを企画しました。これを実現するために、バイオイン

ii シリーズ刊行のことば

フォマティクス分野の最前線で活躍をしている，若手・中堅の研究者に執筆を依頼しております。執筆者の方々には，バイオインフォマティクス研究の基盤となる理論やアルゴリズムを中心に，可能な限り厳密かつ自己完結的に解説を行うようお願いしています。そのため，本シリーズは，大学などにおけるバイオインフォマティクスの講義の教科書として活用可能であるのみならず，読者が独学する場合にも最適な書籍になっていると確信しています。

最後になりますが，本シリーズの企画の段階から辛抱強くサポートしてくださったコロナ社の皆様に御礼を申し上げます。本シリーズが，今後のバイオインフォマティクス研究さらには生命科学研究の一助となることを切に願います。

2021年9月

「バイオインフォマティクスシリーズ」監修者 浜田道昭

まえがき

生物のゲノム配列には複雑な生命情報がどのように書き込まれているのか？この魅力的な問いは我々を惹きつけて離さない。ゲノムはその配列情報のみならず、配列の修飾状態（エピゲノム）や核内における立体的構造などを利用してその機能を適切に発現させている。シークエンサーを用いてそのようなエピゲノム・立体構造情報を全ゲノム的に調査するエピゲノム情報解析は、ゲノムに潜むそのような謎を解明するための手段となりうる。エピゲノム情報解析分野は近年急速に発展しており、統計学、配列解析、機械学習、分子シミュレーション、数理モデルなどさまざまな手法を取り入れながら、分野横断的で多彩な研究が展開されている。この多彩さがエピゲノム情報解析の魅力であるが、エピゲノム自体の種類の高さも相まって、その全体像を把握することがやや難しい側面がある。

本書はそのようなエピゲノム情報解析の全体像を把握するための入門書となることを目指して執筆した。本書は初学者でも読むことができ、また中級者にも学びがあるものとなっている。生命系、理論系いずれのバックグラウンドを持つ方にとっても有用なものとなるように努力したつもりである。エピゲノム情報解析はきわめて進歩が速い分野であり、トレンドの変化も激しい。本書を長く有用なものとするため、個々のサンプル調製プロトコルやツールの紹介などは最小限とし、その背後にあるより普遍的な考え方や歴史的な潮流について説明することに注力した。文献については歴史的に重要なものを押さえながら、深層学習手法など最新のアプローチも適宜取り入れるようにしている。教科書的な位置づけも意識しており、講義の教材として使っていただければ大変嬉しく思う。

本書は全6章から構成される。1章は導入であり、ゲノム・エピゲノム情報

とは何か、どのような種類があるかについて、またエピゲノム情報解析に必須の技術であるシーケンサーの基本的なデータ解析方法について概説している。続く2章では転写因子結合、ヒストン修飾解析、3章ではDNAメチル化解析、4章ではオープンクロマチン・ヌクレオソーム解析を解説している。これらの解析では共通した考え方・解析法も多く、たがいに内容が関連しているため、組み合わせて読むことで理解が深まるだろう。一方、5章のゲノム三次元構造は内容がやや独立しており、解析法も他とは大きく異なる。立体構造解析についての詳細を記した和書はまだ数少ないと思われるので、ぜひ活用していただきたい。6章では国際的なエピゲノムデータベースについて解説し、それらを用いた大規模解析の例について紹介している。

なお、特に断りがない場合、本書は筆者が普段親しんでいるヒトやマウスなどの哺乳類を対象にしている。初学者にとってのわかりやすさを重視したため、個々の用語の厳密な説明については不正確な表現になっている箇所があるかもしれない。多くの文献を引用してあるので、個々のトピックに興味を持たれた読者においてはぜひそれらの文献を参照していただきたい。そのような理解の入口として使っていただき、本書をきっかけにエピゲノム情報解析研究を志す学生や研究者が増えれば筆者にとって望外の喜びである。

本書の執筆の機会を与えていただいた早稲田大学理工学術院の浜田道昭先生にこの場を借りて感謝申し上げる。また本書の執筆にあたり、数多くの先生方の助言をいただいた。東京大学定量生命科学研究所の岸雄介先生、札幌厚生病院の牧野吉倫博士には全章通じてレビューいただき、たくさんの有益なコメントをいただいた。また、千葉大学大学院医学研究院の岡部篤史先生、東京大学先端科学技術研究センターの永江玄太先生、東京大学大学院新領域創成科学研究科の三浦史仁先生、京都大学化学研究所バイオインフォマティクスセンターの阿久津達也先生、理化学研究所生命機能科学研究センターの新海創也博士には、筆者の知識が十分でない箇所について多くのご指南をいただいた。筆者の研究室である東京大学定量生命科学研究所大規模生命情報解析研究分野のメンバー、特に西條栄子博士と長岡勇也博士には用語の表現方法や説明を加えたほ

うがよいトピックに関してさまざまなコメントをいただいた。この場を借りて各氏にお礼を申し上げたい。本書の完成を辛抱強く待ち、サポートいただいたコロナ社の方々にも深く感謝する。最後に、本書の執筆をずっと応援してくれていた筆者の妻と子どもたち、彼らの応援がなければ本書の完成はなかった。心から感謝したい。

2025年6月

中戸隆一郎

【本書ご利用にあたって】

本文中に記載している会社名、製品名は、それぞれ各社の商標または登録商標です。本書では®やTMは省略しています。

目 次

1. エピゲノム情報とは

1.1	ゲノム	1
1.1.1	コーディング遺伝子・ノンコーディング遺伝子	2
1.1.2	ゲノミクス	4
1.2	エピゲノム	4
1.2.1	細胞種の個性を形成するエピゲノム	5
1.2.2	細胞分化におけるエピゲノムの遷移	5
1.2.3	ゲノムリプログラミング	6
1.2.4	環境で変化するエピゲノム	7
1.3	エピゲノムの種類	7
1.3.1	DNAメチル化	8
1.3.2	ヒストン修飾	8
1.3.3	転写因子結合	10
1.3.4	オープンクロマチン	10
1.3.5	ゲノム三次元構造	10
1.4	シーケンサーを用いたエピゲノム解析	11
1.4.1	シーケンサーの概要	11
1.4.2	リードデータの形式 (FASTQ形式)	12
1.4.3	シングルエンドリードとペアエンドリード	12
1.4.4	シーケンサーデータの公開・共有	13
1.5	リードマッピング	14
1.5.1	シーケンサーデータのマッピング	14
1.5.2	ユニークリードとマルチリード	15
1.5.3	マップデータの形式 (SAM・BAM形式)	16
1.5.4	BED形式	17

1.6	リード分布	19
1.6.1	リード分布ファイルの形式	20
1.6.2	マップバビリティ	22

2. ヒストン修飾, 転写因子結合

2.1	ChIP-seq 法とは	23
2.1.1	ChIP-seq 法の流れ	24
2.1.2	その他のアッセイ	26
2.2	ピーク検出	26
2.2.1	ピークの種類	27
2.2.2	リード分布モデル	27
2.2.3	インプットサンプル・ネガティブコントロールとの比較	30
2.2.4	p 値の閾値と多重検定の補正	33
2.2.5	Irreproducible discovery rate (IDR)	34
2.2.6	その他のピーク検出法	34
2.2.7	広域ピークの検出	35
2.2.8	混合ピークの解析	37
2.3	品質評価	38
2.3.1	ChIP-seq データの品質	38
2.3.2	リード深度	39
2.3.3	ライブラリ複雑度	40
2.3.4	S/N 比	42
2.3.5	ストランドシフトプロファイルを用いた S/N 比と DNA 断片長の評価	44
2.4	サンプル間ピーク比較	47
2.4.1	二値比較	47
2.4.2	定量的比較	49
2.4.3	定量的比較のためのリード正規化手法	50
2.4.4	スパイクイン正規化	51
2.5	モチーフ抽出	56

- 2.5.1 モチーフ抽出とは 56
- 2.5.2 モチーフ抽出法の種類 57

3. DNA メチル化

- 3.1 DNA メチル化とは 59
 - 3.1.1 DNA メチル化・ヒドロキシメチル化 59
 - 3.1.2 DNA 脱メチル化機構 60
 - 3.1.3 DNA メチル化の生物学的意義 61
 - 3.1.4 ゲノムインプリンティング 62
- 3.2 DNA メチル化解析アッセイの種類 63
 - 3.2.1 MeDIP-seq 63
 - 3.2.2 バイサルファイトシークエンシング 64
 - 3.2.3 PBAT 法 65
 - 3.2.4 Reduced representation bisulfite sequencing (RRBS) 66
 - 3.2.5 マイクロアレイ (Infinium) 66
 - 3.2.6 1分子シークエンサーを用いたメチル化解析 67
- 3.3 DNA メチル化解析のワークフロー 68
 - 3.3.1 DNA メチル化解析の流れ 68
 - 3.3.2 サンプル間メチル化比較：フィッシャーの正確確率検定 70
 - 3.3.3 サンプル間メチル化比較：ロジスティック回帰 71

4. オープンクロマチン, ヌクレオソーム

- 4.1 オープンクロマチン・ヌクレオソーム解析とは 73
 - 4.1.1 オープンクロマチン解析 73
 - 4.1.2 オープンクロマチン解析とヌクレオソーム解析の種類 74
 - 4.1.3 ChIP-seq との用途の違い 78
- 4.2 オープンクロマチン・ヌクレオソーム解析のワークフロー 79
 - 4.2.1 オープンクロマチン解析の流れ 79
 - 4.2.2 ヌクレオソーム解析の流れ 80

4.2.3	オープンクロマチン解析のピーク抽出	81
4.2.4	モチーフ抽出	82
4.2.5	転写因子フットプリント	82
4.3	シングルセル解析 (scATAC-seq)	84
4.3.1	scATAC-seq とは	85
4.3.2	scATAC-seq の流れ	85
4.3.3	遺伝子発現量の推定	88
4.3.4	疑似バルク解析	89
4.3.5	エンハンサー・プロモーター相互作用推定	90
4.3.6	遺伝子制御ネットワークの構築	91

5. ゲノム三次元構造

5.1	階層的ゲノム構造	94
5.1.1	染色体テリトリー	94
5.1.2	コンパートメント	95
5.1.3	TAD	95
5.1.4	クロマチンループ	97
5.1.5	ストライプ	98
5.1.6	その他のゲノム構造	99
5.1.7	その他の生物種のゲノム構造	99
5.2	ゲノム立体構造解析の原理	100
5.2.1	立体構造解析の流れ	101
5.2.2	立体構造解析で得られるリード分布	103
5.3	種々の立体構造解析手法	104
5.3.1	全ゲノム相互作用の網羅的解析 (Hi-C, Micro-C)	104
5.3.2	ゲノム領域を制限した網羅的解析 (Capture-C)	105
5.3.3	特定のタンパク質を介した立体相互作用の抽出 (ChIA-PET, HiChIP)	106
5.3.4	多点間相互作用検出	107
5.4	Hi-C 解析の流れ	108
5.4.1	コンタクト行列の計算	108

5.4.2	コンタクト行列の正規化	109
5.4.3	コンタクト距離の分布	110
5.4.4	コンパートメント検出	111
5.4.5	コンパートメント強度	113
5.4.6	TAD 検 出	114
5.4.7	Arrowhead	115
5.4.8	インシュレーションスコア	117
5.4.9	Directionality index	119
5.4.10	ル ー プ 検 出	120
5.4.11	ChIA-PET でのループ検出	121
5.4.12	ピーク集計分析	121
5.4.13	その他の解析手法	122
5.4.14	アリアル特異的解析	123
5.5	確率的立体構造	123
5.5.1	ゲノム立体構造の確率的ばらつき	124
5.5.2	ループ押し出しモデル	125
5.5.3	三次元構造モデリング	127

6. エピゲノムデータベースと大規模解析

6.1	エピゲノムデータベース	130
6.1.1	論文公開のためのデータアーカイブ	130
6.1.2	種々のエピゲノムデータベース	133
6.1.3	ヒストン修飾データベース	134
6.1.4	転写因子結合領域データベース	134
6.1.5	三次元構造データベース	135
6.1.6	エピゲノムデータベースの注意点	136
6.2	多サンプルを用いた大規模解析	137
6.2.1	文脈特異的な転写因子相互作用の同定	137
6.2.2	クロマチン状態アノテーション	139
6.2.3	BED12 形 式	142
6.3	機械学習を用いた予測モデル	143

1

エピゲノム情報とは

bioinformatics

ゲノム (genome) は生命の遺伝情報の総体であり、生命の設計図としての役割を持つ。すべての生命活動に必要な情報はゲノム上に存在しており、このゲノムを次世代・子孫に受け継ぐことで、種としての保存を可能にしている。ゲノム上にはタンパク質をコードする多数の遺伝子 (ヒトの場合 2 万~3 万) が存在し、皮膚や脳、消化管、骨などの臓器を構成するさまざまな細胞種はそれぞれに必要な遺伝子を適切に発現させているが、その適切な発現制御に重要なのが DNA 修飾やヒストン修飾などの**エピゲノム** (epigenome) である。ゲノム配列は全身の細胞で基本的な同一であるのに対し、エピゲノムは細胞種やコンディションごとに異なるパターンを示す。このようなゲノム配列やエピゲノムの修飾パターンの異常はある種の疾患を引き起こすことがわかってきている。本章ではこのゲノム・エピゲノムについて概説する。

1.1 ゲノム

ゲノムは生命の遺伝情報の総体である。ゲノム (genome) という単語はすべての遺伝子 (gene) を集めたもの (-ome) を指す概念的な単語として定義されたが、現在では非遺伝子領域も含めた全 DNA 配列のことを指す。わかりやすく言えば、ゲノムとは細胞内に存在するすべての染色体 DNA[†] を集めた (つなげた、と言い換えてもよい) ものである。例として、ヒト細胞核には 1

† 染色体という用語はクロマチンタンパク質を含む定義もあるが、本書では DNA とほぼ同義で用いる。

番から 22 番までの常染色体と 2 本の性染色体 (X と Y), さらにミトコンドリアなどが格納されているが, 「ヒトゲノム配列」とはこれらの染色体配列をすべて足し合わせたものになる。

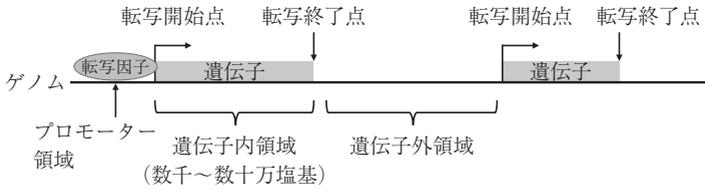
ゲノム配列は**アデニン** (adenine), **グアニン** (guanine), **シトシン** (cytosine), **チミン** (thymine) の 4 種の**塩基** (base) からなる。計算機上ではこれらの塩基の頭文字をそれぞれ取って A, G, C, T と表記する (実際のゲノム配列には非解読塩基を表す N も存在する)。DNA は二重らせん構造を取っており, アデニンはチミンと, シトシンはグアニンとペアとなり, **塩基対** (base pair) を形成する。ゲノムの長さの単位としても塩基対が使われるが (英語では bp と略される), 和文では単に塩基と呼ばれることもある (本書では塩基とする)。例として, ヒトゲノムの長さは約 30 億塩基 (three billion bp) であり, ヒトゲノム解析は A, G, C, T の 4 種の文字 (あるいは N を加えた 5 種) からなる 30 億文字の長大な文字列を解析する文字列解析と表現することができる。

1.1.1 コーディング遺伝子・ノンコーディング遺伝子

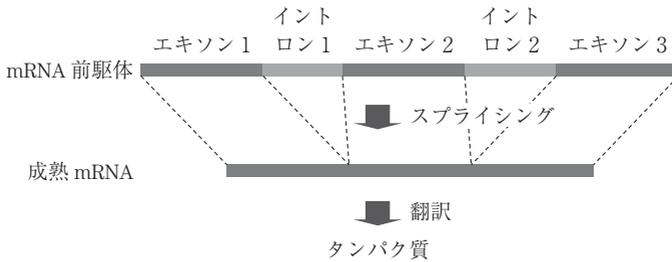
ゲノム上には多数の遺伝子が存在する (図 1.1)。その多くは数千~数十万塩基程度の長さを持つ。遺伝子の転写開始点上流のプロモーター領域に特定の**転写因子** (transcription factor) が結合することで RNA への転写が活性化する。遺伝子にはエキソン領域とイントロン領域があり, 遺伝子から転写された mRNA 前駆体には両者が含まれるが, その後スプライシングによってイントロン領域が除去され成熟 mRNA となり, その成熟 mRNA がタンパク質へと翻訳される (図 1.1 (b))。

ゲノム上には**遺伝子領域** (intragenic region または genic region) の他に, 遺伝子ではない**遺伝子外領域** (intergenic region または extragenic region) も存在する。出芽酵母はほとんどのゲノム領域が遺伝子領域で占められているが, ヒトゲノムの場合, タンパク質を合成する**コーディング遺伝子** (coding gene) のゲノムに占める割合は 5 % にも満たない^{1)†}。残りの遺伝子外領域に

† 肩付き数字は, 巻末の引用・参考文献番号を示す。



(a) ゲノム内の遺伝子領域



(b) エクソンとイントロン (エクソンとイントロンを含む mRNA 前駆体がスプライシングを受け、イントロンが除去され、成熟 mRNA となる)

図 1.1 ゲノムと遺伝子領域

はエンハンサー (enhancer) やサイレンサー (silencer) など各細胞種で転写を適切に制御するための機能領域も存在するが、大部分の領域は同じ配列パターンが繰り返し現れる**反復配列** (repetitive sequence, リピート) 領域などで占められている。このような領域は機能を持たない不必要な領域であると古くは考えられていたが、その後のゲノム・エピゲノム研究によって、これらの領域の多くに機能が存在することがわかってきている²⁾。例えば、タンパク質をコードしないノンコーディング RNA (non-coding RNA) を転写する**ノンコーディング遺伝子** (non-coding gene) がゲノム上に広く存在する。ノンコーディング RNA には siRNA や miRNA などの小分子 RNA (small RNA, 20~30 塩基程度) や、長大な長鎖ノンコーディング RNA (long non-coding RNA, lncRNA と略す。数万塩基程度) などがあり、細胞内でさまざまな機能を担っている。エピゲノムに関する国際プロジェクトである ENCODE (Encyclopedia of DNA Elements) プロジェクトが 2012 年に発表した論文では、ヒトゲノムの 80.4%

の領域は最低一つの細胞種において転写活性または一つ以上のクロマチンマークでカバーされると報告されており³⁾、これらの領域は機能を持つ領域である可能性がある。

1.1.2 ゲノミクス

ゲノムは生物種ごとに異なる。また、同じ生物種でも0.1%程度の個人差・多様性があり、身体的特徴や特定疾患のリスクなどの違いにつながっている。「ゲノムのどの領域にどのような差異が存在するとこれらの身体的特徴あるいは疾患リスクにつながるのか」という問いに答えるための研究が世界中で行われており、このような研究領域を**ゲノム学**または**ゲノミクス** (genomics) と呼ぶ。例えば**ゲノムワイド関連解析** (genome-wide association study, GWAS という略称で呼ばれる) は、数万~数十万人もの人間からゲノムを取得し、個人の持つ種々の特徴とゲノムの個人差の相関を取ることでその特徴に重要なゲノム部位の同定を試みる手法であり、世界中のさまざまな地域で研究が進められている⁴⁾。このようなゲノム個人差の情報を用いて、個々人に最適化された治療方法を提案する**個別化医療** (personalized medicine) の試みが進んでいる。

1.2 エピゲノム

細胞核内においてDNAは**ヒストン** (histone) と呼ばれるタンパク質に巻き付いたかたちで存在しており、このDNAとヒストンの複合体を**ヌクレオソーム** (nucleosome) と呼ぶ。さらに、多数の連続したヌクレオソームのまとまりを**クロマチン** (chromatin) と呼ぶ。**エピゲノム** (epigenome) とはこれらDNAやヒストンの化学修飾や、オープンクロマチンなどのゲノム構造の全体を指す言葉である (図1.2)。「epi」は「後に」という意味を表すギリシャ語であり、ゲノムに後から追加された生命情報という概念を表している。エピゲノムは各細胞種で必要な遺伝子群を適切に発現させるための「スイッチ」の役割を果たす。したがってエピゲノムが正しく制御されることは生命活動を正し

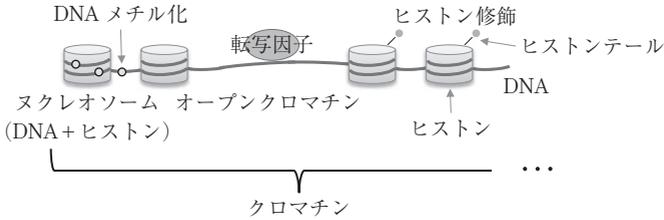


図 1.2 エピゲノムの模式図。DNA はヒストン八量体に巻き付き、ヌクレオソームを形成する。DNA のメチル化、ヒストンのテールと呼ばれる領域の化学修飾（ヒストン修飾）、ヒストンがほどけてフリーになった DNA 領域（オープンクロマチン）などを総称してエピゲノムと呼ぶ。

く機能させるために重要である。

1.2.1 細胞種の個性を形成するエピゲノム

生物の体を構成するさまざまな細胞種は、それぞれに重要な遺伝子を適切に発現させるため、特異的なエピゲノムパターンを持つ。このエピゲノムパターンはその細胞種の個性（アイデンティティ）を決定づける要素とみなすことができる。さまざまな細胞種からそのエピゲノム情報を網羅的に抽出し、細胞種間のエピゲノムパターンの違いを調査することで、それぞれの細胞種の個性、すなわち重要なエピゲノム領域、その領域に結合する重要な転写因子などを探索することが可能である。このような研究領域を**エピゲノム学**または**エピゲノミクス** (epigenomics) と呼ぶ[†]。エピゲノミクスによって、例えば健常者の細胞と特定の疾患患者の細胞を比較することにより、エピゲノムの喪失・変化によって引き起こされる疾患の発症メカニズムを紐解くことが可能になる。

1.2.2 細胞分化におけるエピゲノムの遷移

受精卵から細胞分化を経て各細胞種へと分化していく流れを、谷間を転げ落

[†] エピゲノミクスと類似の分野に**エピジェネティクス** (epigenetics) がある。エピジェネティクスとは、DNA 配列の変化を伴わない「エピジェネティックな」変化による個々の遺伝子制御に焦点を当てた研究分野であるのに対し、エピゲノミクスはより全ゲノムにわたるゲノム修飾の変化やその機能に焦点を当てている。本書は全ゲノム解析をおもな対象とするため、ここではエピゲノミクスとしている。

索 引

<p style="text-align: center;">【あ】</p> <p>アデニン 2</p> <p>アリル特異的解析 123</p> <p>アリル特異的メチル化解析 63</p> <p>アンカー部位 98</p> <p style="text-align: center;">【い】</p> <p>遺伝子外領域 2</p> <p>遺伝子活性スコア 88</p> <p>遺伝子制御ネットワーク 91</p> <p>遺伝子領域 2</p> <p>インシュレーションスコア 117</p> <p style="text-align: center;">【う】</p> <p>ウィンドウ 20</p> <p>ウィンドウサイズ 20</p> <p>ウォディントンのエピゲノム地形 6</p> <p style="text-align: center;">【え】</p> <p>液-液相分離 95</p> <p>エピゲノミクス 5</p> <p>エピゲノム 1, 4</p> <p>エピゲノム学 5</p> <p>エピジェネティクス 5</p> <p>エピジェネティック・リプログラミング 6</p> <p>エピトランスクリプトーム 8</p> <p>塩基 2</p> <p>塩基対 2</p> <p>エンハンサー 3</p>	<p style="text-align: center;">【お】</p> <p>欧州バイオインフォマティクス研究所 131</p> <p>オッズ 71</p> <p>オッズ比 71</p> <p>オープンクロマチン解析 74</p> <p>オープンクロマチン領域 73</p> <p style="text-align: center;">【か】</p> <p>拡張クロマチン状態 アノテーション 142</p> <p>過分散 29</p> <p>がんゲノムアトラス 77, 134</p> <p style="text-align: center;">【き】</p> <p>疑似バルク解析 89</p> <p>偽発見率 33</p> <p>共アクセスシビリティスコア 90</p> <p>境界因子 125</p> <p>偽陽性 15</p> <p>偽陽性ピーク 31</p> <p>局所ボアソン分布 30</p> <p style="text-align: center;">【く】</p> <p>グアニン 2</p> <p>クロマチン 4</p> <p>クロマチンジェット 99</p> <p>クロマチン状態 140</p> <p>クロマチン状態 アノテーション 140</p> <p>クロマチン免疫沈降法 23</p> <p>クロマチンループ 97</p>	<p style="text-align: center;">【け】</p> <p>蛍光 in situ ハイブリダイゼーション法 94</p> <p>ゲノミクス 4</p> <p>ゲノム 1</p> <p>ゲノムインプリンティング 62</p> <p>ゲノム学 4</p> <p>ゲノムリプログラミング 6</p> <p>ゲノムワイド関連解析 4</p> <p style="text-align: center;">【こ】</p> <p>広域ピーク 27</p> <p>格子縞パターン 111</p> <p>抗体 23</p> <p>国立衛生研究所 134</p> <p>国立生物工学情報センター 131</p> <p>コーディング遺伝子 2</p> <p>コヒーシオン 96</p> <p>個別化医療 4</p> <p>混合ピーク 27</p> <p>コンタクト行列 102</p> <p>コンパートメント 95</p> <p>コンパートメント化 113</p> <p>コンパートメント強度 113</p> <p>コンパートメントスイッチング 112</p> <p style="text-align: center;">【さ】</p> <p>サイレンサー 3</p> <p>サドルプロット 113</p> <p>サブコンパートメント 99</p>
---	---	--

【し】

シークエンサー 11
 シークエンス深度 39
 シス制御モジュール 135
 シトシン 2
 受動的脱メチル化 61
 冗長なリード 41
 シングルエンドリード 12
 シングルセル解析 84

【す】

ストライプ 98
 スtrandシフト
 プロファイル 44
 スパイクイン正規化 51
 スーパーループ 123
 スムージング 36
 鋭いピーク 27

【せ】

制限酵素サイト 104
 全ゲノムバイサルファイト
 シークエンシング 64
 染色体テトリール 94
 全リード数正規化 39

【た】

多重検定の補正 33
 多重検定の問題 33

【ち】

チェッカーボードパターン 111
 チミン 2
 重複リード 41

【て】

低マップバビリティ領域 22
 定量的比較 49
 データ補完 148
 転写因子 2
 転写因子フットプリント 83

転写伸長因子 27

【に】

二項検定 32
 二値比較 47

【ぬ】

ヌクレオソーム 4
 ヌクレオソームポジショ
 ニング 76

【の】

能動的脱メチル化 60
 ノンコーディング遺伝子 3

【は】

バイサルファイト処理 64
 配列アラインメント 14
 バックグラウンド 26
 バックグラウンドリード 26
 ハブ 98
 半自動ゲノムアノテ
 ション 140
 反復配列 3

【ひ】

ピーク 23
 ピーク検出 26
 ピーク集計分析 122
 非冗長なリード 41
 ヒストン 4
 ヒストンテール 8
 非特異的結合 32
 ビン 20
 ビンサイズ 20

【ふ】

フィッシャーの正確確率
 検定 70
 負の二項分布 29
 普遍的クロマチン状態
 アノテーション 142
 フラクタル球モデル 104

文脈特異的 137

【へ】

ペアエンドリード 12
 平衡球体モデル 104
 ベータ値 66

【ほ】

ボアソン分布 28
 飽和分析 40

【ま】

マスターレギュレーター 10
 マップバビリティ 22
 マトリックス・バラン
 シング 109
 マルチインシミュレーション
 スコア 118
 マルチリード 15

【み、め】

ミニマイザースケッチ 14
 メガドメイン 123
 メチル化 DNA 免疫沈降法 63
 メチル化検出 69
 メディエーター 97

【も】

モチーフ 56
 モチーフ抽出 23

【ゆ】

有意差判定 33
 ユニークリード 15

【ら、り】

ライブラリ複雑度 41
 リード 12
 リード深度 39
 リード分布 19
 リードマッピング 14

【る】		【ろ】	
ループ押し出し因子	125	ロジスティック回帰	70
ループ押し出しモデル	125	ロジット	71
◆ ◆			
【A】		DNase I	74
Array Express	131	DNase-seq	74
ATAC-seq	76	DNA 断片	24
【B】		DNA フラグメント	24
BAM 形式	17	DNA メチルトランス フェラーゼ	59
BedGraph 形式	20	DOHaD 仮説	7
BED 形式	17	【E】	
Benjamini-Hochberg 法	33	ENCODE	134
bigWig 形式	20	European Nucleotide Archive	131
BLUEPRINT	134	【F】	
BS-seq	64	FAIRE-seq	75
Burrows-Wheeler 変換	14	FASTQ 形式	12
【C】		FDR	33
Capture-Hi-C	106	frequently interacting region	98
ChIA-PET	106	FRiP スコア	42
ChIP-Atlas	135	【G】	
ChIP-seq 法	23	Gene Expression Omnibus	13, 131
CRAM 形式	17	GEO	13
CTCF	61	【H】	
CUT&RUN	26	Hi-C Computational Unbiased Peak Search	121
CUT&Tag	26	HiChIP	106
【D】		hyper-ChIPable region	32
DDBJ Sequence Read Archive	131	【I】	
DDBJ	13	ICE 法	110
Developmental Origins of Health and Disease 仮説	7	IDR	34
differentially methylated region	70	IHEC	134
Directionality index	119	Infinium	66
DMR	70	iterative correction and eigenvector decomposition 法	110
DNA Data Bank of Japan	13, 131	【J】	
		Jensen-Shannon 距離	42
		【K】	
		Knight-Ruiz 法	110
		KR 法	110
		【L】	
		LAD	95
		Loop Catalog	136
		【M】	
		MeDIP-seq	63
		Micro-C	105
		Micro-Capture-C	106
		micrococcal nuclease	105
		MNase-seq	76
		MulTI-Tag	26
		【N】	
		NSC	46
		【P】	
		pausing index	37
		PBAT 法	65
		PCR bias	41
		PCR	25
		Phred 品質スコア	12
		【R】	
		Reduced representation bisulfite sequencing	66
		ReMap	135

RNA ポリメラーゼ II	27	SRA	13	Wig 形式	20
Roadmap	134				
RRBS	66	【T】		【数字】	
RSC	46	TAD	95	3C	100
【s】		TAD の境界	96	4D ヌクレオーム	135
S/N 比	42	tagAlign 形式	18	5-カルボキシシトシン	60
SAM 形式	16	traveling ratio	37	5-ヒドロキシメチル シトシン	59
Sequence Read Archive		【w】		5-ホルミルシトシン	60
	13, 131	WGBS	64	5-メチルシトシン	59

—監修者・著者略歴—

浜田 道昭 (はまだ みちあき)
2000年 東北大学理学部数学科卒業
2002年 東北大学大学院理学研究科修士課程修了(数学専攻)
2002年 株式会社富士総合研究所研究員
2009年 東京工業大学大学院総合理工学研究科博士後期課程(社会人博士)修了(知能システム科学専攻),
博士(理学)
2010年 東京大学特任准教授
2014年 早稲田大学准教授
2018年 早稲田大学教授
現在に至る

中戸 隆一郎 (なかと りゅういちろう)
2005年 京都大学工学部情報学科卒業
2007年 京都大学大学院情報学研究科修士課程修了(知能情報学専攻)
2010年 京都大学大学院情報学研究科博士後期課程修了(知能情報学専攻),
博士(情報学)
2010年 東京大学助教
2019年 東京大学講師
2022年 東京大学准教授
現在に至る

エピゲノム情報解析

Computational Analysis for Epigenome

© Ryuichiro Nakato 2025

2025年8月18日 初版第1刷発行

検印省略

監修者 浜田 道昭
著者 中戸 隆一郎
発行者 株式会社 コロナ社
代表者 牛来真也
印刷所 新日本印刷株式会社
製本所 株式会社 グリーン

112-0011 東京都文京区千石 4-46-10
発行所 株式会社 コロナ社
CORONA PUBLISHING CO., LTD.

Tokyo Japan

振替00140-8-14844・電話(03)3941-3131(代)

ホームページ <https://www.coronasha.co.jp>

ISBN 978-4-339-02737-2 C3355 Printed in Japan

(齋藤)



 <出版者著作権管理機構 委託出版物>

本書の無断複製は著作権法上での例外を除き禁じられています。複製される場合は、そのつど事前に、出版者著作権管理機構(電話 03-5244-5088, FAX 03-5244-5089, e-mail: info@jcopy.or.jp)の許諾を得てください。

本書のコピー、スキャン、デジタル化等の無断複製・転載は著作権法上での例外を除き禁じられています。購入者以外の第三者による本書の電子データ化及び電子書籍化は、いかなる場合も認めません。落丁・乱丁はお取替えいたします。