

東京大学理学部 生物情報科学科

Department of Bioinformatics and
Systems Biology, Faculty of Science,
The University of Tokyo

生物情報科学とは

生命科学はいま、その歴史上もっとも華々しい時代を迎えています。生物情報科学は、生命科学の最先端を切り開く新しい学問領域です。その特徴は、個別の遺伝子やタンパク質を解析するだけでなく、**生命システムを生命科学と情報科学の両面から解き明かす点**にあります。

30億文字のDNA配列からなるヒトゲノム情報（全遺伝子情報）は、いまや、わずかな時間・費用で解読できるようになりました。また、オーミクス解析と呼ばれる、生命現象に関わるデータ全体を計測する技術も、驚くべき発展を続けています。これらの技術革新により、「**生命とは何か**」という問いに対し、**生命現象の全体を俯瞰する新しい視点から迫ることが可能**になりました。そして、**病気のメカニズムを解明し、生命の進化の歴史をひもといていく上で、膨大な生命科学データを情報科学的に解析していくことが不可欠**になりました。

生物情報科学科は、このような背景のもと、既存の学問分野をまたぐ幅広い視点を持ち、生命科学と情報科学の双方の専門性を備えた人材を育成するために、**東京大学理学部におよそ30年ぶりに新設された学科**です。

「バイオインフォマティクス」と「システム生物学」

生物情報科学を代表するキーワードに、上でも触れた**ゲノム**と**オーミクス**のほか、**バイオインフォマティクス**と**システム生物学**があります。これら2つのキーワードは互いに重なり合っていますが、おおまかにいうとバイオインフォマティクスは生命を「情報」として捉え、生命科学の膨大なデータを解析するための情報科学的手法を開発し、生命現象の背後にひそむ法則性や規則性を見つけ出す研究分野です。また、システム生物学は生命を「システム」として捉え、遺伝子やタンパク質など個別の要素である“部分”と生命現象のダイナミックな振る舞いである“全体”との関係を数理モデルなどを使って明らかにする研究分野です。いずれの分野も、いわば、**生命科学を物理学や工学のような視点から理論的に捉えることが共通した特徴**です。

生物情報科学科では、**実験（ウェット）と情報（ドライ）の両方のアプローチ**を学び、生命をシステムとして理解するための研究に取り組みます。生物情報科学とは、生命科学と情報科学を単に足し合わせたものではなく、どちらの基礎にも立脚しつつ新しい視点から生命現象を理解することを目指す学問なのです。

生物情報科学科誕生のインパクト

私たちの最初の取り組みは、平成13年度から始まりました。当初は学科という形ではなく、文部科学省振興調整費・新興分野人材養成「生物情報科学学部教育特別プログラム」により、主に理学部他学科の学生が本来所属する学科の講義に加えて生物情報科学分野の講義を受講する、という形で開始しました。既存の講義のない6限目や夏休み等に開講しましたが、工学部、農学部、薬学部学生や大学院生、更には社会人までさまざまな人が受講し、いざ始まってみると平均40名近い受講者があり予想以上の大盛況となりました。平日の夜や夏休みに行う非常に厳しいカリキュラムであったにも関わらず、約70名の修了者が輩出されました。この数は通常の学科に相当するものであり、生物情報科学という学問が極めて魅力的で、しかもその需要が大きいことを明確に示していました。

上記のプログラムは平成16年度に終了しましたが、学生からの人気の高さが後押しして「生物情報科学学部教育プログラム」に引き継がれ、本学の独力で新規の学科として生物情報科学科が設立されることになりました。学科が設立されたのは平成19年、進学振り分け後に新3年生（第1期生）が初めて配属されたのは平成21年のことです。新しい大学院専攻とは異なり、学部における**新学科設立はその学問分野が新しく発展する場合のみに**限られています。本格的な生物情報科学分野の学科設立は国立大学では国内初で、欧米でさえまだほとんどありません。従来の日本の大学は欧米の動きを追従することが多かったのですが、この生物情報科学科の設置は世界に先駆けたものであり、東京大学が世界をリードしていく意気込みの現れのひとつと言えるでしょう。

生物情報科学は日進月歩の領域であることから、従来のような1専攻の教員のみでの運営では柔軟さに欠け教育のポテンシャルを最大限引き出すことができません。そこで、生物情報科学科は、理学系研究科生物科学専攻・新領域創成科学研究科メディカル情報生命専攻・情報理工学系研究科コンピュータ科学専攻の3つの専攻にまたがる教員により運営されています。また、生物情報科学科の教員のもとで大学院における研究を継続して行う場合には、個々人のニーズに合わせて、適した専攻（大学院）を受験することができます。このように**複数の専攻の教員からなる教育組織は、私たちの新しい教育体制の試みのひとつ**になっています。

メッセージ

最後に私たち教員からのメッセージを。生物情報科学は日進月歩であるだけでなく勃興してから日が浅いため人材が圧倒的に不足しています。つまり、皆さんには**この分野の中心を担う人材となれる空前のチャンスが目の前に広がっている**と言ってよいでしょう。生物情報科学科への進学を志す皆さんが勇気と情熱を持ってこの分野を開拓して、近い将来に国際的なリーダーとなっていくことを私たちは心から期待しています。



生物情報科学

バイオインフォマティクス + システム生物学

最先端の生物学実験：生命システム解析・ゲノム機能解析

遺伝子ネットワークが従うロジックとダイナミクス

疾患の病因を解明するエピゲノム・トランスクリプトーム解析

ゲノムとそこにコードされた生命システムの進化

階層縦断的に分子ネットワークを解析するトランスオミクス

生命現象をモデル化し数理的に特性を解明するシステム生物学

「生命の設計図」ゲノム情報解析

生命現象の法則性を探るオーミクスデータ解析・生物知識処理



カリキュラム

生物情報科学科では右に記載したような授業科目が用意されています。

2年生 2A セメスター専門科目（前期課程）では、理学部生物情報科学科専門科目 18 単位以上（必修 12 単位含む）を学修します。3、4 年生（後期課程）では、理学部生物情報科学科専門科目 60 単位以上（必修科目 41 単位、選択必修科目 8 単位以上、その他の専門科目 11 単位以上）を学修することになります。「生命科学基礎実験」、「生物化学実験」は生物化学科と合同で行います。「生物情報実験法」は、生物情報科学科に進学して、初めてプログラミングをする学生にも対応でき、かつ、上級者にも役立つ内容になっています。



シラバス抜粋

1・2年（Sセメスター）

生物情報科学（総合科目）

生命科学の最先端の分野である生物情報科学についての基礎をオムニバス形式で紹介する。生物情報科学研究に必要とされる生命科学と情報科学の双方の基礎的な知識と技術を習得する。

2年（Aセメスター専門科目）

生物情報学基礎論Ⅰ

最初に生物情報科学とはどんな学問か、その意義、歴史、動向について述べるとともに、それが今後どういうことに役立つのか紹介する。続いて、生物情報学の基本であるゲノム配列解析を例に、生物情報学を理解する上で欠かせない情報科学の基礎知識、具体的には、アルゴリズム、データ構造、確率的手法などについて説明する。最後に、これらの基礎知識がどのように応用されていくのか、巨大な生命情報の解析法や進化現象の解析法などについても述べる。

生物情報学基礎論Ⅱ

生物情報科学における生命科学の基礎について講義する。生化学・分子生物学・オーミクス・システム生物学、およびそれらの医科学研究との関連性についての概論。

3年（Sセメスター／S1，S2ターム）

ゲノム配列解析論Ⅰ

生物配列の情報解析法、特に配列アラインメントとその背景となる理論・アルゴリズム・応用を習得する。配列情報解析の基本を学ぶ科目である。

生物情報ソフトウェア論Ⅰ・Ⅱ

生命科学データ分析で使われるソフトウェアのなかで、中核となっている高速アルゴリズムが、どのように設計されているかを解説する。演習問題を解くことで、将来自分で設計できる力を身につけることを目標とする。

ゲノム生物学

生物の構造と機能はゲノムのDNAによって決められる。ゲノムとは世代を超えて引き継がれていく、生物の持つ遺伝情報の本体である。したがって、その基本構造を知ることがは多様な生命現象を理解するための基礎となり不可欠である。本講義ではまず遺伝情報の複製・発現などについて学び、ゲノム情報がさまざまな環境要因によって修正されることについて理解を深める。

システム生物学

細胞の増殖や分化、生体ホメオスタシス、発生、神経活動などのさまざまな生命現象の特徴を、個々の分子や遺伝子が集まって機能する分子ネットワークの「システム」の振る舞いとして理解する。簡単な分子ネットワークの組み合わせでできる微分回路や積分回路、振動子、メモリ、スイッチ応答などの振る舞いを、具体的な生命現象に照らし合わせながら講義する。さらに、最近急速に発展しつつある多階層オミクスを統合したデータドリブンのシステム生物学についても紹介する。

生物統計論

DNA配列、RNA配列、アミノ酸配列、次世代シーケンサーデータなど、生命データを解析する際に頻繁に使われる統計的手法について解説する。

生物画像情報学

生命科学の研究では、顕微鏡などを用いて取得された画像がしばしば利用される。本講義では、顕微鏡システムの動作原理・特性や、画像データから定量的な情報を引き出す解析手法などについて解説する。

細胞物理学

細胞極性や細胞骨格再編成、形態形成などの様々な生命現象を物理学の視点で理解するための基礎知識を習得する。さらに、物理法則を統計手法に取り込んで、データからその背後にある生命システムの性質を抽出する統計手法についても学ぶ。

生物情報実験法

バイオインフォマティクスを含むプログラミング一般に必要な基礎知識を演習形式で学ぶ。毎週課題をこなすことで、プログラミング能力の向上を目指す。

生物化学実験

生化学および分子生物学における基礎的な実験。タンパク質の単離、精製、大腸菌を用いた遺伝子DNA に関する実験。本実習は、生物化学科と合同で実施する。

生命科学基礎実験

シーケンシング、生化学反応シミュレーション、および遺伝子発現プロファイリング等の生物情報科学的実験を行う。本実習は、生物化学科と合同で実施する。

3年（Aセメスター／A1，A2ターム）

生物データマイニング論

大規模な生命科学データから知識を抽出するための様々な方法を講義する。

生命情報表現論

ゲノム研究の進展により生命科学は仮説駆動型からデータ駆動型の科学に変貌しつつある。このような科学を推進するには、多種多様な複雑なデータや知識を計算機でうまく扱えるようにすることが不可欠である。本講義では、その基礎となる理論や技術について解説する。

オーミクス論

オーミクスは、網羅的解析を基軸とするアプローチによって、従来にはない角度から生命を理解しようとする新しい学問分野である。本講義では、ゲノム、エピゲノム、トランスクリプトーム、プロテオームを対象に、基本的な方法論と戦略を概説するとともに、それらによって得られた代表的成果を紹介する。これらの学習を通して、データ駆動型科学としてのオーミクス、オーミクスにおけるバイオインフォマティクスの重要性、そして生命システムの解明においてオーミクスが果たす役割を理解する。



ゲノム配列解析論Ⅱ

生物配列の情報解析法、特に次世代シーケンサー配列解析とRNA配列・構造の情報解析、その背景となる理論・アルゴリズム、確率モデルの理論を習得する。

進化生態情報学

進化・生態など、マクロな観点から生命現象を解析していく上で必要な生物情報科学的知識について講義する。

情報基礎実験

生物情報科学に関する基礎的な情報処理の実験（プログラミング演習）を行う。

4年（Sセメスター／S1，S2ターム）

生物情報科学特別演習・実験Ⅰ

前半には選択した3つの研究室を2週間ずつローテーションでまわり、後半は研究室に配属され、研究を行うための入門として、各研究室の課題に取り組む。生物情報科学研究の実践を体験することによって、従来の知識を、実際の研究と結びつけて活用できる力を養うことを目標としている。

生物情報科学演習Ⅰ

生物情報科学における重要な原著論文を読むことによって、それぞれの研究の背景や研究法の基礎を学ぶ。

4年（Aセメスター／A1，A2ターム）

生物情報科学特別演習・実験Ⅱ

各研究室に配属され、研究を行うための入門として、各研究室の課題に取り組む。生物情報科学研究の実践を体験することによって、従来の知識を、実際の研究と結びつけて活用できる力を養うことを目標としている。

生物情報科学演習Ⅱ

生物情報科学における研究法の基礎および応用について、様々な演習を行う。

生命現象のモデル化

「生命とは何か」という問いに対し、生命現象の全体を俯瞰する新しい視点から迫る

ランダムな文字列数値列にも見える大規模生命データを統計解析し、生命の本質を探る研究をしています。

数理的生命モデリング

木立 尚孝 — p.19

ゲノム情報の理論的な解析手法とソフトウェアを構築するとともに、様々な応用について共同研究しています。

バイオインフォマティクス

浅井 潔 — p.13

バイオインフォマティクス

生命を「情報」として捉え、生命科学の膨大なデータを解析するための情報科学的手法を開発し、生命現象の背後にひそむ法則性や規則性を見つけ出す

生命進化の背後にあるルールや、多様な非モデル生物が持つ未知のポテンシャルを研究しています。

進化生態情報学

岩崎 渉 — p.18

数学とアルゴリズムを活用して、遺伝子配列の機能や歴史を読み解きます。

ゲノム情報生物学

フリス・マーティン — p.17

ゲノム配列や遺伝子を中心とした大規模なオミクス情報を高速に観測・解析する手法を研究しています。

ゲノム情報科学

笠原 雅弘 — p.20

大量かつ多様なデータから、信頼できる知識を高速に発見するための数理手法に関する研究を行います。

機械学習

津田 宏治 — p.16

行動中の神経活動の計測や数理モデル化を通じて神経回路の情報処理のしくみを解明しています。

システム神経科学

豊島 有 — p.12

物理と統計の眼で、多細胞集団のかたちやパターンを理解することを目指しています。

多細胞秩序形成

杉村 薫 — p.11

ヒトゲノムの変異がどのようにがん、免疫、神経疾患等の表現型を示すのか、システムレベルでの解明を目指しています。

オミクス解析技術の開発と応用

鈴木 穰 — p.15

細胞内シグナル伝達機構の動的な仕組みを実験と数理モデリングを用いたシステム生物学により解析しています。

システム生物学

黒田 真也 — p.8

ゲノムワイドな遺伝子発現制御マシナリーを解析しています。

ゲノム情報生物学

程 久美子 — p.10

がんなどの病気の克服を目指し、先端の数理科学と分子観測技術を用い、病気と免疫などの生体現象との関係を解明しています。

医科学数理

角田 達彦 — p.9

DNA3次元構造の謎の解明にとりくんでいます。

DNA・オミクス・アルゴリズム

森下 真一 — p.14

病気のメカニズムを解明し、生命の進化の歴史をひもといていく上で、膨大な生命科学データを情報科学的に解析していく

大規模生命データの分析

システム生物学・オミクス

生命を「システム」として捉え、遺伝子やタンパク質など個別の要素である「部分」と生命現象のダイナミックな振る舞いである「全体」との関係を数理モデルなどを使って明らかにする

生命を多階層システムとしてとらえる

理学系研究科 生物科学専攻 教授

黒田 真也

Shinya KURODA

私たちの研究室では生命現象を制御するシグナル伝達経路を、細胞外環境の情報を伝達するための「通信路」とみなしその特性を理解することを目的としています。特に、分子の活性化などの時間パターンに情報を埋め込む「時間情報コーディング」の概念を世界に先駆けて提唱しています。シグナル伝達経路は複雑な相互作用により成り立っているため、従来のように個々の分子を生命現象に関連づけるだけでは十分に理解できません。そこで、私たちはシミュレーションモデル構築と細胞内の分子ダイナミクス計測を密接にフィードバックさせることにより、シグナル伝達経路の振る舞いを予測できるモデルの構築を行っています。さらに、複雑なモデルをシンプルにしてシグナル伝達の振る舞いの本質を抽出します。このようにヒトを含む哺乳類を中心としたシグナル伝達機構のシステム生物学が私たちの研究テーマです。

1. シグナル伝達機構の情報コーディング

インスリン作用：インスリンは血糖を下げる唯一のホルモンで、生体のホメオスタシスを制御しています。血中のインスリンは複数の時間パターンからなり、その生理学意義も報告されていますが、その作用の分子メカニズムは不明です。「時間情報コーディング」の概念から、これらの時間パターンに複数の情報がコードされ、情報依存的に標的臓器の応答を個別に制御している可能性が考えられます。我々は、インスリンの時間パターンに埋め込まれた複数の情報が、一旦、AKT の時間パターンに多重にコードされ、下流の分子を個別に制御できることを明らかにしました（図1）。将来的には動物を用いた実験を行い、生体内における「時間情報コーディング」の実証とメカニズムの解明を目指しています。また、時間情報コーディングはインスリンに限らず広くシグナル伝達経路一般に認められる特性であることを見出しています。

細胞運命決定機構：PC12細胞では成長因子の刺激に応じてERKが活性化し、増殖または神経分化が誘導されますが、ERK活性化が一過的な場合には増殖が、持続的な場合には分化が誘導されます。つまりERK活性化の時間パターンにより細胞の運命が決定されます。私たちは刺激の増加速度と最終濃度がそれぞれ一過性あるいは持続性ERKの活性化を制御していることをシミュレーションから予測して実験により実証しました。現在、ERKやc-FOSなどの分子の活性化などを一細胞レベルでの分布データをもとに、シグナル伝達経路がどの程度正確に情報が伝達できるかを情報理論の観点から解析しています。

2. トランスオミクス

私たちは個別の研究によるボトムアップアプローチの限界を克服するため、複数階層を網羅的に測定し、多階層にまたがる大規模ネットワークを同定するトップダウンアプローチである「トランスオミクス解析」の手法を確立しました。現在までに、インスリン作用を題材に、メタボロームやリン酸化プロテオームの共同実験を行い、多階層にまたがる代謝調節経路をデータドリブンに同定しました（図2）。これによりインスリン作用の経路の全体像が初めて明らかになりました。この手法により、さまざまな生命現象のネットワークの全貌を明らかにすることができ

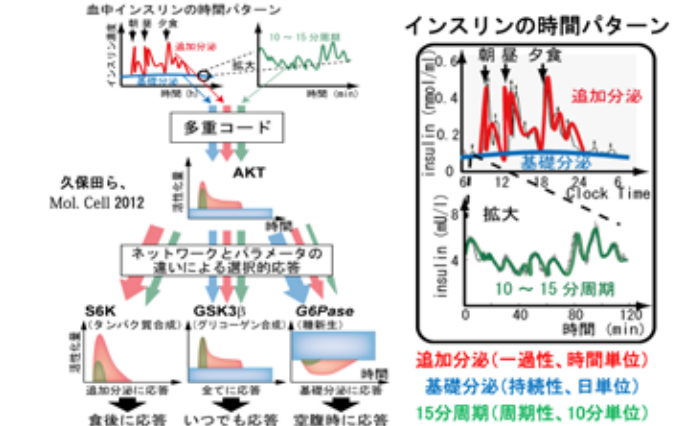


図1 インスリンの波形の違いにより異なる分子を制御

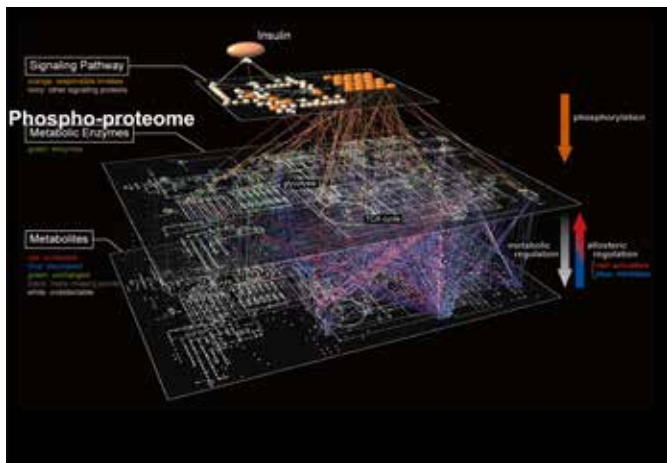


図2 インスリン作用のトランスオミクス

人材の多様性が創造性のカギ

システム生物学には生命科学や物理、工学、情報、数学などの基本的な知識も必要です。私たちの研究室ではさまざまな background の人が参加しており一つのラボで異分野融合を目指しています。

Tatsuhiko TSUNODA

先端の数理科学と分子観測技術で生命現象と病気を理解する

理学系研究科 生物科学専攻 教授

角田 達彦

先端の数理科学と分子観測技術で生命現象と病気を理解する

私たちの研究室では、がんなどの疾患を克服するために、疾患やそれに対する免疫などの動態を分子レベルで解き明かすことを目指しています。患者一人ひとりに合わせた適切な種類と量の治療を施す医療が期待されつつあります。そのため、患者のゲノム・オミックス、画像、臨床情報などのビッグデータを解析し、疾患の原因を発見します。そして統合オミックス解析を行うことで、疾患メカニズムをシステムとして理解します。このように、最先端の観測技術と数理科学と計算科学を駆使した、生命医科学研究を行います。

研究のアプローチと主な研究テーマ

医療機関との連携により患者のオミックスデータや臨床情報・病理画像情報を取得し、それらを深層学習などの手法により解析し、また新たな方法論を提案します（図1）。以下のテーマの他、個々の自由な発想も取り入れた研究を推進します。

1. 統合オミックス解析による、新たな生命現象の発見
2. 統合オミックス解析による、病気の新たな原因の発見
3. 生命医科学の現象のシステム・ネットワーク的理解
4. がん、免疫などの微小環境との相互作用の解明
5. 深層学習による、生体画像とオミックスの構造の理解



図1 患者のオミックス・臨床・病理画像データで病気を解明し予測する

がんとその微小環境との関係の解明

がん細胞集団と、免疫などの微小環境との関係をひもとくことで、個人ごとに治療の奏効や副作用、耐性獲得を予測します（図1）。例えば、私たちは肝がんの患者300人からのがん細胞の全ゲノム配列を解析し、新たな変異をもつクラスタを発見しました。このクラスタの患者は、手術後のがんの再発が起こりにくく良好な経過をたどります。さらに詳細に解析すると、がん細胞の集団的特性や免疫などの微小環境との関係が、クラスタ間で大きく異なることがわかってきました。がん細胞は実は私たち自身の細胞が変化したもので、非自己ですので免疫が排除するわけですが、がん細胞もまた巧みに変化して逃けます。つまり、がんの個性と、微小環境の相性や治療の内容で、その後の挙動に違いが生まれます。このような違いをうむメカニズムをオミックスや病理画像の解析によって解明するとともに、数理シミュレーションによる治療効果予測モデルを作り、患者に合わせた治療を行うプレジジョン医療を目指します。

深層学習による独創的な解析の方法論

深層学習が、画像のみならず、ゲノムやオミックスのデータを解析する能力を極める研究をします。応用例は、病理画像や生体イメージングなどの解析、オミックスデータの解析、それらの統合の解析です。一例として、オミックスデータをうまく変換して画像のように見せ深層学習にかけると独自の離れ技を練り出し、がんのオミックスデータでその種類を見分けられるようになりました（図2）。さらに深層学習の内部を解析して深層学習が何を見て判別しているのかを見出す技も編み出しました（図3）。それにより、がんの個性を示す新たなシグナル伝達系を発見しました。つまり、深層学習で新たな科学的発見ができることを示せました。

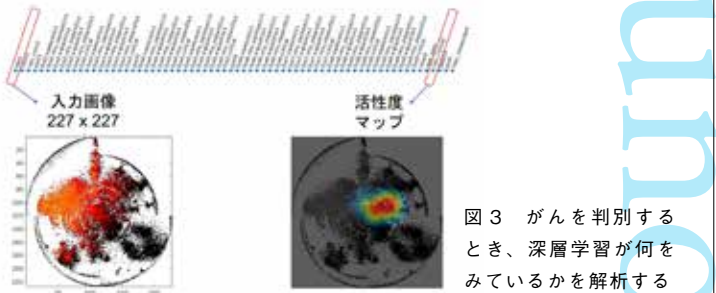
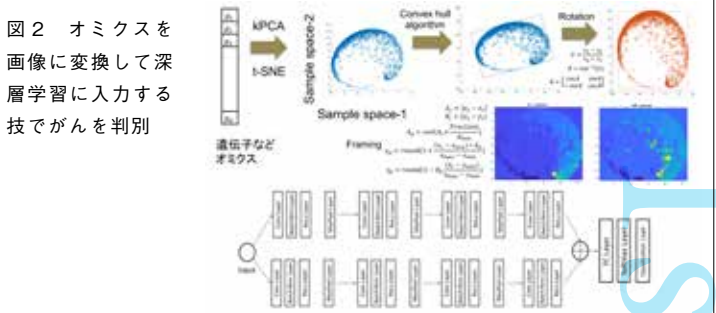


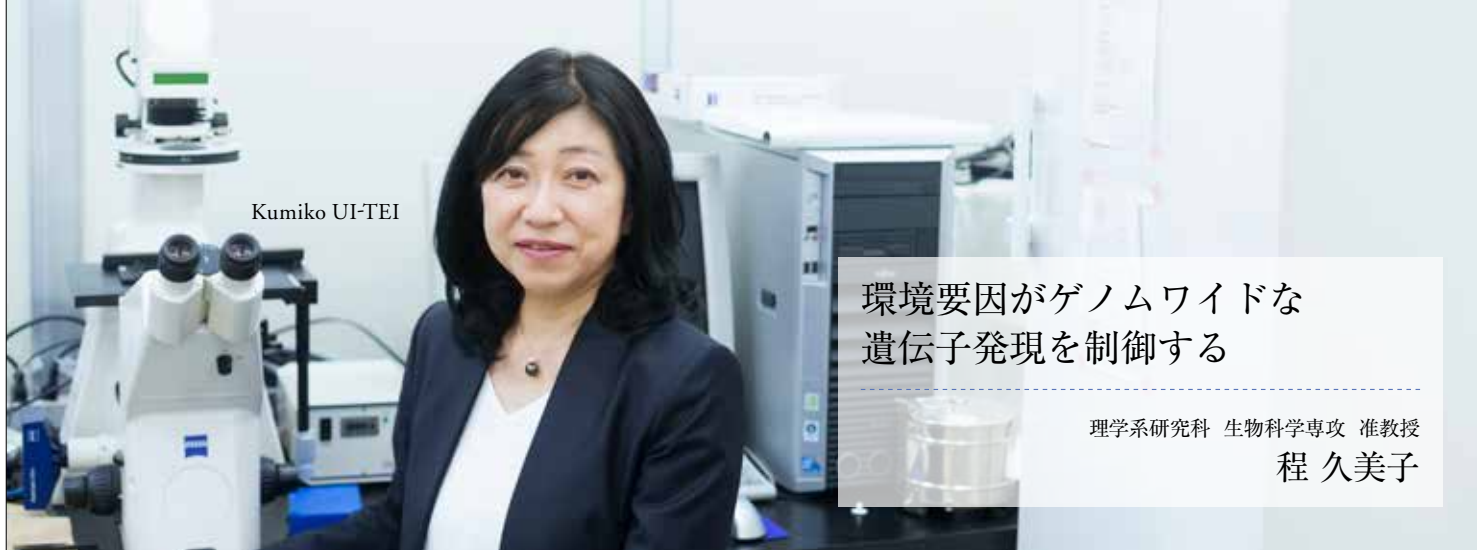
図3 がんを判別するとき、深層学習が何をみているかを解析する

トランスオミクス研究

上記のがん細胞と免疫などの微小環境の相互作用を理解するため、がんのゲノムのみならず、エピゲノムや環境、遺伝子発現などを合わせてシステムの解析する、トランスオミックス解析の方法論を研究しています。この方法は、がんだけでなく糖尿病などの生活習慣病にも使え、生体や病気の描像がより理解しやすくなります。

いろいろな人が集まって研究しています

私たちの研究室は、東大以外にも理研にもあり、また多くの連携研究をしています。そのため、生命情報研究者以外にも、臨床医や、遺伝学、シークエンス解析、ネットワーク解析、数学が好きな研究者など、また、外国人研究者も多く、お互いにさまざまな考えを出しあって日々研究しています。



Kumiko UI-TEI

環境要因がゲノムワイドな遺伝子発現を制御する

理学系研究科 生物科学専攻 准教授
程 久美子

さまざまな環境要因がRNAを介して遺伝情報を制御する

私たちの体は細胞という生命の最も基本的な構成単位からなりたっています。細胞内ではゲノムを構成するDNAからRNAが転写され、RNAからタンパク質が翻訳されるといった、セントラルドグマと呼ばれるゲノムの遺伝情報を読み解く流れが存在し、それによって生命体が機能します。しかし、ゲノムプロジェクトがすすみ、ゲノム情報のもつ全体像が明らかになるにつれて、生命機能はゲノムだけで決まってくるのではなく、環境に由来する要因も重要であることが明らかになってきました。すなわち、複雑で高次な生命現象や加齢に伴う遺伝子発現の変動は、光、重力、放射線や温度、さらには微生物の感染やストレスといった環境要因が大きな影響をもたらすのです。近年、RNAに関する研究が急速に進展し、RNAは単なる遺伝情報の仲介役ではなく、その柔軟な性質を利用して配列や構造に依存した多様な機能を示すことが明らかになってきました。このようなRNAはノンコーディングRNAと呼ばれ、これまで研究対象となっていたRNA群とは異なるものと位置づけられています。また、RNAは様々な化学修飾によってエピトランスクリプトームという新しい遺伝子発現調節の過程を構築する分子としても注目されています。私たちは環境の影響をうまく伝える新しい遺伝子制御の機構についてRNAを中心に研究しています。

小さなRNAのゲノムワイドな遺伝子発現制御

ノンコーディングRNAの中には約20塩基ほどの小さなRNAから数10キログラム塩基におよぶ長いRNAまであり、それぞれ多種多様な生命現象に関わっています。microRNAやsmall interfering RNA (siRNA) という約

20塩基のRNAは、短いながらも相補的な塩基配列をもつメッセンジャーRNA群を識別し、一気に抑制するというグローバルな働きをしています。われわれは、マイクロアレイなどの細胞・分子生物学の実験と情報科学的解析を組み合わせることにより、ヒトでは2,000種にもおよぶmicroRNAの配列やそれらのもつ熱力学的性質に依存して、ゲノムワイドな遺伝子サイレンシングが起こることを明らかにし、そのような遺伝子制御は生物の体温の影響を受けることを見出しました。

自然免疫応答とRNAサイレンシング

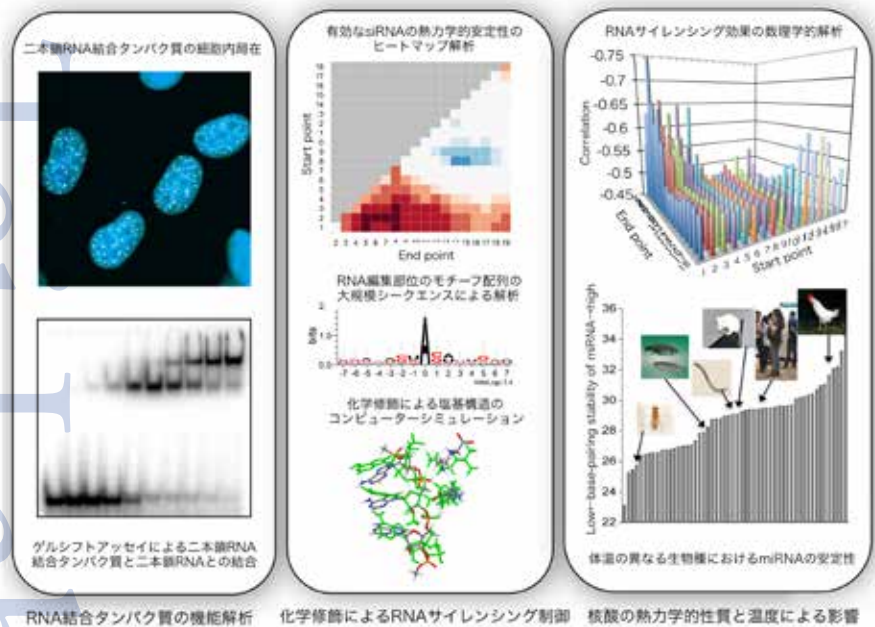
RNAウイルスが感染するとヒトを始めとする哺乳類では自然免疫応答と呼ばれる生体防御機構が駆動します。そのため、内在的なRNAサイレンシングのメカニズムは何らかの影響をうける可能性が考えられますが、その機構については不明でした。われわれは、TRBPというRNAサイレンシングに関わるRNA結合タンパク質がウイルス感染時には自然免疫応答機構にも関わり、両経路のバランスを調節する可能性を見出しました。環境要因の1つとして、特に哺乳動物特異的な免疫応答系との関連を明らかにしていきます。

エピトランスクリプトーム

古くはトランスファーRNAやリボソームRNAなどで見出された、メチル化や脱アミノ化などのRNAの化学修飾がノンコーディングRNAにも多く存在し、遺伝子発現調節機構に大きな影響を与えていることが明らかになってきました。われわれは、RNA結合タンパク質の免疫沈降実験や大規模シーケンシによる網羅的な解析により、ADARというRNA脱アミノ化酵素がmicroRNAのアデノシンを脱アミノ化することで、遺伝子発現をゲノムワイドに変換することを明らかにしています。さらに、核酸の構造学的シミュレーションをコンピュータシミュレーションで行うことで、microRNAの化学修飾がもたらすエピトランスクリプトームにおける調節機構を解明しようとしています。

RNA干渉法とCRISPR法

siRNAによるRNAサイレンシングはRNA干渉と呼ばれます。我々は、目的以外の遺伝子には影響を与えず、目的の1遺伝子のみを特異的に抑制するRNA干渉法の構築に成功しています。RNA干渉法は現代の分子生物学では欠かせないツールであると共に、その臨床応用にも大きな期待がもたれています。近年、CRISPRシステムというゲノム編集技術が開発されました。CRISPRシステムは、まだ標的特異性や効率などのさまざまな問題点が残されているため、それらの作用機構解明から技術開発を行い、両者の長所・短所を組み合わせることで、目的に合わせた効率的な遺伝子解析技術を構築します。



Kaoru SUGIMURA

物理と統計の眼で、生き物のかたちとパターンを理解する

理学系研究科 生物科学専攻 准教授
杉村 薫

学際的アプローチで多細胞秩序形成の謎に迫る

細胞が集団として動き、パターンを作ることは生き物の最も根幹的な性質の一つであり、その原理を解き明かすことで生命の本質に迫れると期待されます。その際に鍵となるのが、変形や運動を駆動する機械的な力です。

私たちの研究室では、機械的な力による多細胞秩序形成原理を解き明かすことを目指しています。この目的を達成するために、実験生物学と統計・データ解析、物理学を統合した学際的アプローチを採用しています。研究テーマは、「細胞の頂点から読み解くショウジョウバエ上皮形態形成」、「哺乳類培養細胞集団パターン形成の構成的理解」、「多細胞集団力学の統計解析手法の開発」など多岐に亘ります。

画像データから力学情報を定量する

多細胞秩序形成の力学制御を解明するためには、細胞の力や機械特性を計測することが必須です。体の表面を構成する上皮組織では、細胞接着面の張力と細胞の圧力が細胞のかたちを決定することが知られています。これらの力を非侵襲的に計測する手法は確立していませんでした。私たちは、力とかたちの逆問題をベイズ推定で定式化することで、細胞形態が可視化された画像から細胞接着面の張力と細胞の圧力を推定する手法を開発し、推定の正確性と安定性を実験的に確認しました。力のベイズ推定法は、見えない「力」を見る「かたち」から知ることが可能になります。その特長として、非侵襲で、かつ、1万細胞を超える多点同時計測が可能なが挙げられます。これらの特長を生かすことで、私たちはWhole-tissueでの力・応力の時空間動態を計測することに世界に先駆けて成功しています。

私たちは最近、画像データから得られる細胞の力・形態特徴量の相関関係にもとづいて力学モデルを構築し、モデルに含まれる力学パラメータを数値計算なしで推定する統計手法を開発しました。開発手法はその簡易さから、力学パラメータという新しい指標に基づく遺伝学スクリーニングに応用可能であり、力学情報と遺伝情報の相互作用を明らかにする上で強力なツールになると期待されます。

データから背景にある系の性質を推定する統計学手法は、蛍光イメー

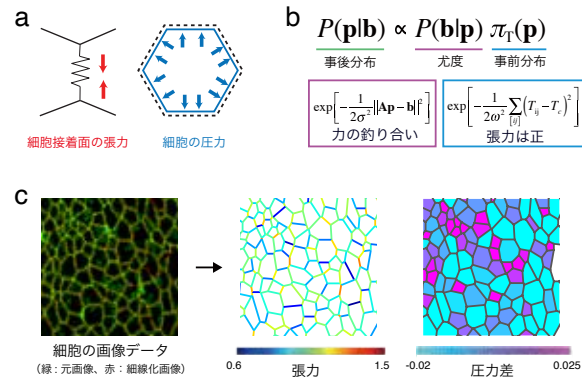


図1 力のベイズ推定法

ジングや次世代シーケンシングなどの計測技術が発展し、大量の定量データが得られるようになった現在、その重要性を飛躍的に増しています。私たちが開発してきた画像データから力学情報を推定する手法は、生命システムの挙動を支配する物理法則・化学法則を統計機械学習に統合するアプローチの有効性を示しています。

「力」を介した生体内情報処理原理の分子メカニズムを解明

生物学は主に、モルフォゲン勾配や細胞極性に代表される化学的情報処理という視点から、生き物のかたちやパターン形成を理解してきました。私たちは、組織応力場が方向情報をコードすること、すなわち、ショウジョウバエの翅上皮細胞が組織引張り応力を感じて、細胞を並び替える向きを揃えることで、細胞の六角格子化が促進されることを見出しました。

ここ数年で、脊椎動物、無脊椎動物の多くの組織で、組織応力場が個体発生を司ることがわかってきました。引張り応力による細胞六角格子化の促進もその一例です。この機構が働くためには細胞が力を感じる仕組み（力感知）と細胞が組織応力による機械的な負荷に打ち勝って動的に変形する仕組み（力抵抗）が必要です。しかし、組織応力駆動型の発生プロセスにおける細胞の力感知と力抵抗の分子メカニズムの多くは不明でした。私たちは細胞六角格子化過程を解析対象に用いて研究を進め、1. マクロスケールの力の場の情報がミクロスケールのアクチンフィラメント構造に変換される、2. アクチン脱重合因子が細胞の力抵抗と動的な変形に最適なバランスを実現するという二つの新しいコンセプトを提唱しました。組織の変形は必ず細胞の力感知と力抵抗を伴うことから、本研究で明らかになった分子メカニズムは発生・恒常性維持の様々な局面で働いている可能性があり、今後の研究の発展が期待されます。

学生へのメッセージ

実は、小学校5年生にして「人生詰んだ」と思っていました。中学受験のための塾に通い始めたところ、成績はよいものの、抜群に得意科目がなかったからです。人生で何かを成し遂げる人には突き抜けたものがあるはずで、自分には何もないのだと。何年か経ち、文理問わず、いろいろな科目をそれなりにこなせるのもある種の能力なのではと思うようになりました。物理学の視点で生き物を理解したいと漠然と考え始めていたときでもあり、複数の学問を組み合わせれば、自分にも面白いことができるかもしれないと希望がもてました。いわば、合わせ技一本を狙う戦略です。その後、苦労もたくさんしましたが、16歳のときにやりたかったことをやれているので、合わせ技一本狙いは大失敗だったというわけでもなさそうです。

学際的アプローチで生命の謎に迫る生物情報科学はまさに、合わせ技一本の学問です。学術分野の壁を飛び越えて新しい思考や技術を柔軟に取り込み、複数の学術分野を統合して研究を大きく展開することが求められます。皆さんも生物情報科学科で「合わせ技一本」に挑戦してみませんか？

Yu TOYOSHIMA

環境に応じて行動を生み出す 神経回路の情報処理のしくみ

理学系研究科 生物科学専攻 准教授
豊島 有

神経は自発的に活動しているようにみえることや、線虫の動きとの相関が強いことなどが明らかになりました。そこで私たちは、4D 顕微鏡を電動ステージと組み合わせて tracking 4D 顕微鏡を構築し、自由行動中の線虫を自動的に追尾しながら全脳神経活動を計測しています（図2）。また独自に開発した微小流路を用いて、特定の姿勢や行動を誘発しながら神経活動を計測することもできます。このようにして神経活動と行動を同時に計測することで、行動と関係する神経活動のパターンを特定し、その生成メカニズムを調べています。

数理モデル化とシミュレーション

神経細胞はお互いに接続して複雑なネットワークを形成しています。また神経活動は筋肉の収縮を通じて姿勢や行動を生み出しますが、その結果は固有受容感覚や外部環境の変化として、神経系へフィードバックされます。こうした多階層・多要素のシステムのダイナミクスを理解するには、定量的な数理モデル化とシミュレーションが重要です。

線虫は酸素濃度に勾配のある環境下では特定の酸素濃度に移動する性質を示します。私たちは、新たな実験手法と数理モデル化手法を開発して、酸素受容神経の神経活動を幅広い酸素濃度域にわたって忠実に再現することに成功しました。さらにこの神経モデルと既知の神経科学的知見に基づいて、酸素濃度の感知からナビゲーション行動までの実験データを忠実に再現する包括的な数理モデルを開発しました（図3）。

また深層学習を使った時系列モデルによって線虫の行動パターンを予測し、光遺伝学的な摂動と強化学習を組み合わせ、線虫の行動を制御する戦略を自動的に学習させることにも成功しました。

メッセージ

私たちは、ゲノム編集や一細胞 RNA 配列解析、実験自動化などの生物学的な技術開発も進めています。さらに、数理モデルや画像解析、機械学習などの情報科学的手法を積極的に活用し、生物学的手法と融合していきます。生物情報科学の新たな地平を一緒に切り開く、強い志のある方を歓迎します。

脳・神経系は生物がもつ最も高度な情報処理システムです。私たちの研究の目標は、脳・神経回路における情報処理の動作原理を明らかにすることです。感覚器を通じて取得された外部環境の情報は、神経回路において記憶などの内部状態と組み合わせて処理され、適切な行動の生成につながります。このときの神経活動と行動のダイナミクスを網羅的に計測し、相互の関係を定量的に調べることが、動作原理の解明に必要です。

生物は周囲の匂いや音などを頼りにして、餌場など好みの環境へ移動するナビゲーション行動を示します。こうしたナビゲーション行動は、外界の環境を感知して必要な情報を取捨選択し、行動として出力するという入出力関係が明確であり、神経回路における情報処理のしくみを明らかにするのに適した現象です。私たちが研究に用いている線虫 *C. elegans* は、餌とともに経験した塩の濃度を記憶し、塩濃度勾配のある環境ではその塩濃度の領域に向かいます。また線虫の神経回路は 302 個の神経細胞から構成されており、それぞれの神経細胞の特徴や互いの接続も詳しく調べられています。しかしこのように、神経科学の基盤的情報が最も充実した生物であっても、それぞれの神経細胞がどのように情報を処理して行動を生み出しているかという、神経回路の動作原理については、あまりよくわかっていません。

全中枢神経活動の同時計測

私たちは、全神経活動の同時計測が、神経回路の動作原理を解明するためのブレイクスルーになると考えました。線虫の体は小さく透明なので、微小な流路に線虫を閉じ込めて、蛍光顕微鏡で、神経活動を生きたまま計測することができます。そこで私たちは線虫の全中枢神経の活動を同時に観察できる 4D 顕微鏡などの実験技術（図1）と、撮影された立体動画画像から神経活動を抽出する画像解析手法を開発してきました。また抽出された神経活動を個体間で比較したり、神経回路に対応付けるために、画像中の個々の細胞を既知の神経と対応付ける細胞同定の手法も開発してきました。こうした技術を統合し、全中枢神経活動の同時計測を実現しました。

神経活動と行動の同時計測

全脳神経活動の時系列データを独立成分分析などの手法で解析したところ、外部から与えた刺激はごく一部の神経にしか影響せず、他の大半の



図1 4D 顕微鏡

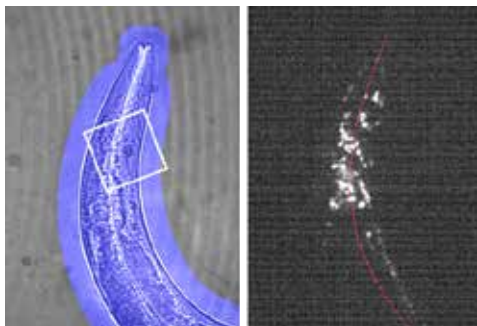


図2 自由行動中の線虫の全脳イメージング

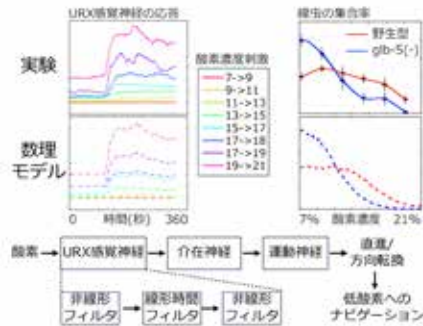


図3 酸素走性行動の包括的な数理モデル

生命を理解するための 数理モデルとアルゴリズム

新領域創成科学研究科 メディカル情報生命専攻 教授
浅井 潔

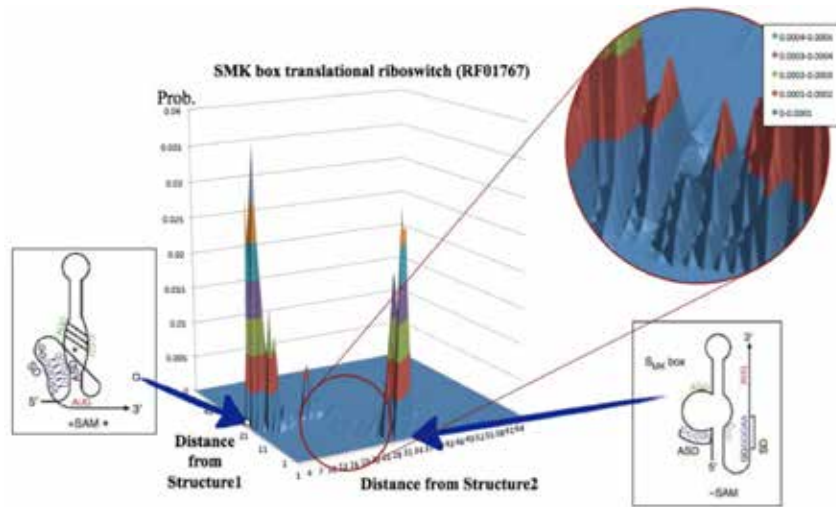
生命現象の全体像はさきわめて複雑で、実験で確かめたり、数理的なモデルで説明したりすることは限界があります。一方、ゲノム DNA 配列やその転写物である RNA 配列は、新型のシーケンサーで超高速に同定することができます。浅井研究室では、塩基配列情報を主な対象として、生命の理解につながる情報を得るための、確率的な基盤に基づいた理論とアルゴリズムの研究を行っています。また、アルゴリズムをソフトウェアとして実装し、公共データや共同研究先のデータの解析に活用しています。

ゲノム情報解析

ゲノム塩基配列は、単なる文字列ではなく、その背後には物質としての構造があり、複雑な生命現象のメカニズムが隠れています。「見る」ことができる文字列の背後に、「意味」が隠されているのです。ゲノム配列は、我々が手にすることができる生命情報の中で最も精度の高いデータですから、ゲノム配列に隠された「意味」を見出す研究は極めて重要なのです。浅井研では、新型の高速シーケンサーの巨大配列データから、信頼できる情報を効率的に抽出するための手法を研究しています。また、アルゴリズムをソフトウェアとして実装し、公共データや共同研究先のデータの解析に活用しています。

RNA インフォマティクス

RNA 干渉やマイクロ RNA による翻訳抑制の発見以降、ゲノムから転写されるがタンパク質には翻訳されず、RNA として機能を発揮する非コード RNA が、様々な細胞内の現象に重要な役割を担っていることが明らかとなりました。細胞内の RNA の機能は、RNA が形成する 2 次構造と密接な関わりがあるため、塩基配列から RNA2 次構造を高精度に予測することが重要です。我々は、信頼性の高い RNA 2 次構造予測ソフトウェア CentroidFold をはじめ、RNA 配列・構造の情報解析で世界をリードする理論・ソフトウェアを数多く生み出してしてきました（<http://www.ncrna.org>）。RNA の構造は室温でも熱揺らぎにより変化しており、最も



安定な 2 次構造ですら存在確率は極めて小さいので、RNA 2 次構造の確率分布を様々な角度から解析する手法を研究しています。

近年、ゲノム DNA の修飾が細胞分化などに重要な役割を果たすことが明らかになりましたが、RNA でも転写後の修飾が構造や機能に大きな役割を果たしています。現状では困難な、修飾を受けた RNA の構造予測に取り組んでいます。その結果は、RNA 情報解析のツール群に組み込んで RNA 修飾の役割の解明に役立てたいと考えています。

RNA 情報解析は近年に進歩したため、大規模なゲノム情報解析には、最新の RNA 情報解析技術を駆使することによって解決しそうな研究テーマが数多く残されています。

ゲノム配列設計

微生物に目的物質を効率的に産生させることを目的として、人工的なゲノム配列を設計する研究も行っています。これまで、抗体を微生物に産生させるための DNA 配列設計技術の研究や、DNA 配列の膨大な組み合わせを実験で生成し、機械学習などの人工知能技術と組み合わせ、有用物質の生産に最適な DNA 配列を探索する技術に取り組んできました。これらの DNA 配列設計では、遺伝子群の mRNA への転写効率とともに、mRNA の翻訳効率も目的とする物質の生産が向上するように最適化しなければなりません。配列情報解析技術、RNA 情報解析技術をさらに発展させるだけでなく、翻訳効率と RNA の構造の関係などにも着目しています。

秘匿化計算技術

個人ゲノムの塩基配列情報など膨大なデータからは、機械学習など最先端の人工知能技術を駆使して、価値のある情報を得ることが期待できます。そこで、個人情報や知財情報などの秘匿すべき情報を漏えいすることなく、安全に解析する技術として、データを暗号化したまま処理する秘匿化計算技術が注目されています。CREST のプロジェクトでは、多様なアプリケーションを秘匿化情報処理として社会実装することを可能とする「汎用秘匿化依頼計算技術」の研究を行っています。

研究環境

浅井研究室では、学生の自主性を尊重し、学生自身が興味を持つテーマを掘り下げ、研究室内外のメンバーとの議論を通じて研究を進めていく方式を取っています。数理的な理論や、プログラミング技術を習得しながら、学生自身の興味と資質に応じて研究を進めていくことができます。生物学の知識の乏しい学生でも、研究テーマが見つけれられるように配慮しています。

研究活動では、(国研)産業技術総合研究所の人工知能研究センター（お台場）と連携しています。お台場では、専門の研究者と日常的に交流し、研究討論やセミナーにも参加することができます。

情報学・数学・観測技術の 魅力的な実験場としての生命科学

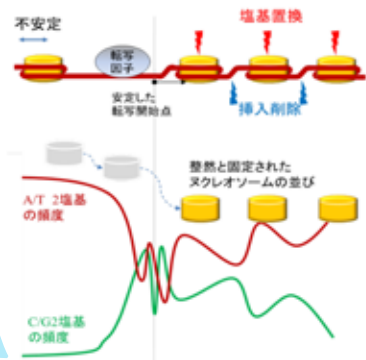
新領域創成科学研究科 メディカル情報生命専攻 教授
森下 真一

Shinichi MORISHITA

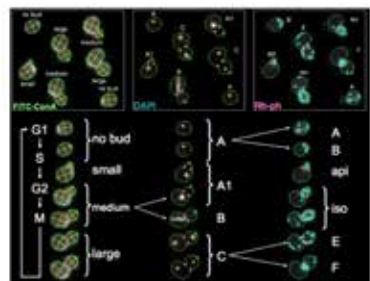
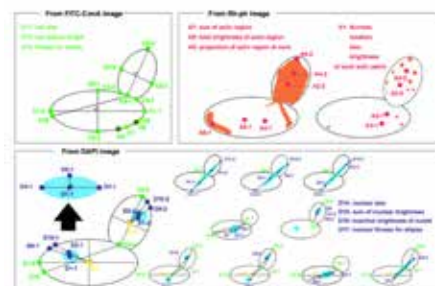
生命科学は、情報学・数学・1分子計測技術の魅力的な実験場です。私たちは、これらの学問を生命科学に取り入れ、DNA とオミクス周辺の生命科学の研究課題に取り組んでいます。

情報学・数学からのアプローチ

数学的思考等の知的活動を計算機で自動化することには限界があることは、たとえばゲーデルの不完全性定理等で知られています。この限界を意識しつつ、自動化できる範囲が機械学習、データマイニング等の情報学分野で徐々に広がっています。一方、私たちが取り組んでいるのは、生物における自動化メカニズムの理解です。たとえば、ゲノムがどのように情報をコードして細胞や組織を構築し、また進化してきたか？ 核内でDNA はどのように畳み込まれ遺伝子の発現を制御しているか？ などを深く知りたいと考えています。このような歩みから、生命科学は情報学と数学の魅力的な実験場であると感じています。特に最適化アルゴリズム、並列計算、機械学習、定理証明技術、計算幾何学 代数幾何学は、生命における自動化の仕組みを分析することに役立ってきています。



クロマチン構造が遺伝子転写や遺伝的多様性と相関することを報告したが、そのメカニズムは？ (出典 Sasaki et al. Science, 323(5912):401-404, 2009)

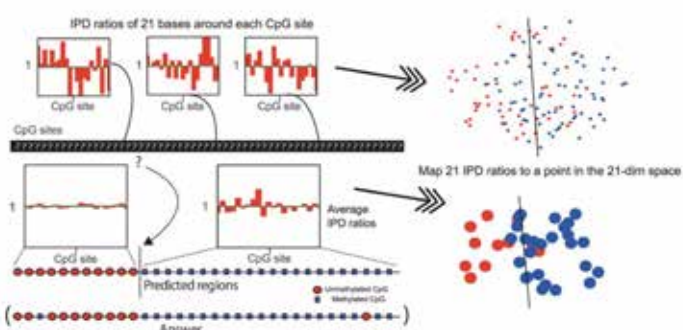


表現型の幾何学 - 出芽酵母遺伝子破壊株の1細胞画像分析 - (出典 Ohya et al. PNAS, 2005)

1分子・1細胞を計測する技術の開拓

新しい観測技術が新しい発見を導くことが多いことはご存知かと思いますが、1分子や1細胞の動きを直接観測する技術はそのような例です。1分子計測技術を研究開発しているシリコンバレーの企業 Pacific Biosciences 社と、私たちは長年共同で研究しています。同社の Zero-Mode Waveguide (ZMW) 技術は1分子計測の傑作で、たとえばDNA ポリメラーゼの動きを実時間で直接観測できるため、これまで観測が困難だった難読領域の CpG メチル化を測定するアルゴリズムを一緒に研究開発しました。

現在は、メタゲノム(細菌叢)および多細胞生物(線虫、ヒト等)で、ZMW を用いて多様な DNA 修飾(6mA, 4mC 等)を観測し、それらの機能を分析しています。



難読 DNA 領域では、どのような DNA 修飾(例 CpG メチル化)が起こっているのか? 1分子計測のノイズを除去し、精度良く観測する最適化アルゴリズムを導いた。(出典 Suzuki et al. Bioinformatics 32 (19): 2911-2919, 2016)

DNA とオミクス周辺の生命科学の研究課題

21世紀に入り、ヒトを始めとする巨大DNAが解読されるようになり、いろいろな謎が解明されています。たとえば遺伝学者の大野は、脊椎動物祖先で2回の全ゲノム重複が起こったと1970年に予測しました(大野の仮説)。この全ゲノム重複が遺伝子集合を豊かにし、脊椎動物の繁栄をもたらしたと考えられています。私たちは魚類(メダカ)ゲノムを解読し、哺乳類ゲノムと比較することにより、大野の仮説を立証しました[Nature, Genome Res 2007]。

その後は、ヒト個人ゲノムの解読に力点を移し、東大病院にて約5千人の個人ゲノムを解読し、脳疾患および癌に関連する遺伝子変異を分析しています。このような解析は、一昔前までは難しかったのですが、現在では容易になりました。いま取り組んでいるのは、解読困難なDNA領域の解明です。難読なセントロメア配列、ヒト腸内細菌叢、個人特異的な繰返し配列(LINE, LTR等)などを研究対象にしています。最先端のDNAシーケンサー(第三世代・長鎖型PacBio Sequel, Oxford nanopore, 10X Chromiumおよび次世代・短鎖型Illumina HiSeq)を利用し、研究室内でDNAデータを収集しています。特に、数万塩基の長いDNA断片を解読できる長鎖型は、謎の理解に役立っています。



Yutaka SUZUKI

研究は自分の価値観。大切に でも修正することを恐れずに。

新領域創成科学研究科 メディカル情報生命専攻 教授
鈴木 穰

当研究室では、ヒトゲノム配列に生じた多型、突然変異がどのような生物学的意義を持って、最終的に疾患の病因として寄与するのか、エピゲノム、トランスクリプトームといった様々なオミクス階層について、実験的、情報学的手法を駆使し多角的に解析を進めている。

近年の次世代シーケンサー技術の進展により、ヒトの全ゲノムあるいは全エキソーム解析(後述)から疾患関連遺伝子を同定、あるいは診断に応用しようという試みが急速に加速している。しかし見出された変異がどのような分子機序を持って疾患に寄与しているのか、その生物学的意義は必ずしも明らかではない。一方で、圧倒的なシーケンス産生能力を背景に、DNAのメチル化部位、あるいは特定の修飾を受けたヒストン部位および種々の転写因子に結合するゲノム部位を解析するいわゆるエピゲノム解析、あるいは転写産物を解析するトランスクリプトーム解析も広く行われている。当研究室では以下に示す応用について、多階層から得られたオミクスデータの統合を行うことで、疾患と関連付けられたゲノム変異の生物学的意義の解明を試みている。また同時に、次世代シーケンス技術、シングルセル解析技術、1分子シーケンス技術等の新技術を基盤に、現在、計測することができないオミクス階層について新解析技術の開発を行い、国内有数のシーケンス拠点として全国の研究者に提供している。

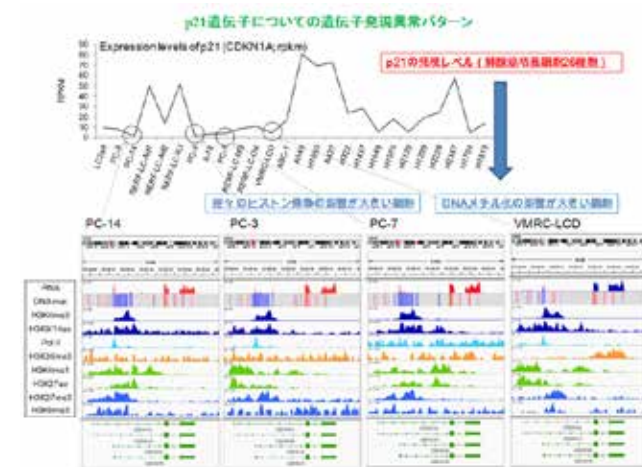
鈴木研究室一次世代シーケンサーを用いた 遺伝子発現制御因子の網羅的理解



1. がんゲノム解析

当研究室では多くの病院、臨床研究機関と連携して、肺がん、大腸がん、食道がんをはじめとして多くのがん種について、ゲノムに生じる体細胞突然変異の同定を行ってきた。その結果、p53, KRAS, EGFR遺伝子等の代表的な遺伝子を例外に、多くのがん症例においてそれぞれに共通する変異遺伝子はまれであることが明らかになった。症例間相互の共通項の比較が困難であるために、発がんの主たる駆動力になる「ドライバー変異」と、がんにおけるゲノムの不安定化の結果生じる機能的に中立な「パッセンジャー変異」から区別するのは依然として困難である。特に遺伝子発現制御に異常をきたす変異については、重要性が指摘されつつもその同定から検証にいたる方法論に定式がない。当研究室では、培養がん細胞株をモデルにゲノム変異、エピゲノム変異、トランスクリプ

トーム変異を相互に対応させて計測する系を構築、その相互依存性を検証することで、これらの問題解明に取り組んでいる。また、それらががんの転移、薬剤耐性の獲得といったがんの再発時にどのように改変されるのかについてのデータ収集、データベースの公開を進めている。



2. 汎遺伝子発現制御についての 計測技術開発とモデル化

近年のトランスクリプトーム研究から、遺伝子発現は転写開始段階だけでなく、RNAの伸長、移送、分解の各段階において、精密に制御されていることが明らかになっている。当研究室では、最新のオミクス計測技術を基盤にこれらの制御要素の各段階において対応する新技術を開発、得られたデータを解析することでそのモデル化を行っている。例えば、ゲノム塩基の置換がどのようにプロモーター活性に影響を及ぼすか、人為的に変異を導入したDNA配列について飽和規模で配列-活性相関を測定する実験系を開発し、機械学習等の手法を用いてデータ解析を行うことで予測モデルの構築を行っている。最終的には、遺伝子発現量を規定する要素としてのクロマチン構造、mRNA分解速度等についてもモデルに取り込み、汎発現制御機構についてのモデル化を目指す。これにより臨床検体において見出されるゲノム多型・変異について、モデルの演繹から生物学的意義を推定することが可能になると考えている。

3. 野外株を用いた感染症ゲノム解析

感染症ゲノム解析の分野において、野外株感染時のヒト免疫細胞の応答は、研究室環境でモデル化したものと大きく異なる。本研究室はインドネシアに野外活動拠点をもち、特にマラリア原虫感染時の寄生虫-宿主応答を中心に多階層オミクス解析を行っている。同一環境の中で相克する生物種間での遺伝子発現プログラムの相互干渉を明らかにしようという試みである。

参考文献
次世代シーケンス目的別アバスタリファレンス: 秀潤社.: 菅野純夫、鈴木穰
RNA-Seq 実験ハンドブック: 羊土社: 鈴木穰

生命の生み出す大量のデータに
機械学習で立ち向かう

新領域創成科学研究科 メディカル情報生命専攻 教授
津田 宏治

Koji TSUDA

本研究室では、大量かつ多様なデータから、信頼できる知識を高速に発見するための数理手法に関する研究を行う。データ中心科学の拠点として、生物学、化学、薬学、エネルギー・環境など、どのような科学分野からのデータでも、その本質を素早く見抜き、高度なアルゴリズムを駆使して、重要な知見を発見することができる人材の育成を目指す。

組合せ効果の多重検定法

自然科学で得られるデータ量は増加の一途をたどり、これらを有効に解析できる方法が望まれている。しかし、従来の統計検定手法では観測できる対象が増えれば増えるほど、発見の基準を厳しくしなくてはならない。特に、複合的な組み合わせ因子に対して極めて保守的な検定値（P値）を出すことが多く、有意義な実験結果が不当に低く評価されることがあった。本研究室では、超高速アルゴリズムの技法を用いて、従来法より、格段に精度の高いP値を算出する新手法 LAMP (Limitless Arity Multiple-testing Procedure)を開発した[1]。転写因子の組み合わせ効果の研究をはじめ、複数の遺伝子が原因となっている疾患の同定や多数の部位が関わる脳の高次機能の解明など、複合要因に起因する現象の解明が加速されることが期待される。

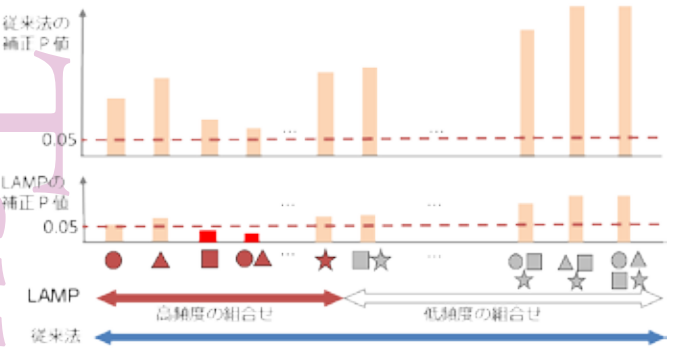


図1 LAMPによる多重検定

分子・物質の自動設計技術の開発

核酸・タンパク質などの生体高分子や、物質・材料研究における金属・セラミック・ナノ粒子など、望みの機能を持つ分子を設計することは、科学的・産業的に大きな意義を持つ。津田は、理化学研究所革新知能統合研究センター分子情報科学チーム長を兼任しており、機械学習技術による分子・物質の自動設計の研究を行っている。ここでは、ベイズ最適化などの機械学習技術を用いて、物理シミュレーションや、合成実験の計画を行うことによって、新規分子・物質開発の効率を向上させることを目的としている。本研究室では、実験科学者が簡単に使うことができるベイズ最適化パッケージ COMBO[2]を開発するなど、自動設計手法の普及に努めている。

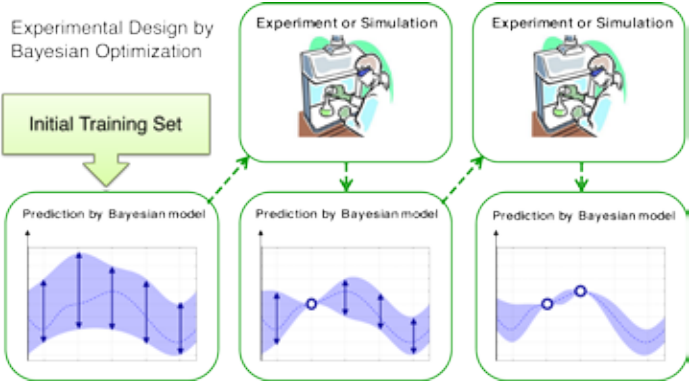


図2 機械学習による分子・物質の自動設計

機械学習・データマイニングの基礎研究

本研究室では、生命科学での応用に向けて、高速な機械学習・データマイニングの基礎研究も行っている。例えば、分子構造などグラフ構造を持つデータから、毒性などの性質を予測し、同時に、重要な役割を持つ部分構造を特定するアルゴリズム gBoostの開発を行った[3]。また、対象データから高速に類似ペアを発見するアルゴリズムにも力を入れている。これらの手法は、ゲノム・エピゲノム・代謝物データなど、マルチ・オミックスデータの解析に利用できると考えられる。

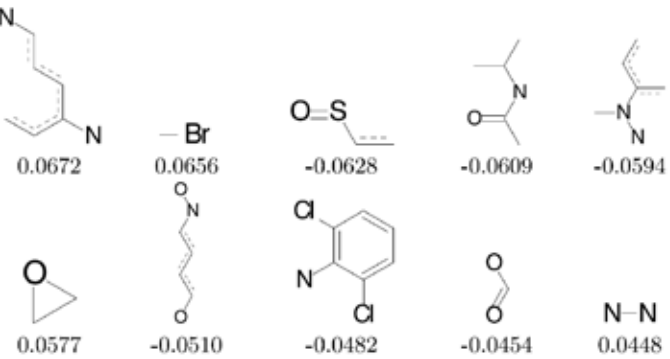


図3 gBOOST[3]によって発見された部分グラフ特徴

参考文献
[1] A. Terada, M. Okada-Hatakeyama, K. Tsuda* and J. Sese*. Statistical significance of combinatorial regulations, Proceedings of the National Academy of Sciences of the United States of America, 110(32):12996-13001, 2013.
2009.
[2] T. Ueno, T.D. Rhone, Z. Hou, T. Mizoguchi and K. Tsuda. COMBO: An efficient bayesian optimization library for materials science, Materials Discovery, 4:18-21, 2016.
[3] H. Saigo, S. Nowozin, T. Kadowaki, T. Kudo, and K. Tsuda. gBoost: A mathematical programming approach to graph classification and regression. Machine Learning, 75:69-89,

Martin FRITH

Finding interesting things in
genetic sequences

新領域創成科学研究科 メディカル情報生命専攻 教授
Martin FRITH フリス・マーティン

We are entering the age when we know the genetic sequences of many living things. These DNA, RNA, and protein sequences are the blueprints of life: they hold the instructions for making a brain, for maintaining a healthy organism, and for the functioning of ecosystems (interactions between organisms). They have evolved for ~4 billion years by gradual changes and repurposing of genetic elements: so these sequences hold evidence of their evolutionary history.

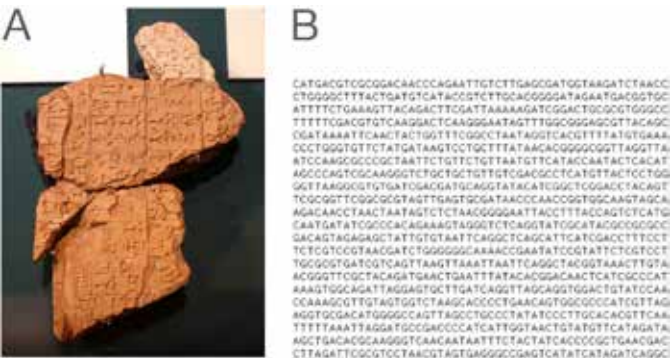


Fig.1 A The Instructions of Shuruppak, ~4500 years old. B Genome, up to ~4000000000 years old: parts have remained recognizably similar for billions of years.

Our research group aims to develop and use methods for finding interesting functional and historical information in these sequences. Everyone's own creativity and original ideas are especially welcome; breakthroughs will likely come from weird or overlooked aspects of genetics and biology.

Genetic fossils

Genomes contain decaying relics of ancient genes, which often come from long-ago insertions of DNA from viruses or virus-like entities (transposable elements). These "fossils" enable paleovirology: the study of ancient viruses and their evolution. We have developed methods to find extremely old and decayed remnants, which are revealing a diverse ecosystem of transposable elements in our aquatic ancestors.

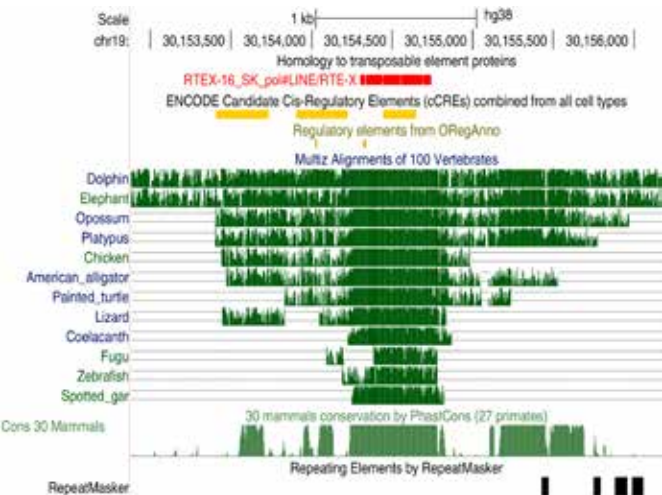


Fig.2 An ancient genetic fossil (red) in human chromosome 19, which has been repurposed as a regulatory element.

Wataru IWASAKI

生命現象を俯瞰的な視点から
捉えなおす新領域創成科学研究科 先端生命科学専攻 教授
岩崎 渉

生物学における未発見の法則・可能性を求めて

私たちの研究室では、特定の生物や個別の問題にフォーカスした従来の生物学から、俯瞰的な「データサイエンス」へと変貌する生物学への大きな流れに注目しています。これまでアプローチできなかった非モデル生物や生命現象を自在に研究できるようになる中で、バイオインフォマティクス、理論・数理、実験、フィールドサンプリング、また他分野の技術や手法も含めたあらゆるアプローチを取り入れつつ、以下の5つの本質的な問いに挑むことをミッションとしています。

- (1) 生命システム・ゲノム進化学
極めて精巧で複雑な生命システムはどうして進化できたのか？
- (2) エコシステム戦略学
生態系はどうして今の姿で安定に維持されているのか？
- (3) 機能未知遺伝子学・DNA 配列空間学
40 億年の生命の歴史はどのような未知の遺伝子機能を進化させたのか？
- (4) 対偶遺伝学・二次元オーミクス
ビッグデータから未知の遺伝子機能に迫るための方法論は？
- (5) 遺伝子誕生学
この広い地球上で遺伝子は今どれだけ生まれているのか？

生命システムとゲノムの進化

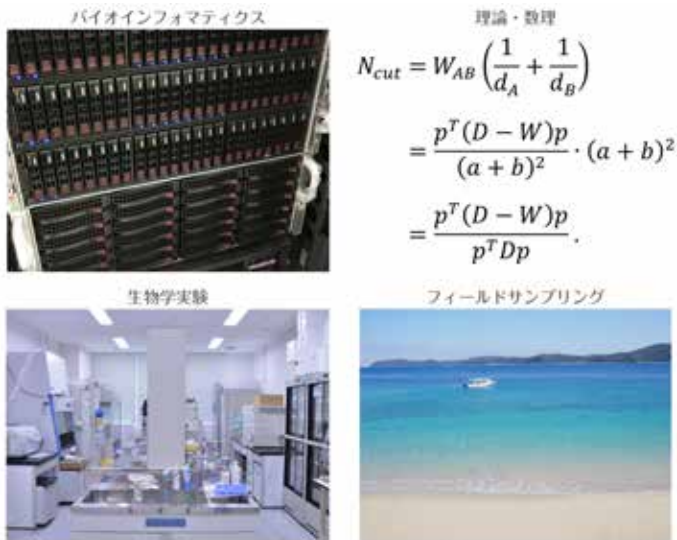
ゲノム情報は生命活動の礎となるものであり、また、祖先生命から私たち現代の生命に至る長い歴史の記録でもあります。ゲノム進化解析によって、生命の共通祖先から現代に至る生命がどのように進化してきたか、ゲノムやそこにコードされた生命システム・遺伝子ネットワークはどのように進化してきたか、さらに、生命と地球が長い時間の中でどのような歴史を相綴ってきたか、などを明らかにするための研究を行っています。

エコシステム戦略学

環境中の微生物群集（マイクロバイーム）をまるごと分析するメタゲノム解析や、生物が環境中に排出する DNA を分析する環境 DNA 解析など、生態系を解析するための技術の革新によって、近年、これまでは決して得られなかった豊富なデータを得ることが可能になりました。私たちの研究室では、新たな技術開発を行うとともに大規模生態系データを独自の視点から解析することで、生態系のダイナミクスが生命と環境のどのような相互作用により生み出されているか、探求しています。

機能未知遺伝子学から遺伝子誕生学へ

現代生物学における最大のボトルネックの一つであり、さらに、医学・薬学・農学等への応用の観点からも巨大なポテンシャルを秘めているフィールドが、多様な微生物ゲノムにコードされた機能未知遺伝子の機能解析です。私たちの研究室では、既存の研究手法ではアプローチが難しいこのフィールドに挑戦しています。さらに「いま、この広い地球上で



情報も実験もわかる研究者を目指す



研究室でのディスカッションの様子

遺伝子は今どれだけ生まれているのか？」という問いに答える未来の研究領域としての「遺伝子誕生学」を切り開こうとしています。

学科に参加される方へのメッセージ

既存の学問分野をまたぐ幅広い視点を踏まえた研究に挑戦したい方、特に、自分自身のユニークなアイデアに基づいて新たな領域を開拓することを強く志す方を歓迎します。学問を志すやる気と自ら考えて行動する積極性に加えて、柔軟な発想と遊び心、さらに、日本語か英語かによらず自分の考えを明確に表現しつつ仲間と議論・協調していく姿勢を持つことを期待しています。

Hisanori KIRYU

人工知能・機械学習技術を用いて、
数理的な生命モデルを構築する新領域創成科学研究科 メディカル情報生命専攻 准教授
木立 尚孝

木立研究室では、人工知能 (AI)・機械学習などの高度な情報技術を用い、爆発的に増え続ける生物データを計算機で解析し生物学的発見を行う、バイオインフォマティクス（生物情報学）研究を行なっています。また、このような解析が必要となる、計算機上の生物モデルの構築にも力を入れています。

1990 年代に初めて生物の全ゲノムが解読されてから、これまで微生物からヒトまで数千の生物種についてゲノムの解読が行われました。大規模なデータ計測の対象は、ゲノム配列にとどまらず、RNA、タンパク質、代謝物質、DNA の化学修飾、細胞動画像、など、生命活動を構成するさまざまな要素へ広がっています。しかし、これらの大量データの包括的な解析は始まったばかりで、まだまだ多くの課題が残されています。

私たちは、人工知能・機械学習技術を用いて、生命の高度な数理モデルを構築することにより、統合されたシステムとしての生命の深い理解や、生命を細胞レベルで自在に制御する技術につなげることを目指しています。

一細胞解像度の遺伝子発現データから
細胞分化進行度を推定

シーケンシング技術の進歩により、様々な条件下で細胞がどのような遺伝子を活性化するかを網羅的に計測できるようになっています。特に最近では一細胞 RNA-seq 法により、数万の細胞の一つ一つについて RNA 活性プロファイルを測定することが可能になりました。私たちは、まず、オルンスタイン・ウーレンバック過程と呼ばれる確率過程で細胞の内部状態をモデリングし、その未知パラメータを機械学習により最適化する手法を開発しました。次に、刺激を与えられた細胞の分化過程の一細胞 RNA-seq データにこの手法を適用し、各細胞の分化進行度を推定する新しい技術を開発しました。

植物生長データからマクロな環境要因と
ミクロな細胞状態との相関をモデリング

現在、主に次世代シーケンサー技術により細胞レベルのデータが爆発的に蓄積しています。しかし、これらのミクロスコピックなデータから、生体組織の特性や個体の行動など、生物のマクロスコピックな性質を予測することはほとんどできていないのが現状です。一方、一般に植物は動かないため動画による形態計測が比較的容易で、また遺伝子発現データと外部環境（気温、雨量、日射量等）との相関が見やすいなど、数理モデリングがしやすい特徴を備えています。そこで私たちは、制御工学の分野で発展し飛行機自動制御やカーナビで用いられている、カルマン・フィルターの理論を、現代のデータサイエンスの時代に適した形へと刷新し、イネなどの植物のミクロの細胞状態とマクロの環境要因とを機械学習により関係づける手法の開発を行っています。

がんゲノムの進化モデリングによる、
がん進行履歴の推測

がんは細胞が無制限に増殖する病気であり、DNA への変異の蓄積がおも

な要因です。多くのがん種では、細胞分裂のたびに、ゲノム中にさまざまな塩基変異を蓄積します。このがんゲノムの変化の過程は、生物種分化の際のゲノム進化と似ており、進化学や遺伝学的手法を用いて、がんの進行過程を調べることができます。私たちは、集団遺伝学で使われるライト・フィッシャーモデルや合祖理論を用いて、がんゲノムのシーケンシングデータから、がんの増殖過程を推定する手法の開発を進めています。また、推定されたがん組織の定量的なデータからがんが転移・再発する確率を予測する手法の開発を目指しています。

RNA 構造の情報解析と
エピ・トランスクリプトームの解明

DNA がもつ遺伝情報は、DNA からメッセンジャー RNA が転写され、さらにタンパク質へと翻訳されて、実際の生理的機能を発揮します。この反応過程は定常的ではなく、あらゆるステップで反応を促したり遅らせたりする制御機構が働き、外部環境変化や刺激への柔軟な細胞応答が可能となっています。特に最近では、メッセンジャー RNA がタンパク質に翻訳される過程に働く様々な制御因子（RNA 結合タンパク質の結合状態、RNA の化学修飾位置、RNA の二次構造）の測定が行われるようになり、翻訳制御の総体（エピ・トランスクリプトーム）を解明する研究が重要となっています。私たちは、確率文脈自由文法 (SCFG) とよばれる情報科学理論に基づく RNA 二次構造モデルと、次世代シーケンサーからのデータを組み合わせることにより、RNA の化学修飾による二次構造変化や、RNA 結合タンパク質が結合しやすい二次構造モチーフは何か、などを推定する手法を開発しています（図 1）。

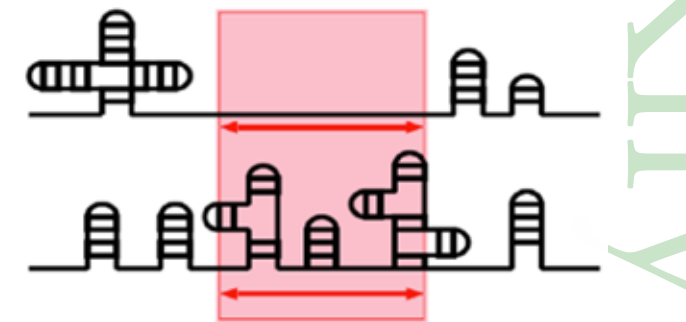


図 1 RNA のアクセシビリティを計算するツール Raccess によるゲノムスケールの配列解析。Raccess は RNA 転写物の全ての領域について、その領域が二次構造を形成しない露出した領域であるかどうかを確率値として与えます。

研究体制

当研究室は、実験設備を持たないドライ系の研究室であるため、自前で生物実験を行うことはなく、公開データ及び共同研究先の実験データを用いて、計算機を使った研究を行うことになります。

新しい観測技術から 新しいゲノム解析技術を創る

Masahiro KASAHARA

新領域創成科学研究科 メディカル情報生命専攻 准教授
笠原 雅弘

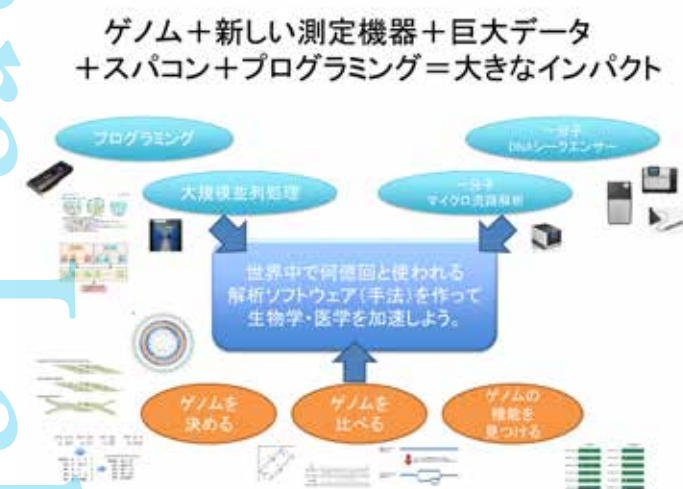


今、ゲノムが最も面白い

一つの面白い生物学的メカニズムを発見した人より、数千の研究室が面白い生物学的メカニズムを倍の速度で発見できる道具を創った人の方が一般的に言って科学により貢献していると思う。Google が凄いのは少人数で開発したプロダクトが世界中に住む数億の人に大きなインパクトを与えているからだ。私はゲノム情報解析の Google 的な存在を目指したい。誰もが当たり前のように便利に使い、無い世界は考えられない、そんな存在に。

ここ数年の DNA シークエンシング技術の発展は、そんな「汎用的な道具」という存在が生物学や生物情報科学にもあり得ることを示してくれた。私が、ゲノム解析で最も面白いと思うところは原因と結果の区別を非常に付けやすいところである。例えば、孤発性アルツハイマー病の患者の脳を観察すると老人斑にアミロイドβが蓄積していることは分かるが、アミロイドβの蓄積によって病気になったのであろうか、それとも病気になったのでアミロイドβが蓄積したのであろうか？ 条件を揃えた介入実験が難しいヒトを含む生物種についてのあらゆる研究にはこの問題が必ずつきまとう。

しかし、ゲノム情報はこの呪縛とは関係ない。「病気になったのでゲノムの塩基配列が変化した」ということは（一部の例外はあるが）ほぼあり得ない。相関と因果の区別は依然として難しいものの、因果の向きはほぼ固定されていることは大きな強みであり、データを増やせば基本的には答えにたどり着ける自信を与えてくれる。また、近年のゲノム編集技術を使えば様々な生物種でゲノム配列への介入実験も可能になっている。そんなゲノム情報の読み取り費用はここ10年ほどで10万分の1ほどに、所要時間も10万分の1未満になっている。数年のうちには1回のシークエンシングコストが数千円レベルのスマートフォン接続型の DNA シークエンサーすら現れる。10年前と比べて10億倍の速度で生まれるゲノム情報を相手に、優れた解析アルゴリズム・ソフトウェアを作って我々が配ることで、世界中の研究者に10億倍の速度で知識を生み出して欲しい。



新しい観測技術から新しい解析技術を作る

DNA 観測技術の発達がキーとなっているのであれば DNA の情報を軸にしてあらゆる生命の情報を整理・分析できるようにしたいと考えるのはとても普通の発想だろう。我々の研究室では PacBio/Oxford Nanopore などの一分子 DNA シークエンサーや 10X Genomics などの一細胞解析機など、次々にあらわれる新しい観測技術から得られるデータを解析する情報技術を研究している。現在は、数千人・数万人規模のヒト全ゲノムを解析するには1億円規模のスーパーコンピュータが必要となっており、指をくわえて傍観すれば、ほんの数年でデータ量は1000倍になり、1000億円のスーパーコンピュータが必要になってしまうだろう。

ゲノム配列を解読する（ゲノムアセンブリ）、遺伝子配列を解読する（トランスクリプトームアセンブリ）、種や個体間のゲノム比較を行う（比較ゲノム・グラフゲノム）、機能や原因を推定する（関連解析）、といった基礎的ではあるが従来はデータ量が少なく大規模には解析できていなかった重要で面白い生物学的問題に対する情報解析技術を、現代の大量観測技術に対応していち早くスケールアップし、観測技術から科学的知見までを短時間で繋げるためのアルゴリズム・ソフトウェアを我々は開発している。ほとんどのソフトウェアはウェブページから公開されている。

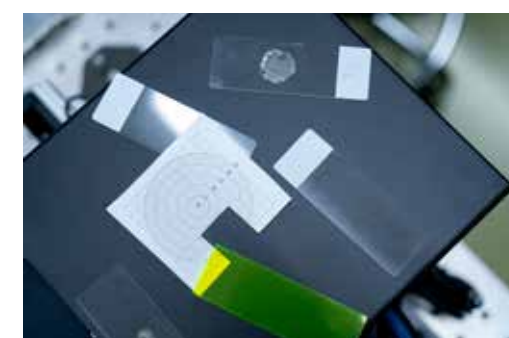
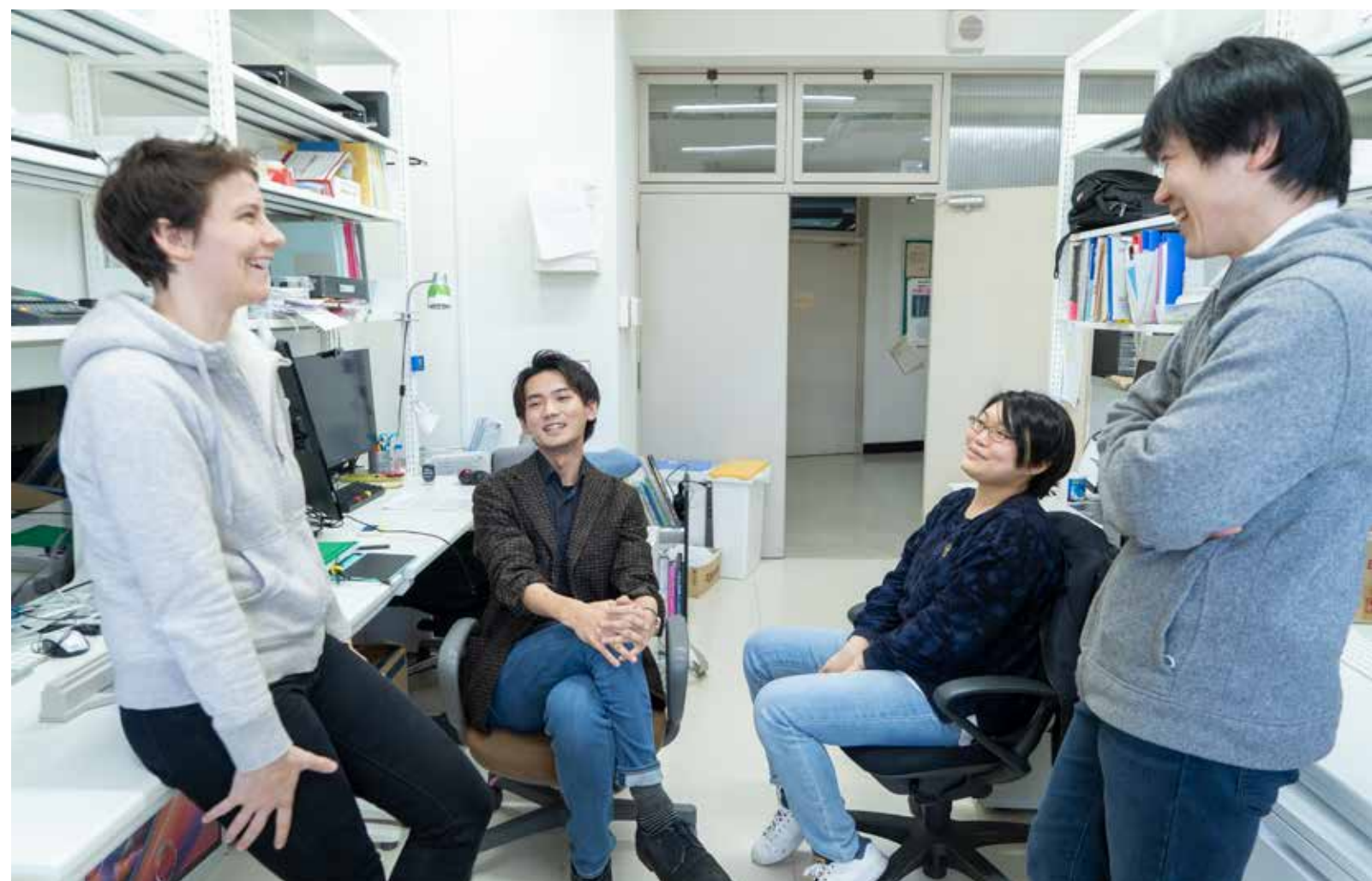
大規模観測データ解析のための プライベートクラウドミドルウェアの開発・研究

データ量の増加にともなって、アルゴリズム開発やプログラミングのコストはそれ以上に大きく増加している。端的に言うと、3～6ヶ月毎に新しい DNA シークエンシング技術が発表されるこの激動の時代には、一つ一つのアルゴリズムやソフトウェアを3ヶ月で仕上げてリリースする必要がある。物理シミュレーションのように3年間掛けてアルゴリズムを高速化することは完全に無駄である。我々は、このような生命科学で発生している「研究者の時間がボトルネックとなる」問題に対処するため、大量の計算資源を並列計算で（ある程度）無駄に使いつつ、計算効率を上げることを至上とする従来型のハイパフォーマンスコンピューティングから脱却し、科学的知見のアウトプットを増やすことを目指す「ハイパフォーマンスサイエンス」を目標とする、研究者に便利な並列分散計算ミドルウェアの開発を行い、実際に我々がその技術を応用して生命科学研究に役に立つことを証明していく。

研究体制など

我々の研究室は、なるべくウェット実験を行っている他研究室と共同研究を行い、最新の観測技術を用いた最新のデータを解析できるエキサイティングな環境を提供することを目指している。自分のペースで進めたい学生のために他研究室とは独立のテーマも希望に応じて用意している。

One day...



学部卒業生の進路 (2016-2020 修了)

卒業人数	大学院進学				海外	学士入学	就職	その他 (未回答含む)
	生物科学専攻	新領域	他研究科	他大学				
67	10	27	5	0	1	0	4	20

修士課程修了者の進路 (2016-2020 修了) ※生物科学専攻のデータより

修了人数	進学		就職	研究生・その他 (未回答含む)
	生物科学専攻	他研究科・他大学		
414	132	2	150	130

博士課程修了者の進路 (2016-2020 修了) ※生物科学専攻のデータより

修了人数	ポストドクトラルフェロー (学振特別研究員など 国内研究機構)	ポストドクトラルフェロー (国外の大学・ 研究所など)	国立研究所・ 民間企業 (研究所含む)	大学 (助教など)・官公庁	就職・研究生・その他 (未回答含む)
221	27	2	39	39	114

学部卒業生の就職先

(2016-2020 修了)

マッキンゼー・アンド・カンパニー／NEC ソリューションイノベータ／
C&G システムズ／警察庁
※「株式会社」表記省略

修士課程修了者の就職先、進学先

(2016-2020 修了)

東京大学／九州大学／琉球大学／西大和学園／厚生労働省／文部科学省／外務省／農
林水産省／環境省／特許庁／国税庁／千葉県警科学捜査研究所／埼玉県警察科学捜査
研究所／国立遺伝学研究所／国立環境研究所／国立国際医療研究センター／理化学研
究所／土木管理総合試験所／野村総合研究所／大和総研／産業技術総合研究所／日本
学術振興会／日本政策金融公庫／参天製薬／東洋新薬／久光製薬／中外製薬／第一三
共／小野薬品工業／アステラス製薬／佐藤製薬／Meiji Seika ファルマ／花王／テルモ
／東京応化工業／ウェルネット／SBI バイオテック／アサヒビール／サッポロビール
／キリン／味の素／協和キリン／ヤクルト／タカナシ乳業／加美乃素本舗／ヒガシマ
ル／昭和産業／ケンコーマヨネーズ／日本アルトマーク／日本水産／東レ／富士フ
ィルム／日立製作所／カネコ種苗／木屋製作所／日本郵便／東京電力エナジーパートナ
ー／NTT コミュニケーションズ／伊藤忠商事／日本 IBM／富士通／三井住友銀行／
三菱東京 UFJ 銀行／クレディ・スイス証券／住友生命／ボストンコンサルティンググ
ループ／日本ユニシス／新興出版社啓林館／博報堂／日本たばこ産業／ドコモ・シス
テムズ／mars&co／東陽テクニカ／オー・ジー／響国際特許事務所／ビズリーチ／
住友林業／PwC あらた有限責任監査法人／WDB エウレカ／サーモフィッシャーサ
イエンティフィック／モリコーポレーション／農林中央金庫／マッキンゼー・アンド
・カンパニー／アカツキ／出光興産／National Innovation Agency, Thailand／パレク
セルインターナショナル／コムチュア／キャノンIT ソリューション／データ・フォア
ビジョン／グリー／ラーニングエージェンシ／テンタムス・ジャパン／ADEKA／セ
ック／ワールドインテック／セガゲームス／ニトリ HD／貝印／アコーディア・ゴル
フ／データフォーションズ／コスモ・バイオ／リージョナルフィッシュ

※「株式会社」表記省略

生物情報科学科に
進学した
学生の声から

なぜこの学科を選んだか

3 年 A さん

Students' VOICE 001

私は大学入学前から生物に興味があり、進学振り分けでは生物系の学科に進学したいと思っていました。医療や生命工学などの応用的な利用法よりも、むしろ生物はどのような仕組みで生きているのか、というメカニズム的な分野に興味があったので、学部としては理学部を希望しました。同じ理学部にある生物学科や生物化学科ではなく生物情報科学科を選んだ一番大きな理由は、生物情報科学科は出来たばかりの新しい学科であったので、他の人とは違う新しいことをやりたいという自分の性格に最も合っているように感じたことです。また、学科の定員が少ない分密度の高い教育を受けられるという期待もありました。私はもともと高校時代にプログラミングを勉強したことがあり、その知識を活かすこともできると思ったことも理由の一つです。

4 年 B さん

Students' VOICE 002

「生物学はいつ終わるのか」私が気にしていたことはそれだった。当時私がよく聞いていた生物研究は一般性のない発見だったり、定性的な発見にとどまっていて、この研究スタイルが生物の全容解明に到底辿りつけるものとは思えなかった。氷山を崩したい。氷山の一角ではなく、氷山そのものを破壊したい。生命現象全てを正確に記述し、個々の現象の一般法則を導き出し、全体像を理解したい。それには膨大なデータとデータ処理が必要で、それは最早手作業では不可能、絶対に情報科学の力が必要だと思った。幸運にも時は十分に満ちていて、それができるのが生物情報科学科だった。これは入るしかないと考えた。

4 年 C さん

Students' VOICE 003

生物と言うと、高校や駒場のときの授業では、ただ生命現象に名前をつけて暗記するだけで、退屈なものだと思っていました。しかし、大量の実験データを元に生物現象を定量的に解析することで、ニュートンの法則のような基本原理を探ろうとしているという話を聞き、今まで持っていた生物に対するイメージとは全く異なったビジョンを示されて驚愕しました。わたしが学科を選んだときは、進路に関する情報もあまりなかったのですが、生物情報を志望することにしました。今は学生演習室も充実しているので、パンフには載っていないような話を是非聞きにきてください。

4 年 D さん

Students' VOICE 004

私がこの学科を志望した動機は一言で言うなら「何か新しく面白そうだったから」です。私はごく小さな頃は生物の図鑑を一日中飽きずに眺めたり、いろいろな昆虫を飼育したりしていましたが、中高と進学するにつれて生物学への興味は薄れて行きました。授業での生物学はただ生物にまつわる細々とした現象を暗記する科目にしか思えなかったのです。私は高校では物理と化学を選択し、理科一類に入学しました。しかし入学後に進学する学科を考えていた時、生物情報科学科の存在を知りました。調べるにつれて、この学科で勉強する生物学ならもっと面白いかもしれない、という期待を抱いて私はこの学科を志望しました。

(学年は当時のものです)





東京大学理学部 生物情報科学科

所在地 〒113-0032
文京区弥生2-11-16 理学部3号館

電話番号 03-5841-3045

FAX番号 03-5841-4400

ホームページ <http://www.bs.s.u-tokyo.ac.jp/bioinfo/>

お問い合わせ 生物情報科学科教務
jimu-biochem.s@gs.mail.u-tokyo.ac.jp

