東京大学理学部
生物情報科学科

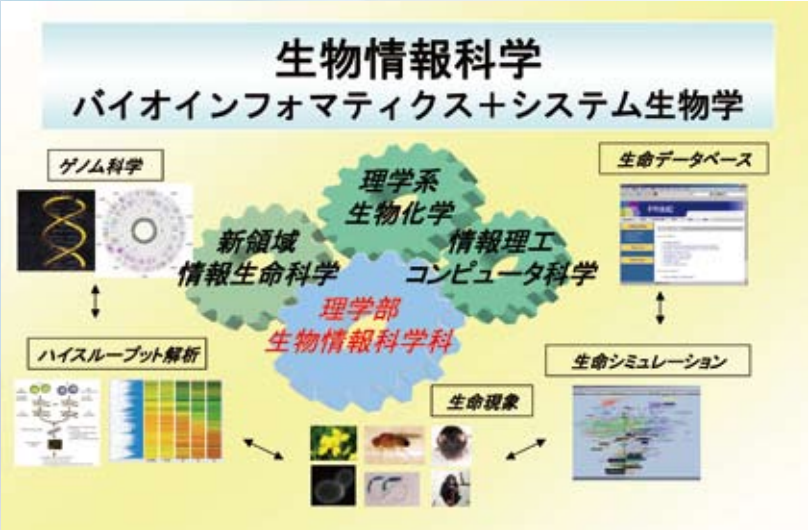
Department of Bioinformatics & Systems Biology
School of Science
The University of Tokyo

生物情報科学科について

生物情報科学科は平成 19 年度に新設され、
平成 20 年度から進学振分けを行い、平成 23 年3月には一期生が卒業しました。
生物情報科学科については設置後間もないため、
皆さんの手元までなかなか情報が入らないかもしれません。
そこで、生物情報科学科について概要をお伝えしたいと思います。

おそらく多くの人は生物情報科学という言葉さえ聞いたことがないでしょう。まずは、生物情報科学とは何かを最初に説明したいと思います。現在、生命科学はヒトを含む多くの生物でゲノムプロジェクトが終了したことを受け、その歴史上もっとも華々しい時代を迎えつつあります。生物情報科学は今後の生命科学の最先端を切り開く新しい学問領域として期待されています。生物情報科学とこれまでの生命科学との大きな違いは、個別の遺伝子やタンパク質を解析するだけでなく、それらの要素が相互作用することによって生命というシステムがどのように構築されているかを実験科学と情報科学の両方を駆使して解き明かす点にあります。生物情報科学は、大きく分けてバイオインフォマティクスとシステム生物学の 2 つの近接する分野からなっています（図）。

まずはバイオインフォマティクスについて説明しましょう。生命科学の分野では、ゲノム配列解析プロジェクトによって、2000 年前後に、ヒトを始めとする主要な生物種の全ゲノム配列が解明され、生命科学的な情報や知識が、近年著しく蓄積されてきました。一方、電子情報技術は、社会的にみても、インターネット、携帯電話などに見られるように人間社会のありかたを大きく変えるほどの速さで急激に進歩しています。バイオインフォマティクスとは、生命科学において新たに蓄積されてきた膨大なデータを、最新の情報科学的手法によって定量的に扱う学問分野です。次にシステム生物学について説明しましょう。システム生物学とは遺伝子やタンパク質の個別の要素である「部分」と生命現象のダイナミックな振る舞いである「全体」をシステムの視点から捉える学問分野です。生命現象を実験



により観測して、その背後に潜むロジックを数理的なモデルを用いて明らかにしていきます。生命科学を物理や工学のような視点から捉えて理解する感じに近いので、高校で数学や物理は学んだが生物を学んでこなかった人にもその考え方自体は馴染みやすいものと思います。これらの両分野の特徴は、実験（ウェット）と計算機（ドライ）の両方を密接にフィードバックしながら研究を行う点にあります。また、生物情報科学の分野では、生命科学と情報科学のどちらの手法も用いますが、単にそれらを足し合わせたものではなくて、どちらの基礎にも立脚して新しい視点から生命現象を理解する点がこれまでの生命科学と異なるユニークな点であり、10 年前には領域そのものが存在しなかった最先端の学問領域です。

生物情報科学の進学振分けとカリキュラムについて説明します。進学定員は 10 名です。平成 26 年度の進学振分けは、第一段階で 7 名（理科全類 6 名、全科類 1 名）、第二段階で 3 名（全科類）となっています。第二段階では生物化学と a.b 制を取っており幅広い選択が可能となっています。できるだけ偏りのないように理科全類、全科類となっており、主に理系を対象としていますがどこからでも進学することができます。また、カリキュラムについては、生物学、情報科学の基礎講義に加えて、新しい領域である生物情報科学（バイオインフォマティクス・システム生物学関連）の講義を中心に行います。実験についても、それぞれの基礎実験だけでなく、コンピュータプログラミングとゲノム実験を融合させた生物情報科学実験を行い、生物情報科学の実践力の養成も重視したカリキュラムになっています。生物学、情報科学の基礎的な実習については、生物化学科および情報科学科と合同で行ってきましたが、平成 24 年度からは生物情報独自の科目（生物情報実験法）を増やし、よりきめ細やかな対応ができるようになりました。理学部の多くの学科と同様に、3 年生の間は、午前中に講義が集中して行われ、午後はほとんど毎日実習です。4 年生になると、各研究室に配属されて、研究を行うことになります。生物情報科学科の具体的なカリキュラムについては、生物情報科学のホームページ（www.bs.s.u-tokyo.ac.jp/bioinfo/）に随時アップしていきます。

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るため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、生物情報科学の運営は、理学系研究科生物科学専攻、新領域創成科学研究科情報生命科学専攻および情報理工学系研究科コンピュータ科学専攻の複数の専攻にまたがる教員により行います。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）を受験することになります。このように複数の専攻の教員からなる教育組織は、従来の 1 専攻・1 学科から進化した新しい教育体制の試みのひとつです。また、本格的な生物情報科学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随することが多かったのですが、この生物情報科学の設置については世界に先駆けたものと言えるでしょう。独立法人化後、各大学の自助努力によりさまざまな取り組みができるようになりましたが、生物情報科学科の設立も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ます。また、理学部における新学科設立は 30 数年前の情報科学科以来です。一般に新しい大学院専攻は割りと頻繁に設置されたりしますが、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ることが多く、1 世紀に 1 度か 2 度あるくらいの頻度です。その意味でも生物情報科学の生命科学におけるインパクトがうかがえると思います。

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。生物情報科学科に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本当に真面目に期待しています。

生物情報科学科 カリキュラム

生物情報科学科では表のような授業科目が用意されています。平成 27 年度 3 年生進学者は、2 年生冬学期（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 10 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 38 単位、選択必修科目 1 単位、その他の専門科目 21 単位）を学修することになります。

なお、「情報科学実験」は情報科学科と合同で行い、「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。

また、平成 24 年度には、カリキュラムの大幅な改正を行い「生物情報実験法」を開設しました。生物情報に来て、初めてプログラミングをするという学生にも対応でき、かつ、上級者にも役立つ内容となっています。

時間割

1～2年・第1～3学期

日本バイオインフォマティクス学会認定技術者を目指せます。

時	午前		午後		
曜	1限	2限	3限	4限	5限
	駒場9：00～10：30				
水	総合科目一般「生物情報科学」				

2年・第4学期

□は、必修科目。

進学後に届出をし、認められた場合は、ここに記載された科目以外でも選択科目とすることができます。4 学期科目は、所定の単位を落とさず進学してきて下さい。

時	午前		午後		
曜	1限	2限	3限	4限	5限
	駒場9：00～10：30	駒場10：40～12：10	駒場13：00～14：30	駒場14：50～16：20	駒場16：30～18：00
月		生物情報学基礎論 I	計算機システム	遺伝学	生物情報学基礎論 II
火					
水	地形・地質学	物理実験学 細胞生理学		アルゴリズムとデータ構造	情報科学基礎実験
木		生物化学概論 I 物理数学 I ／物理数学 II	分子生物学		生物統計学
金	有機化学 I		情報数学 生物化学概論 II	形式言語理論 無機化学 I	情報科学基礎実験 化学熱力学 I

３年・夏学期

※集中講義　生物情報科学特別講義Ⅲ, Ⅳ。集中実習　地球惑星環境学野外調査Ⅰ。□は、選択必修科目。□は、必修科目。

月曜以外はウェット系の実験です。

時 曜	午前 1　限 8：40～10：10	午前 2　限 10：30～12：00	午後 3　限 13：00～14：30	午後 4　限 14：50～16：20	午後 5　限 16：40～18：10
月		ゲノム配列解析論Ⅰ ゲノム配列解析論Ⅱ	情報科学実験または生物情報実験法		
火	細胞分子生物学Ⅱ	離散数学 生物物理化学Ⅰ	生命科学基礎実験／生物化学実験		
水	細胞分子生物学Ⅰ	生体物質化学Ⅰ	生命科学基礎実験／生物化学実験		
木	バイオメトリックス	生物情報ソフトウェア論Ⅰ 生物情報ソフトウェア論Ⅱ	生命科学基礎実験／生物化学実験		
金		ゲノム生物学 システム生物学	生命科学基礎実験／生物化学実験		

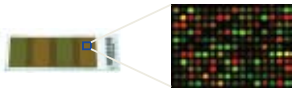
生命科学基礎実験 / 生物化学実験（３年夏学期 ◆火～木曜日３，４限 ◆金曜日３，４，５限）

前半　生物化学科担当（生物化学実験）

■基本実験操作　■蛋白質酵素の取り扱いの基礎　■遺伝子DNAの取り扱いの基礎 を身に付ける

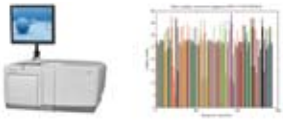
後半　生物情報科学科担当（生命科学基礎実験）

■マイクロアレイ実習
遺伝子発現プロファイル

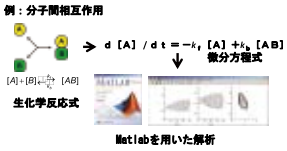


■トランスクリプトーム解析実習
次世代シークエンサーによる
遺伝子発現制御解析

■シークエンス実習
ホールゲノムショットガン
シークエンシング



■シミュレーション実習
シミュレーションモデル作成



３年・冬学期

□は、必修科目。

ドライ系の実験です。

月		細胞情報学	生物データマイニング論 生命情報表現論	連続系アルゴリズム	
火		生物物理化学Ⅱ 計算量理論	情報基礎実験		
水	海洋生物学	オーミクス論 環境ゲノム情報学	情報基礎実験		
木	分子生命科学Ⅲ	知能システム論 分子遺伝学	情報基礎実験		
金	生物物理学	生物統計論 生命情報科学Ⅰ	情報基礎実験		

４年・夏学期

□は、必修科目。

いよいよ卒業研究。まずは３研究室のローテーションから入ります。

月	生物情報科学特別演習・実験Ⅰ	情報論理 動物発生学Ⅰ	生物情報科学特別講義Ⅰ 生物情報科学特別講義Ⅱ		
火		生体数理モデル論 分散・並列システム論	情報科学とバイオインフォマティクス	理論生物学	生物情報科学特別演習・実験Ⅰ
水	生命分子化学		生物情報科学特別演習・実験Ⅰ		
木		コンピュータグラフィクス論 計算数理Ⅱ	生物情報科学特別演習・実験Ⅰ		
金		計算数理Ⅰ 統計力学Ⅰ	生物情報科学演習Ⅰ	生物情報科学特別演習・実験Ⅰ	

４年・冬学期

□は、必修科目。

卒業研究中心の生活です。

月					
火	生物情報科学		生物情報科学特別演習・実験Ⅱ		
水			生物情報科学特別演習・実験Ⅱ		
木		統計解析法	生物情報科学特別演習・実験Ⅱ		
金	生物情報科学演習Ⅱ	統計力学Ⅱ 細胞生理化学	生物情報科学特別演習・実験Ⅱ		

シラバス抜粋

■第１～３学期

生物情報科学（総合一般科目）

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。本講義は、バイオインフォマティクス技術者認定制度（Bioinformatics certificate）に準じたバイオ産業界において即戦力となりうる人材育成のカリキュラムである。

■第４学期

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、将来展望について述べるとともに、これを学ぶためには何を勉強すればいいか、どういう講義をとればいいかを紹介する。次に、生物情報科学のもっとも基礎であるゲノムの配列解析と、そこで使われる確率的な手法について、できるだけ分かりやすく解説する。最後に、データ解析のさまざまな場面で頻繁に使用されるデータ構造とアルゴリズムの基本的な考え方を紹介する。その際、一見関係ないと思われるOS やハードウェアの知識がどう活かされるかについても説明する。

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・システム生物学についての概論

■３年・夏学期

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・アルゴリズムを習得する。

システム生物学

細胞の増殖や分化、発生、神経活動、魚の縞模様などのパターン形成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、時間・空間パターン形成などの振る舞いを、具体的な生命現象に照らし合わせながら講義する。

生物情報ソフトウェア論 I

生命科学のデータ（ゲノム、ＲＮＡ，蛋白質等）は著しい速度で蓄積され、現在１日あたり約 500 億塩基対を解読できる超高速 DNA 解読装置が利用されている。このような大規模なデータの分析から、
・パーソナルヒトゲノムの多様性と疾患の関係、
・DNA のクロマチン構造、
・DNA 修飾と遺伝子発現と機能の関係
等について数多くの知見が得られつつあり、生命科学はデータ駆動型のサイエンスになりつつある。生命科学データの分析で使われるソフトウェアの中核は高速アルゴリズムにある。本講義はこれらの基本アルゴリズムが、どのような着想から設計されているかを解説する。そして将来自分でも設計できる力を身につけることを目標とする。

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身につける。C++ と Java におけるプログラムの書き方、デバッグ法にはじまり並列プログラミングの基礎までを演習形式で学ぶ。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。

情報科学実験

Linux の機能を利用したプログラミングを通じて現代の OS の利用方法を身につける。
さらに OS の基礎的な機能の実装を通じて OS の仕組みを学び、新たな OS の機能などを構築するための基礎力を獲得する。

生物化学実験

生化学、分子生物学および生物情報科学における基礎的実験。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

シークエンシング、質量分析、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・冬学期

ゲノム生物学

生物の構造と機能はゲノムによって規定されている。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、まず遺伝情報の総体としてのゲノム構造の一般的特性から始まり、遺伝情報の複製・発現などについて学ぶ。最近注目されているノンコーディングRNA についても概説する。さらに、ゲノム生物学における生物情報科学の発展においては、それらのゲノム情報をどのように解析し応用するかという点が重要である。そこで、このような視点から、ゲノム情報解析の意義と方法、今後の展開、さらには付随する問題点についての理解を深める内容を学ぶ。

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、多種多様で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。

生物統計論

DNA 配列 ,RNA 配列 , アミノ酸配列、マイクロアレイデータ、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。

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。言語は C++ あるいは Java を用いる。

■４年・夏学期

環境ゲノム情報学

ゲノム微生物学の基礎的な知識と考え方を習得し、微生物ゲノムの情報生命科学を理解する基盤を形成する。

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。（前半：選択した３研究室におけるローテーション。後半：配属研究室における演習。）

■４年・冬学期

生物情報科学　特別演習・実験 II

各研究室に配属され、指導教員の下で特定のテーマの課題に取り組む。研究テーマの設定、遂行、発表およびまとめの一連の過程を経験する。研究テーマの設定や遂行については指導教員からの指導を受けるが、自立的に研究を発展させる力を養う。

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。（各研究室でのセミナー。）

学生生活タイムテーブル

	４年Ａさん （卒研はドライ系@柏）	４年Ｂさん （卒研はウエット系@浅野）	４年Ｃさん （卒研はウエット系@柏）	３年Ｄさん
6：00	起床			
6：30	勉強会などでの発表準備			
8：00			起床、朝食とメールのチェックなど	
8：30		起床 朝シャワーを浴びて身だしなみを整えてお出かけ準備		
9：00	柏キャンパス に移動		家を出て柏キャンパスに向かう	
9：30				起床　寝起きは寒い
9：45		学校へ出発 自転車でキコキコ		
9：50				家を出る 自転車で登校
10：00	Bioinformatics 勉強会 個々人の得意分野を解説したり、生物の基礎知識を復習したり、論文紹介を行ったりしています	学校着 研究室に寄って荷物を置き講義へ		
10：15		講義		２限の授業に出る
10：30			研究室に到着、実験開始	
11：45		講義終わり		
12：00	昼食 普段昼食はカフェテリアで取っていますが、いいことがあるとキャンパス内のお寿司屋さんに行くこともあります	ランチセミナー 研究室のセミナー　ご飯食べながらジャーナル紹介を聞きます	研究室の方達と一緒にお昼ごはん（柏の生協食堂で日替わりランチなど）	昼食 生協の 280 円弁当を買ってくる たまに奮発して 380 円の弁当を買う　500 円弁当は買ったことがない
12：30				
13：00	データ解析	セミナー終わり なかなかの内容でござった	実験を続行	今日は学生演習室で講義がある 大画面モニターで授業が行われる
13：15		実験開始 真剣にやったり、時におしゃべりしながら　基本的に超忙しいけど、待ち時間に控え室に戻って一息入れたりもします		
14：00	卒研生がそれぞれ進捗状況をまとめてきて、先生や他の卒研生とディスカッションを行います　研究について説明し、フィードバックをもらうのは、とても貴重で重要な機会です			早めに講義が終わった　出された課題にすぐ取り掛かるのは気が引けるので、ニコニコ動画を見て遊ぶ
15：00			ラボ内ミーティング･論文輪講会	
16：00			実験の続きを行ったり、解析をするラボ内ミーティングや論文輪講会など	
16：45		バドミントンへ 実験をいい頃合いで切りをつけて、友だちと御殿下へ　バドミントンをします　運動は大事よね		
17：00	柏キャンパスから帰還		必要があれば教授や助教、ポスドクの方と研究について打ち合わせを行ったり 解析結果についてディスカッションを行う	そろそろ生産的なことをしないといけないなと思い、課題に取り掛かる
17：30				みんながゲームで遊びはじめた 課題はいつでもできるがこれは今しかできないので、自分も混ざる
18：00	夕飯			
19：00	データ解析	夕飯 バドミントンした仲間と近くの行きつけの中華料理屋でちょっとリッチなご飯　ヒロタケはどう食ってもうまいものだ　雑談やら研究の意見交換したりすることも		
19：30			ラボを出て家へ	
20：00		実験再開 今日できる実験は今日する！		夕飯を食べに中央食堂へ行く　今日はエコノミー定食の大盛り たまにはデラックス定食とか食べてみたい
20：30				
21：00			帰宅 & 晩ご飯	演習室に帰ってきて、やっとこさ課題にまともに取り組み始める　この時間に演習室にいるメンバーは大体決まっている
22：00			解析の続きなど	だんだん演習室にいる人々が帰っていく　さみしい
24：00	就寝	実験終了　帰宅 実験内容の解析とか、今後の計画とか、だらだらネットサーフィンしたり		４年の先輩がゲームで遊んでいるのを横目に、必死に課題のプログラムのバグ取りをする　つらい
25：00				下校、就寝
26：00			就寝	
27：00		就寝		

なぜ生物情報科学科を選んだか

3年Aさん

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を選んだ一番大きな理由は、生物情報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

「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思った。

4年Cさん

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、まだ卒業生もおらず、進路に関する情報もほとんどなかったのですが、生物情報を志望することにしました。今は学生演習室（通称 302）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来てください。

4年Dさん

高校の時の生物の先生に、生物系の研究がきっとこれからは面白いから是非大学に入ったらやってみると良い、という風に熱を込めて薦めていただいてから、なんとなく生物系の研究に興味を向けていました。大学に入って、研究室を訪問したり最先端の研究に触れる機会が時折あると、高校の時はただの暗記科目と思っていた生物の分野が、まだまだ未探索の地の多い宝島のように感じるようになりました。

生物情報科学科は、私が入学した頃にはまだ学科としてスタートしよう、としている段階でした。将来生物の研究をやる上で、今何を勉強するのがいいのか、非常に迷ったのを覚えています。生物情報は新しい学科（私たちが2期生）でとても活気があること、生物の知識を増やすのみではなく、研究のツールとして非常に有用になるだろうプログラミングの技術を学べることから、興味を持っていましたが、学科を決める段階ではまだよくわからない部分も多かったと思います。

実際に入ってみると、情報科学の課題など大変なことも多かったのですが、4年生になると、最先端の研究にそれらの授業内容が直結していて、大学入学の時に思っていたのよりずっと早く自分の夢でもあった生物の数理的な探求に従事できることが、この学科の大きなメリットの一つと感じるようになりました。

4年Eさん

私がこの学科を志望した動機は一言で言うなら「何か新しくて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。

生物情報科学の教員組織

理学系研究科　生物科学専攻

■ 黒田　真也 教授

skuroda@bs.s.u-tokyo.ac.jp
Tel:03-5841-4697

専門分野：システム生物学

「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」

■ 高木　利久 教授

tt@bs.s.u-tokyo.ac.jp
Tel:03-5841-8350

専門分野：バイオインフォマティクス

「複雑な生物情報を計算機で統一的かつ統合的に扱えるようにするための研究を行っています」

■ 程　久美子 准教授

ktei@bs.s.u-tokyo.ac.jp
Tel:03-5841-3043

専門分野：ゲノム情報生物学

「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」

■ 岩崎　渉 准教授

iwasaki@bs.s.u-tokyo.ac.jp
Tel:03-5841-3047

専門分野：バイオインフォマティクス

「生命現象を“俯瞰”する様々なデータを解析し、生命システムが従う法則性を見出します」

■ 柚木　克之 助教

yugi@bs.s.u-tokyo.ac.jp
Tel:03-5841-4699

専門分野：システム生物学

■ 西　賢二 助教

kennishi@bs.s.u-tokyo.ac.jp
Tel:03-5841-3044

専門分野：分子生物学

■ 日野　公洋 助教

khino@bs.s.u-tokyo.ac.jp
Tel:03-5841-3044

専門分野：RNA 分子生物学

新領域創成科学研究科　情報生命科学専攻

■ 浅井　潔 教授

asai@k.u-tokyo.ac.jp
Tel:04-7136-3986

専門分野：バイオインフォマティクス

「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」

■ 森下　真一 教授

moris@cb.k.u-tokyo.ac.jp
Tel:04-7136-3984

専門分野：バイオインフォマティクス

「DNA3 次元構造の謎の解明にとりくんでいます」

■ 服部　正平 教授

hattori@k.u-tokyo.ac.jp
Tel:04-7136-4070

専門分野：環境ゲノム科学

「ヒト・環境を取り巻く膨大未知な微生物ワールドをゲノムの視点から解いています」

※卒業研究の配属はできません。

■ 鈴木　穣 教授

ysuzuki@k.u-tokyo.ac.jp
Tel:04-7136-3607

専門分野：メディカルゲノム、トランスクリプトーム

「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」

■ 津田　宏治 教授

tsuda@k.u-tokyo.ac.jp
Tel:04-7136-3983

専門分野：バイオインフォマティクス

「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」

■ 木立　尚孝 准教授

kiryu-h@k.u-tokyo.ac.jp
Tel:04-7136-3972

専門分野：機能ゲノミクス解析

「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」

■ 笠原　雅弘 講師

mkasa@k.u-tokyo.ac.jp
Tel:04-7136-4110

専門分野：大規模バイオ情報解析

「ゲノム配列や遺伝子を中心とした大規模なオー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」

情報理工学系研究科・コンピュータ科学専攻

■ 玉田　嘉紀 助教

tamada@is.s.u-tokyo.ac.jp
Tel:03-5841-0517

専門分野：バイオインフォマティクス
大規模並列計算

「スーパーコンピューターを用いた並列アルゴリズムや膨大な計算により生命現象を解明する研究をしています」

※卒業研究の配属はできません。

理学系研究科・生物科学専攻

システム生物学

Shinya KURODA
黒田真也 教授
TEL: 03-5841-4697
skuroda@bs.s.u-tokyo.ac.jp
http://www.kurodalab.org

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（図１）。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。

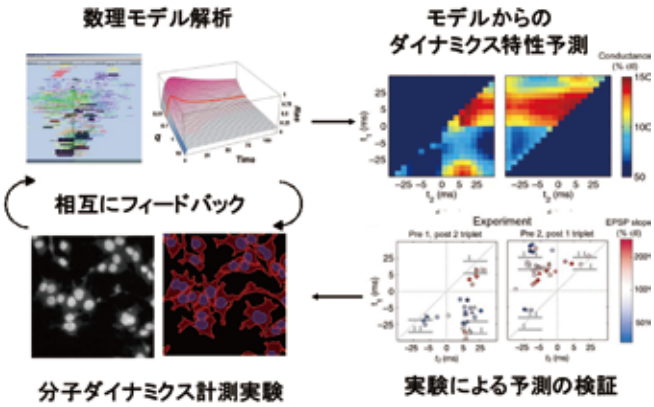


図１　システム生物学の解析の流れ

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２）。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。

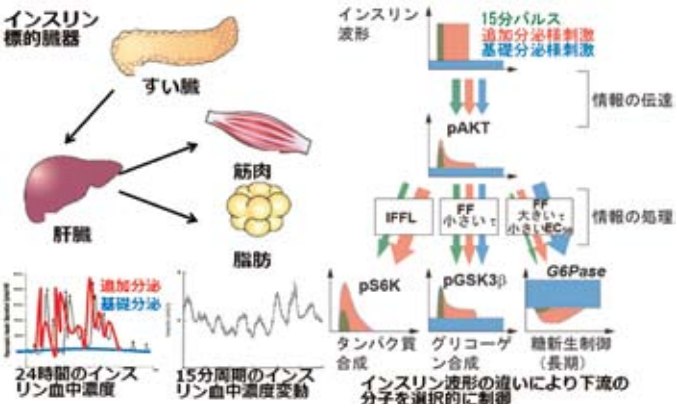


図２　インスリンの波形の違いにより異なる分子を制御

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３）。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ます。

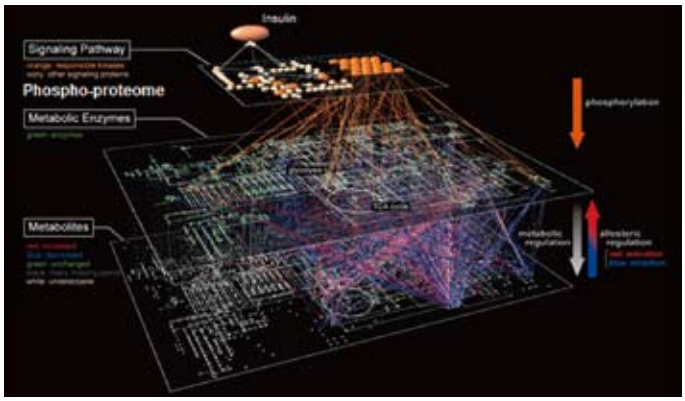


図３　インスリン作用のトランスオミクス経路

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。

理学系研究科・生物科学専攻

バイオインフォマティクス

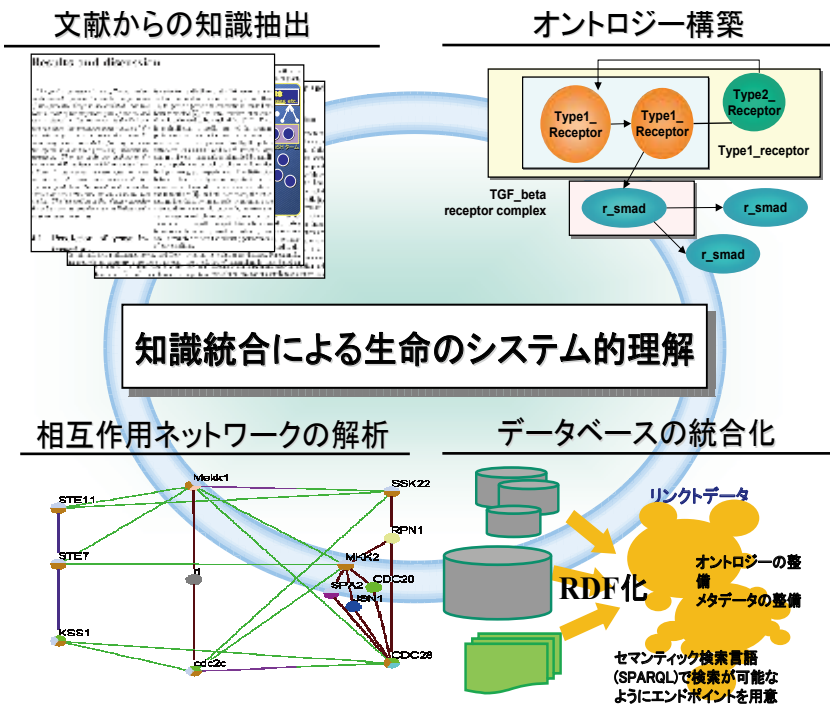
Toshihisa TAKAGI
高木利久 教授
TEL: 03-5841-8350
tt@bs.s.u-tokyo.ac.jp
http://www.cb.k.u-tokyo.ac.jp/takagilab/

知識処理技術を用いた生命システムの再構築

生命をシステムとして理解するためには、ゲノム配列やタンパク質立体構造だけでなく、発現、局在、相互作用、パスウェイ、ネットワーク、表現型などのさまざまな種類のデータ、および、それらの間の関係や生物学的な制約や文脈などに関する知識などを統合し解析すること、すなわち、計算機上に生命システムを再構築し、その性質、特徴、振る舞い、などを調べることが不可欠である。このような考えのもとに、我々の研究室では、次に掲げる研究テーマに取り組んでいる。

1. 文献からの知識抽出

パスウェイやネットワークなどの生命メカニズムに関する知識やそれの根拠となった実験事実の多くは、論文や教科書にテキスト（英文）や図表の形で書かれている。そこで、膨大な文献の中に埋もれたこれらの知識を自然言語処理や情報検索の技術を使い効率的に取り出す手法を開発している。より具体的には、タンパク質や遺伝子、化合物、疾患等の様々な医学生物学的概念に関する情報と概念間の関係性の自動抽出のための技術を開発している。また、抽出した知識を用いて実験データを解釈し、新たな知識発見を支援するシステムの開発を進めている。その他、複数の文献に書かれている内容を俯瞰するための手法、ポンチ絵や画像の検索システム、自然文での質問に回答するシステムの開発などをあわせて行っている。



2. 相互作用に基づく生命システムの解析

生命システムは生体分子間の相互作用を通して高度な機能を実現している。従って、複数の生物種における生体分子間の相互作用やネットワークを比較し、複数の生物に共通する、あるいは、ある生物特有の局所構造を切り出し、それらが実現している生体機能と対応づけることが重要である。我々の研究室ではこのような研究に必要な情報技術の開発およびシステム開発に取り組んでいる。

3. オントロジー構築

ゲノムの配列は ATGC の 4 種類のアルファベットの並びで簡単に表現できる。しかしながら、生体機能や生体メカニズムなどに関する複雑な知識をその本質を損なうことなく計算機上に表現することは容易ではない。また、複数の生物のシステムを比較解析するには、機能に関する表現などの統一を図る必要がある。生体機能や生体メカニズムに関する記述法の開発やその標準化のことをオントロジーと呼ぶが、我々の研究室では、パスウェイやネットワークなどを中心にオントロジー構築を進めている。

4. データベースの統合化技術

生命システムを理解するには、さまざまな種類のデータや知識を統合し、そこから新たな知識発見を行うことが必要である。そこで、DBCLS や NBDC などのデータベースセンターと連携して、RDF などの知識表現技術を用いて複雑な生物情報を統一的に表現する技術の開発を行っている。また、これらを用いて多種多様なデータベースの統合化に取り組んでいる。

5. 大規模データの表現手法の開発

個々の研究者がこれまで考えられなかった量のデータを扱うようになる中、大規模データを効果的に生物学的な解釈につなげるためのデータ表現手法の開発を行っている。例えば、オーミクスデータはしばしば巨大で複雑なネットワークとして、直感的な解釈が不可能な形で表現されるが、これを Google Maps のように動的に情報を抽象化、解釈可能な形で表現するネットワーク解析法を開発している。

理学系研究科・生物科学専攻

ゲノム情報生物学

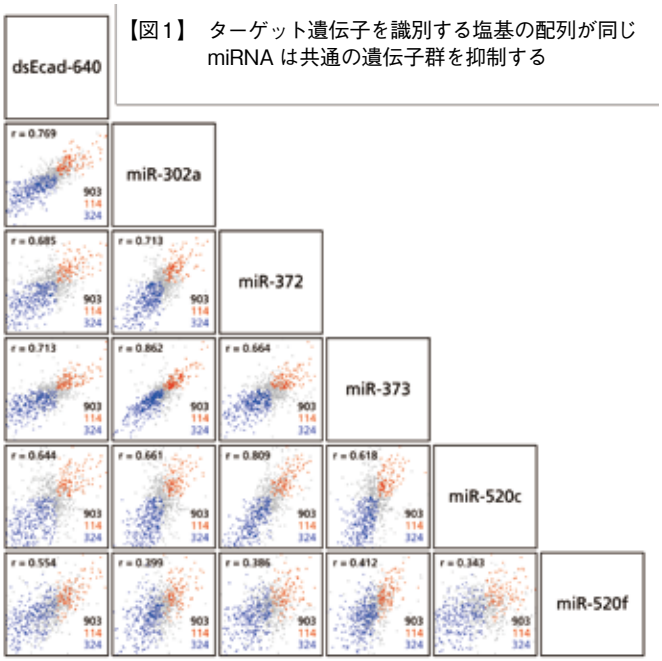
Kumiko UI-TEI
程 久美子 准教授
TEL: 03-5841-3043
ktei@bs.s.u-tokyo.ac.jp
http://ui-tei.RNAi.jp/

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デオキシリボ核酸（DNA）からリボ核酸（RNA）が転写され、RNA からタンパク質が翻訳されることによって、ゲノムの遺伝情報に基づいた機能が発揮されます。このような分子生物学の中心原理はセントラルドグマと呼ばれます。長い間、RNA は DNA からタンパク質へと遺伝情報を伝達するための仲介分子であると認識されていました。しかし、近年、RNA に関する分子生物学の研究が急速に進展して、RNA は単なる遺伝情報の仲介役ではなく、自身にも様々な機能がある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、これまでの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ます。私たちは、ノンコーディング RNA を対象とした新しい RNA 研究を行っています。

ゲノムワイドな RNA サイレンシング機構の解明

ノンコーディング RNA の中には約 20 塩基ほどの小さな 2 本鎖 RNA から数 10 キロ塩基という長い RNA まで様々なものがあり、それぞれ多様な生命現象に関わっていることが明らかになってきています。microRNA や small interfering RNA（siRNA）という約 20 塩基の小分子 RNA による RNA サイレンシングでは、小分子 RNA 全長の約 3 分の 1 の領域がターゲット遺伝子を配列依存的に識別し、相補的な配列をもつ数 100 ～ 1,000 種のメッセンジャー RNA の発現を一挙に抑制します。このような遺伝子全体の挙動は、ゲノムワイドな実験から明らかにすることができます。たとえば microRNA には癌由来細胞を分化多能性をもつ細胞へと分化誘導できるものがありますが、これらは同じターゲット識別部位をもつものが複数存在し、共通の遺伝子群の発現をコントロールしています（図 1）。

さらに、私たちはターゲットを識別する配列が異なる microRNA は、その配列によって規定される熱力学的安定性に依存してサイレンシングの程度が異なることを明らかにしました。このような分子生物学的および物理化学的性質から、ゲノムワイドな遺伝子発現制御機構が推定できることを期待しています。



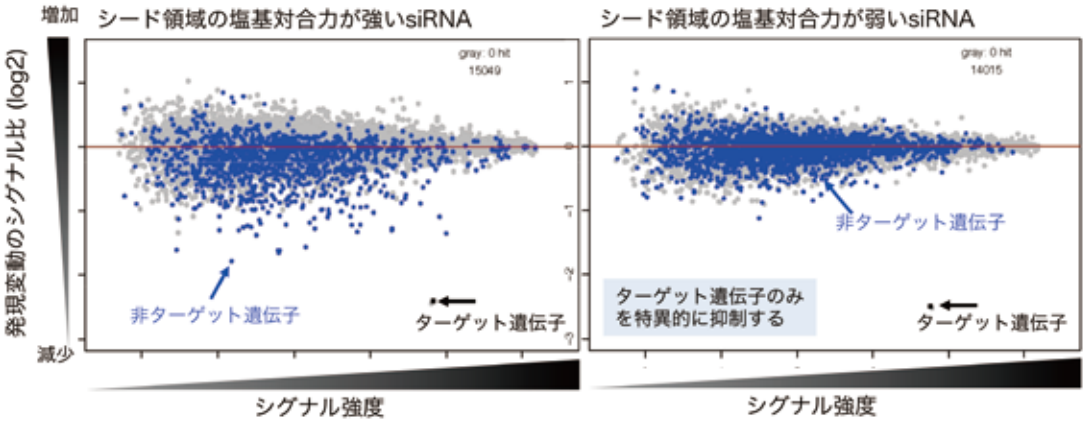
【図1】 ターゲット遺伝子を識別する塩基の配列が同じ miRNA は共通の遺伝子群を抑制する

RNA 結合タンパク質の機能解析

RNA による遺伝子発現制御機構には RNA 結合タンパク質が重要な役割を担うことがわかってきました。RNA 結合タンパク質は、RNA サイレンシングの過程やターゲット遺伝子識別機構、RNA の編集機構、さらには RNA ウイルスによる免疫応答反応などの様々な生体機能に関わっている重要なタンパク質です。それらの機能を明らかにするとともに、大規模シークエンスによる網羅的な結合領域の特定を行い、遺伝子の発現をどのように制御しているのかを解析しています。

RNA 干渉法を利用した遺伝子ネットワークの解析

siRNA による RNA サイレンシングは RNA 干渉と呼ばれます。RNA 干渉法は現代の分子生物学では欠かせないツールになってきています。また、ゲノム遺伝子には影響を与えないため、その臨床応用にも大きな期待がかかっています。しかしながら、RNA 干渉では、ターゲット遺伝子を抑制するだけではなく、部分的に相補的な配列をもつ多くの非ターゲット遺伝子も非特異的に抑制する作用があります。私たちは RNA 干渉の機構を明らかにすることにより、目的とする 1 つの遺伝子のみを特異的に抑制する方法を構築しています（図 2）。この方法は遺伝子機能解析法、およびウイルス感染防止法などの応用研究にも利用されています。それだけではなく、本手法は、従来のトランスクリプトーム解析の手法のみでは困難とされていた遺伝子ネットワーク解析法として利用可能であり、その全体像を明らかにすることを目指しています。



【図2】 標的遺伝子特異的 RNA 干渉法の構築とマイクロアレイによる確認

理学系研究科・生物科学専攻

バイオインフォマティクス

Wataru IWASAKI
岩崎　渉 准教授
TEL: 03-5841-3047
iwasaki@bs.s.u-tokyo.ac.jp
http://iwasakilab.bi.s.u-tokyo.ac.jp

生命システムの設計メカニズムをひもとく

ライフサイエンス分野における技術革新は、これまでは得ることすら想像できなかった規模のデータを俯瞰的に解析し、生命システムの性質やその進化過程に迫ることを可能にしました。

数億年スケールの遥かな時間を越え、生命の過去の歴史を現在に伝える「ゲノムデータ」。遺伝子発現、調節、そして相互作用など、ゲノム配列が機能するメカニズムを語る種々の「オーミクスデータ」。生物学者がこれまでに蓄えてきた膨大な知識を、コンピュータが理解できる形式で表現した「パスウェイ・ネットワークデータ」。生物の表現型、行動や社会性を定量的に解析することを可能にする「生物画像・動画データ」。そして、環境と生命の複雑な相互作用ネットワークを描き出す「メタゲノム・生態系ゲノムデータ」。

生命をシステムとして理解するためには、生命の限られた側面に注目するだけではなく、これらのデータを自在に解析することが不可欠です。私たちの目標は、生物学と情報学の双方の理解にもとづいて、複雑な生命システムの成り立ちに関する新しい概念、仮説、法則性をこれらのデータから見いだしていくことです。

ゲノムと生命システムの進化

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在に至る生命がどのように地球上で進化してきたか、ゲノムやそこにコードされた生命システム・代謝ネットワークはどのように進化してきたか、また、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。

さらに、トランスクリプトーム情報にはゲノム中で機能している遺伝子全体についての、メタゲノム情報には環境微生物の生態系についての、それぞれ豊富な知識が埋もれています。環境の変化に応じた様々なオーミクスデータの変動を解析することで、生命が環境の変化にどのように応答するか、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、などを探求しています。

生物行動・表現型の動画解析

生物を撮影した画像・動画データから、表現型や行動といったこれまで定量的に解析することが難しかった性質を定量的に解析することができるようになりました。この研究分野をバイオイメージインフォマティクスと呼び、近年、注目が高まっている分野です。

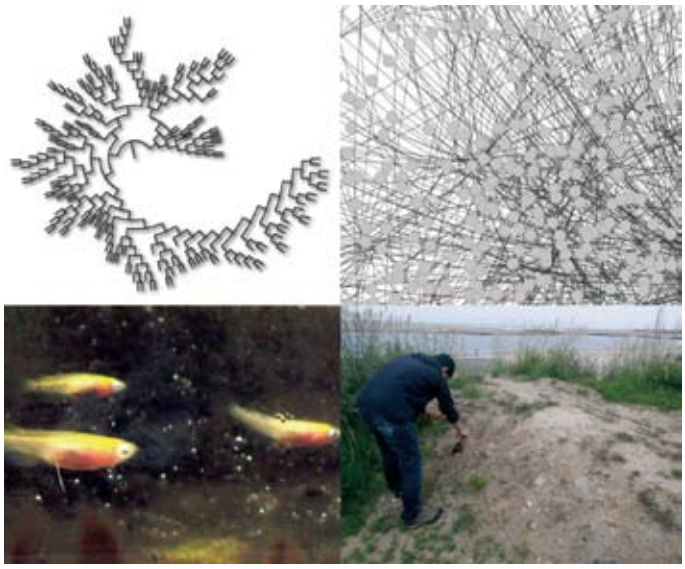


図. 生命システムをひもとく（左上）ゲノム進化解析（右上）遺伝子ネットワーク解析（左下）メダカ行動解析（右下）メタゲノム解析

私たちは、動画データから生物の行動を解析するための新しい情報技術を開発するとともに、生物行動に隠された法則性を見いだすための研究をすすめています。

またそのほかに、データビジュアライゼーション、テキストマイニング・書誌情報解析に関する研究を行っています。

生命を情報システムとして理解する

生命を俯瞰するデータを自在に解析し、新しい概念や仮説へ結びつけていくためには、生物学と情報学の双方に対する理解と興味が必要となります。私たちの研究室では、生物学と情報学にまたがった教育・研究を特に重視しています。

様々なデータから生命システムに関する新しい概念や仮説を導いていくための解析手法は、まだまだ研究が始まったばかりの段階です。そこで、必要な数理的手法や情報科学的手法を自ら見いだし開発することが、新しい問題を解くための重要な手段となります。その一方で、数理や情報に関するバックグラウンドが無くとも、基本的なプログラムの書き方をマスターする熱意があれば、自由な生物学的発想に基づいて網羅的データを縦横に解析し、新しい発見を行うことも可能です。また希望される方には、コンピュータを用いた解析に加えて遺伝子解析を中心とした生物学実験にも取り組んでもらうほか、実験系研究者とアイデアの交換や共同研究を行い、融合領域から面白い研究を生み出すことを目指しています。

生物学と情報学の双方を深く学ぶとともに、独自の発想に基づいた研究に挑戦する熱意ある方を歓迎します。研究室訪問を希望する方はまずはメールにてご連絡ください（進学前でも歓迎します）。

新領域創成科学研究科・情報生命科学専攻

バイオインフォマティクス

Kiyoshi ASAI
浅井　潔 教授
TEL: 04-7136-3986
asai@k.u-tokyo.ac.jp
http://www.cb.k.u-tokyo.ac.jp/asailab/

当研究室では、確率的な枠組みを中心とした数理的な理論とその応用によって、生命科学、特にゲノムに関連した生命現象の解明を目標に研究している。隠れマルコフモデル（HMM）、確率文脈自由文法（SCFG）などの確率文法、サポートベクターマシン（SVM）などのカーネル法を活用するとともに、動的計画法（DP）に関する新しいアルゴリズムを構築し、基本的なソフトウェアの開発に取り組んでいる。生物学に役立つ計算機利用技術を開拓するだけでなく、情報学の立場から生命現象をどのようにモデル化・抽象化するかを重視している。研究・教育においては個人の自主性を重んじ、研究室外の研究者との議論、共同研究を奨励している。

ゲノム情報解析と高速シークエンサー

ゲノム配列は、生命の設計図であるとともに動作プログラムでもある。近年の研究で、ゲノムが動的に制御され、発生・細胞分化におけるゲノム構造の変化にも注目が集まっている。他方、ゲノム配列は生物情報の中で最も精度の高い（デジタル）データであるという点が重要である。浅井らは、麹菌（A. oryzae）ゲノムプロジェクトにおける情報解析および近縁ゲノム（A. nidulans, A. fumigatus）との比較ゲノム解析に参加した。岡田（2008 博士卒）らは、比較ゲノム解析によって脊椎動物の過去に起こった全ゲノム重複と、神経系・獲得免疫系・内分泌系の遺伝子群の関係を明らかにした。

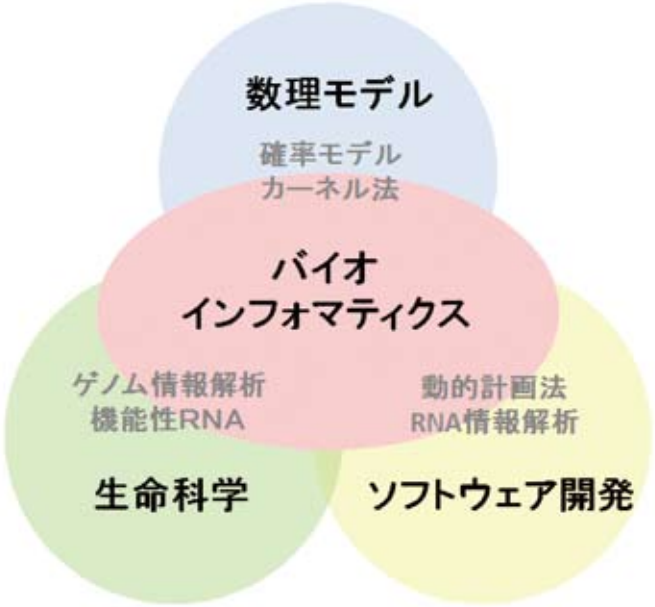
近年、シークエンシング技術が劇的に進歩し、DNA 配列のデータ量が爆発的に増加している。この状況に対応するため、シークエンサーから得られる膨大な配列データを高速に解析する技術の開発が重要となっている。そのため、情報生命科学専攻を中心として 2009 年度から開始されたグローバル COE「ゲノム情報ビッグバンから読み解く生命圏」（拠点リーダ：森下真一教授）が開始された。当研究室では、膨大な配列データを網羅的に解析して意味のある情報を抽出するための研究を中心に取り組んでいる。

情報解析と確率モデル

DNA、RNA、タンパク質アミノ酸配列などの生物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る。配列データとして情報解析する場合には、文字列しか「見る」ことができないが、その背後に、「意味」が隠されているのである。タンパク質アミノ酸配列では、部分文字列に対してタンパク質 2 次構造という物理的な実体が対応しているし、ゲノム DNA では、部分文字列に対してタンパク質コード領域（遺伝子）という生命現象の機能単位が対応している。このような状況を自然に説明できる確率モデルとして、浅井らは 1990 年代の初頭から隠れマルコフモデル（HMM）に注目し、タンパク質 2 次構造予測や遺伝子発見に用いてきた。また、RNA 配列情報解析のための確率文脈自由文法（SCFG）上の周辺化カーネルなどの新しい概念を提唱した。また、バイオインフォマティクスにおける高次離散空間上での予測問題について、予測精度を最大化するための理論的な枠組みについて研究し、γ セントロイドという新しい推定方法を提唱した。

RNA のバイオインフォマティクス

ごく最近まで、mRNA として転写されてタンパク質に翻訳されるタンパク質コード遺伝子と、その転写制御を行う制御配列のみが、ゲノム配列中で重要な領域であると考えられていた。しかし、RNA 干渉や miRNA の発見以降、転写されるが翻訳されず、かつ機能を持つ非コード RNA が翻訳制御その他の生命現象に重要な役割を担っていることが明らかとなった。浅井らは、NEDO の「機能性 RNA プロジェクト」（2005 年度～ 2009 年度）で RNA のバイオインフォマティクスに取り組んだ。田部井（2009 博士卒）らが、2 次構造を考慮した RNA 配列の高速アラインメント手法を開発した他、γ セントロイドによる世界最高精度の 2 次構造予測手法など、RNA 配列情報解析技術で世界をリードする理論・ソフトウェアを開発している。プロジェクトで開発されたソフトウェア群および機能性 RNA データベース、ブラウザは、http://www.ncrna.org で公開されている。



研究プロジェクト

グローバル COE「ゲノム情報ビッグバンから読み解く生命圏」
新学術領域研究「生命科学系 3 分野支援活動」（情報解析支援）

生命情報工学研究センターとの連携

産総研生命情報工学研究センター（CBRC）と連携し、共同研究拠点を産総研臨海副都心センター（お台場地区）に設けており、学生の多くはそこで研究活動を行っている。

新領域創成科学研究科・情報生命科学専攻

バイオインフォマティクス

Shinichi MORISHITA
森 下 真 一 教授
TEL: 04-7136-3984
moris@cb.k.u-tokyo.ac.jp
http://mlab.cb.k.u-tokyo.ac.jp/

生命の設計図ＤＮＡで変化する部分、不変な部分

ＤＮＡ（ゲノム）には幅広いスケールの変化が起こることが知られています。

- 数億年にわたる染色体の大規模な再編成
- 種として分かれた後の比較的小規模な変化（遺伝的多様性）
- 発生、分化の過程でおこる DNA 修飾等

これらの変化が、遺伝子の機能を豊かにし、多様な生物を生み、さまざまな疾患を引き起こします。そこで

「ＤＮＡにはどのような変化が起きるか？」

「不変量および不変な性質はなにか？」

を主題に膨大なＤＮＡデータおよび画像データを情報分析してきました。同時に高速に分析するためのアルゴリズムも研究してきています。

染色体の大規模な変化

過去 6 ～ 10 億年、脊椎動物や昆虫の染色体がどのように進化してきたかを分析し、とくに 1970 年に Ohno が提唱した「脊椎動物初期に DNA 全体が２回重複し遺伝子セットを豊かにした」という仮説を、ヒト、チキン、メダカ、フグのＤＮＡを比較して解明しました（図１）。

個人ＤＮＡ解読と疾患関連遺伝子の探索

数百万年から数十年のスケールでは、置換、挿入、削除等の比較的小さな変化が生まれ、異なる表現型、たとえば遺伝病に関与しています。この変化をとらえるには個人ＤＮＡを解読し比較します。約３０億塩基対の２倍体ＤＮＡを読むのは大変なことです。私たちは超高速解読装置（Illumina, PacBio）および超並列計算機を使って、個人ＤＮＡに起こる変化を１週間程度で分析できるシステムを作成しています。これをもとに、東大病院と共同で日本人固有の遺伝的多様性を分析しています。この結果、「日本人の標準的ＤＮＡ像」が明らかになりつつあります。また、遺伝的要因の強いことが知られている脳疾患を対象に、関連する遺伝子の変化を探索しています。

クロマチン構造、ＤＮＡメチル化、遺伝的多様性

伸ばすと約２m のヒトＤＮＡはヒストン８量体に巻きつきクロマチン構造を形成し、直径 10 μ m ほどの小さな核の中に折り畳まれています。ＤＮＡがコピーされ２つの核に分配される際に長いひものＤＮＡはなぜ絡まらないのか？　遠く離れた位置にコードされる遺伝子が協調して働くとき、折畳みにより近い位置に配置されるのか？　などの疑問にアプローチしています。クロマチン構造に関して報告した面白い現象は、転写開始点下流において、遺伝的多様性には周期性があり、クロマチン構造と相関する結果です（図２）。また脊椎動物では CpG の C がメチル化されると T へ変異しやすいですが、その周辺の塩基も変異する傾向も報告しました（図３）。ＤＮＡ変化を導く背後のメカニズムには謎が多いです。

遺伝子破壊株のイメージ分析

DNA 配列がわかると、遺伝子をノックアウトもしくは強制的に発現するように DNA を改変することが可能です。必須遺伝子の破壊は致死的ですが、非必須遺伝子の場合は表現型が微妙に変化します。そこで出芽酵母のすべての非必須遺伝子を破壊した株の画像を撮影し、わずかな変化を捉え、機能推定に活かしています（図４）。

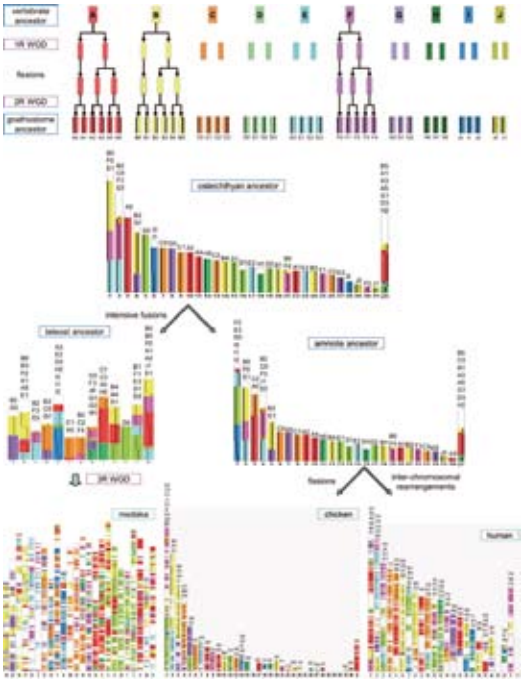


図１　脊椎動物の染色体進化
（出典　Nature, 447:714-719, 2007, Genome Research, 17（9）:1254-1265, 2007）

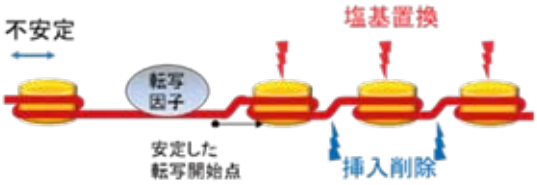


図２　クロマチン構造とＤＮＡ変化は相関する
（出典　Science, 323（5912）:401-404, 2009）

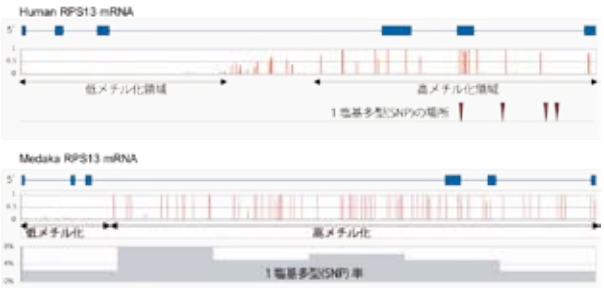


図３　低メチル化領域周辺ではＤＮＡ変異が少ない
（出典　Genome Research, 22（8）:1419-1425, 2012）

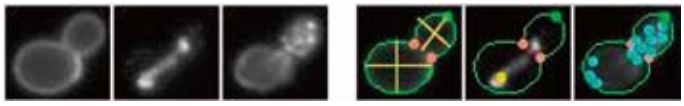


図４　出芽酵母遺伝子破壊株（左）と画像分析（右）
（出典　PNAS, 102（52）:19015-20, 2005）

新領域創成科学研究科・情報生命科学専攻

環境ゲノム科学

Masahira HATTORI
服部正平 教授
TEL: 04-7136-4070
hattori@k.u-tokyo.ac.jp
http://www.cb.k.u-tokyo.ac.jp/hattorilab/

私たちの研究室は、大量ゲノムシークエンス（DNA 塩基配列）解析やバイオインフォマティクス技術などのオーミクス研究を基盤にして、ヒト常在菌を含めた自然環境中に棲息する細菌集団（叢）の生命システムの包括的解明を行っている。

元来、細菌は数種～数万種の細菌種からなる細菌叢を形成して自然環境中やヒト体内に棲息している。そこには、細菌-細菌間や細菌- 環境／宿主間の相互作用などの複雑な生命システム（細菌が生きるための総合的なしくみ）が存在する。しかし、細菌叢を構成する細菌種の 99%以上は実験室で分離培養できず、その全体像は明らかでない。メタゲノム解析は各構成細菌種を分離せず、細菌叢をそのままゲノム解析する方法をいう。メタゲノム解析によって、培養、難培養性にかかわらず、そこに存在する細菌の遺伝子情報（ゲノム配列）を大量に獲得することができ、これまで解析不可能であった細菌種、新規な機能をもった遺伝子や代謝物の発見が見込まれ、細菌叢の生命システムや環境生態系の包括的な解明が期待できる。

ヒト腸内細菌叢のメタゲノム解析

私たちは、これまでにヒトゲノムやヒト病原菌、産業有用菌、昆虫の共生細菌等のさまざまな生物種のゲノム解析研究を行ってきた。最近では、このゲノム解析技術を基盤にして、より複雑で未開分野であるヒト常在細菌叢ゲノム（ヒトマイクロバイオーム）の研究を展開している。

ヒトの腸内、口腔、皮膚等に常在する細菌叢のうち、ヒト腸内細菌叢は約 1000 種の細菌から構成され、ヒトマイクロバイオームの中で圧倒的多数を占めている。ヒト一人の腸内には、総数で数百兆個（重さにして 1kg）の腸内細菌（ビフィズス菌や大腸菌など）が棲み、これら細菌がもつ遺伝子の総数はヒトの 100 倍程度にもなる。そして、善玉菌や悪玉菌として知られているように、食事成分からのビタミンやアミノ酸などの必須栄養素の代謝生産能力、腸管上皮細胞の増殖や腸管免疫系の成熟化、病原菌の感染に対するバリアー、一方で、そのアンバランスは炎症性腸疾患、アレルギー、肥満、大腸がんなどの疾病と関係しており、ヒトの健康と病気に密接に関係している。つまり、ヒトはヒトゲノムとヒトマイクロバイオームからなる超生命体（superorganism）とみなすことができる。しかしながら､腸内細菌叢がもつ（ヒト細胞との相互作用も含めた）機能、細菌叢の形成機構、菌種組成を決定する因子等の分子機構の詳細はわかっていない。現在、大人や子供、乳児の健康な腸内細菌叢および炎症性腸疾患などのさまざまな病態腸内細菌叢のメタゲノム解析、またそれらの比較解析を進めている。これらの解析を通して、細菌叢の全体像把握、健康と病気に関わる細菌種、遺伝子、代謝物等の同定、宿主側遺伝子の応答機構、細菌遺伝子と宿主遺伝子の間にある相互作用の分子機構等の解明をめざしている。

上記の生物学的テーマに加えて、メタゲノム解析で得られる大量のゲノム情報の解析のための、新しいバイオインフォマティクス技術の開発も欠かせない研究テーマの一つである。

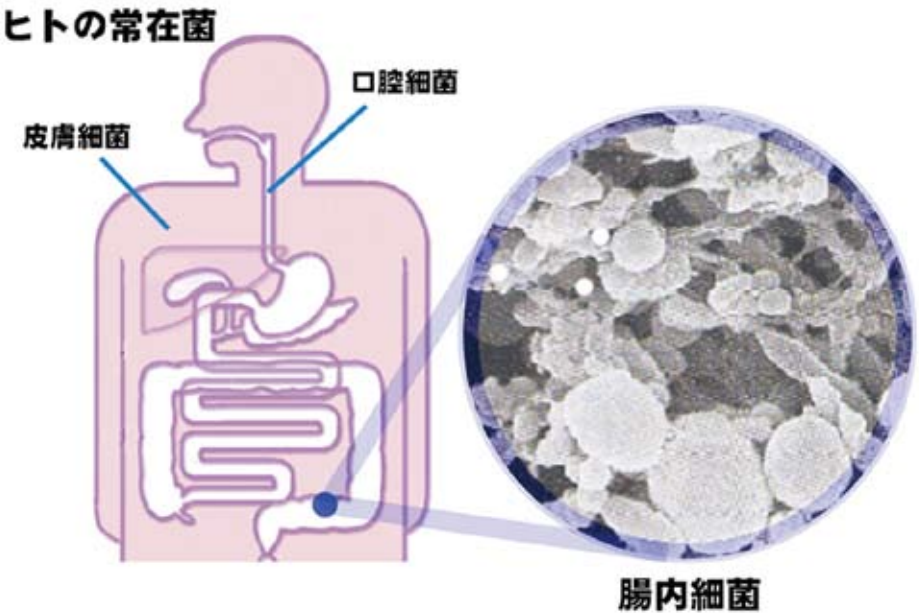


図 . ヒト腸内細菌叢

新領域創成科学研究科・情報生命科学専攻

メディカルゲノム　トランスクリプトーム

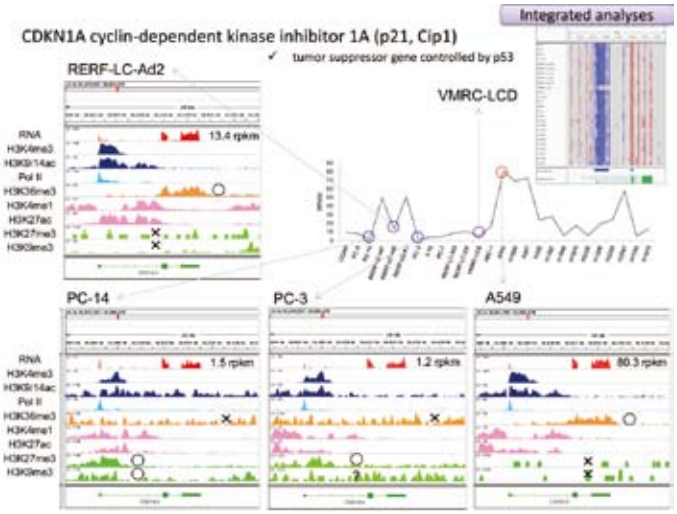
Yutaka SUZUKI
鈴木　穣 教授
TEL: 04-7136-3607
ysuzuki@k.u-tokyo.ac.jp
http://www.cb.k.u-tokyo.ac.jp/suzukilab/

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。

近年の次世代シークエ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、圧倒的なシークエンス産生能を背景に、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途について、多階層から得られたオー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、次世代シークエンス技術、シングルセル解析技術、1分子シークエ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、国内有数のシークエンス拠点として全国の研究者に提供している。

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 配列を飽和規模で配列 - 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持ち、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。同一環境の中で相互に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

次世代シーエクエンス目的別アドバンストリファレンス：
秀潤社 .：菅野純夫、鈴木穣

新領域創成科学研究科・情報生命科学専攻

バイオインフォマティクス

Koji TSUDA
津田宏治 教授
TEL: 04-7136-3983
tsuda@k.u-tokyo.ac.jp
http://www.cbrc.jp/~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。

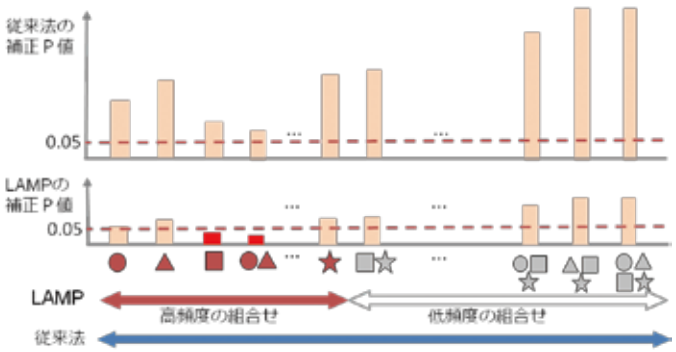


図１　LAMP による多重検定

高機能暗号によるプライバシ保護マイニング

本研究室は、JST CREST プロジェクト「自己情報コントロール機構を持つプライバシ保護データ収集・解析基盤の構築と個別化医療・ゲノム疫学への展開」に参加しており、ゲノムデータに関するプライバシ保護マイニング技術の研究を行っている。このプロジェクトにおいては、ゲノムデータを扱う際に個人情報が漏えいすることを防ぐため、準同型暗号、属性ベース暗号等の高機能暗号を用いて、暗号化したデータを一度も復号しないまま処理する秘密計算技術を開発する。特に、GWAS データを秘匿したまま、形質に関連した多型を特定する方法や、秘密計算による配列アラインメントなどの課題に取り組んでいる。

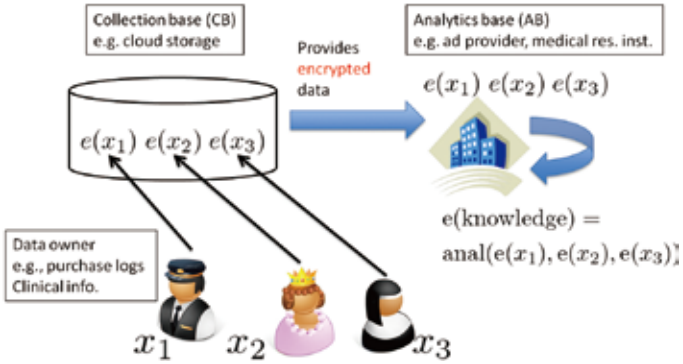


図２　秘密計算によるプライバシ保護データマイニング

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 [2]。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 [3]。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。

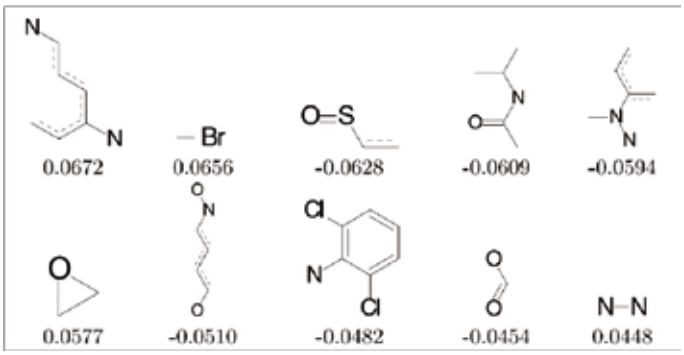


図３　GBOOST[2] によって発見された部分グラフ特徴

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.

[2] 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, 2009.

[3] K. Shimizu and K. Tsuda. SlideSort: All Pairs Similarity Search for Short Reads. Bioinformatics, 27(4):464-470, 2011.

新領域創成科学研究科・情報生命科学専攻

機能ゲノミクス解析

Hisanori KIRYU
木立尚孝 准教授
TEL: 04-7136-3972
kiryu-h@k.u-tokyo.ac.jp
http://www.cb.k.u-tokyo.ac.jp/kiryulab/

木立研究室では、ゲノム配列、マイクロアレイ発現データ、次世代シーケンサーデータなど、ゲノムスケールの大規模生物情報から、統計的手法により、新しい生物学的事実を発見する研究を行なっています。また、そのために必要となる確率的、数学的手法を開発すること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、ゲノム配列のみにとどまらず、RNA、タンパク質、代謝物質、DNA の修飾状態など、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、これらの大規模データの本格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、ぱっとみただけでは分からないようなデータの微妙な特徴を、複雑な統計計算であぶりだし、データに潜む、生物学的に深い真実を発見することを目指しています。

RNA 構造の機能と進化

ゲノムにコードされているタンパク質遺伝子の発現には、メッセンジャー RNA、転移 RNA、マイクロ RNA など、様々な RNA 分子が関わっています。多くの RNA 分子は、A-U、C-G、G-U などの塩基ペアが水素結合した、二次構造を形成します。この二次構造形成のエネルギーは比較的大きく、遺伝子発現の制御や効率に大きな影響を与えています。RNA 二次構造については、確率文脈自由文法とよばれる情報科学的概念に基づく非常に精密なモデルがあり、これを用いて、RNA 構造のいろいろな性質を計算機上で調べることができます。私たちはこのモデルを駆使して、マイクロ RNA・RNA 結合タンパク質の結合、選択的スプライシング、メッセンジャー RNA の翻訳など､RNA が関わるさまざまな生物過程について、構造科学的な観点から研究をしています（図１）。また、脊椎動物やヒトの集団ゲノムデータを用いて、RNA の構造進化に関する研究を行なっています（図２）。

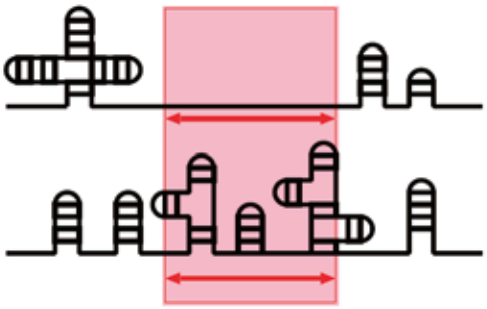


図１　RNA のアクセシビリティを計算するツール Raccess によるゲノムスケールの配列解析。Raccess は RNA 転写物の全ての領域について、その領域が二次構造を形成しない露出した領域であるかどうかを確率値として与えます。

がんゲノムの進化

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過程と似ており、進化学や遺伝学の手法を用いて、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マルコフ過程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。

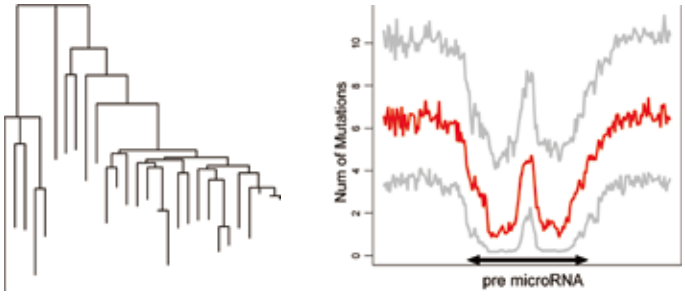


図２　左：開発した系統樹プログラム Fdur を用いて計算されたヒトを含む２８種の脊椎動物の局所進化系統樹。
右:Fdur で計算したマイクロ RNA 領域の塩基置換パターン。

胚発生と細胞分化のシミュレーション

動物の胚発生では、受精卵の卵割に始まり、原腸形成、胚葉分化などを経て、臓器、骨、筋肉などが形成されます。このような動物のマクロスコピックな変化は、細胞内の転写ネットワークと、モルフォゲンなどを介した生体環境との相互作用により、精密に制御されています。しかし、このような制御メカニズムを、転写因子結合・ヒストン修飾などのシーケンシングデータから推定する技術は、まだまだ未熟な状態です。　私たちは、数理生物学における、微分方程式を用いた胚発生モデルと、バイオインフォマティクスにおける、ベイズ推定に基づく遺伝子ネットワーク解析手法とを融合することにより、マクロスコピックな胚発生過程とミクロスコピックなシーケンシングデータとを関連付ける手法の開発を進めています。これを用いて、シーケンシングデータから、動物の発生過程の推定やシミュレーションを可能にすることを目指しています。

共同研究、連携先

産業技術総合研究所・生命情報工学研究センターのグループと緊密に連絡をとり共同研究を行っています。

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。

新領域創成科学研究科・情報生命科学専攻

大規模バイオ情報解析

Masahiro KASAHARA
笠原雅弘 講師
TEL: 04-7136-4110
mkasa@k.u-tokyo.ac.jp
http://ka.cb.k.u-tokyo.ac.jp/

当研究室では、第二世代以降の DNA 配列シークエンサーをはじめとする大規模観測データを用いた生命科学のための解析アルゴリズムや解析基盤ソフトウェアの研究を通じ、ゲノム配列・転写・表現型の関係を探っていく。当研究室はいわゆるドライ系研究室で、実験設備は持っていない。

生命科学分野では DNA 配列シークエンサーの進歩により、ここ数年で DNA 配列の解読スピードが 1 塩基あたりで約 4 桁上がっている。1990 年にスタートした国際ヒトゲノム計画では、ヒトゲノムの解読に 13 年間と 3000 億円を要したが、2010 年 1 月に Illumina 社が発表した HiSeq 2000 は、1 回のランで 2 兆塩基の解読を可能とし、精度に大きな差はあるものの、曲がりなりにもヒトゲノム 2 人分を僅か 8 日間で解読することができる。

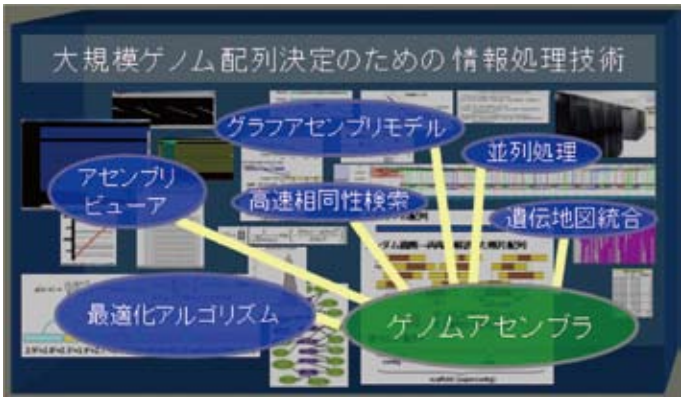
ゲノム配列決定のためのアルゴリズム研究

DNA 配列シークエンサーの最も直接的な応用はゲノム配列の決定である。ゲノム配列は分子生物学で最も基本的な情報として様々な研究に活用されている。

しかし、ヒト染色体で一番短い 21 番染色体が数千万塩基対あるのに対し、DNA シークエンサーが出力する塩基列は 1 本あたり数百塩基対しかない。このため、ランダムに裁断したゲノム配列を数億本読み、シークエンサーから出力された塩基配列をつなぎ合わせてゲノム配列を復元する情報処理（ゲノムアセンブリ）が必要となる。

我々は、より高精度かつ高速にゲノム配列を復元するアルゴリズムや実験の方法を開発することでゲノム研究を大きく加速することを目指している。

また、付随するテーマとしてゲノム配列と同様の手法を用いた転写産物の復元アルゴリズムや、第二世代 DNA シークエンサーを用いた短期間で低コストな染色体地図作成法にも取り組み、ゲノム配列を中心とした基礎データを効率良く収集するシステムを構築していく。



ゲノム多様性とその意味を理解するためのアルゴリズムやシステムの研究

新しい DNA 配列シークエンサーは生物のゲノム配列を新規に決定するのみならず、集団の遺伝的多型を調べる目的にも使うことができる。また逆転写と組み合わせて遺伝子の発現量を調べたり、バイサルファイト法や免疫沈降法と組み合わせて DNA やヒストンの修飾状態を調べたりと、観測対象を DNA の形に変換することで DNA 以外の様々な情報を観測することができる。

ゲノム配列や転写産物を効率よく決定できるようになれば、続いて個体・組織間のゲノムやエピゲノムの多様性を前述の様な方法を用いて調べ、転写や表現型との間の関係を探ることで、ゲノムやエピゲノム情報の意義や意味をより定型的な処理を用いて短時間に調べるシステムを作ることができるようになると考えられる。

このために、DNA 配列シークエンサーの出力からゲノム多様性の意味の理解を行うまでの間に必要な情報処理、すなわちアラインメント等の配列情報処理からゲノムの進化モデル・転写モデルの考案や関連解析に統計処理、可視化やパイプラインまで広範囲の研究やソフトウェア・アルゴリズムの統合を行うことを目指している。

大規模観測データ解析ためのプライベートグリッドミドルウェアの開発・研究

前述のように、大規模な観測データが産生されるようになったものの、データ量の急激な増加に対してその解析手法は充分に追いついていない。

「仮説を立て実験観測に基づきそれを立証する」という研究サイクルは自然科学分野の研究で広く行われているが、これは言い換えると、実験観測技術と仮説立証のための解析手法の限界に応じて調べることのできる仮説の範囲は決まってしまうということでもある。

我々は、膨大なデータから効率よく仮説検証を行うために、大容量データに対して並列計算を行うプログラムを研究者が短期間に開発することができるようなミドルウェア（研究者と OS の間に位置し、ユーザーアプリケーションの作成を容易にするためのソフトウェア）を開発し、実際の分子生物学研究に応用していく。

研究体制など

研究は、なるべく実験を行っている他の研究室と共同して進めるようにし、やりがいのある研究ができる環境を目指している。

情報理工学系研究科・コンピュータ科学専攻

バイオインフォマティクス 大規模並列計算

Yoshinori TAMADA
玉田嘉紀 助教
TEL: 03-5841-0517
tamada@is.s.u-tokyo.ac.jp
http://olab.is.s.u-tokyo.ac.jp/~reiji/sudalab.

情報理工学系研究科コンピュータ科学専攻に所属する須田研究室では、シミュレーション、並列アルゴリズム、数値計算や高性能計算に関する研究を行っており、その一つとしてスーパーコンピュータを用いた大規模遺伝子ネットワーク推定アルゴリズムの研究を行っています。生命科学分野では近年、遺伝子発現データやゲノム配列データなど、大量の観測データが蓄積されるようになってきており、計算機の性能向上と共にこのような大規模データから生命現象の謎を解き明かそうとする研究が盛んです。特に 2011 年に TOP500 において世界一位の計算速度を達成した「京」に代表されるスーパーコンピュータの性能向上は著しく、これまで不可能であった複雑かつ大規模な計算が可能になりつつあります。遺伝子ネットワーク推定問題もそのような計算問題の一つで、スーパーコンピュータを用いた並列アルゴリズムにより、これまで不可能であったヒト全遺伝子を含む大規模な遺伝子ネットワークを推定することが近年可能になっています。

スーパーコンピュータによる大規模並列計算

スーパーコンピュータを用いた計算では並列化が重要な鍵になります。「京」などの最近のスーパーコンピュータは、CPU 自体は私たちが普段使用している PC と比べ速いわけではなく、超並列型と呼ばれる膨大な数の CPU を高速なネットワークでつなげた構成になっています。このような計算機では適切な並列化なしにはその計算能力を活かすことができません。生命科学分野での計算問題は一般的にはデータ並列性があるものが多く並列化が比較的容易とされています。しかし遺伝子ネットワーク推定の大規模化は、これには当てはまりません。したがって、利用しようとしているスーパーコンピュータに適したアルゴリズムの研究開発がその計算能力を活かすためには必要になります。

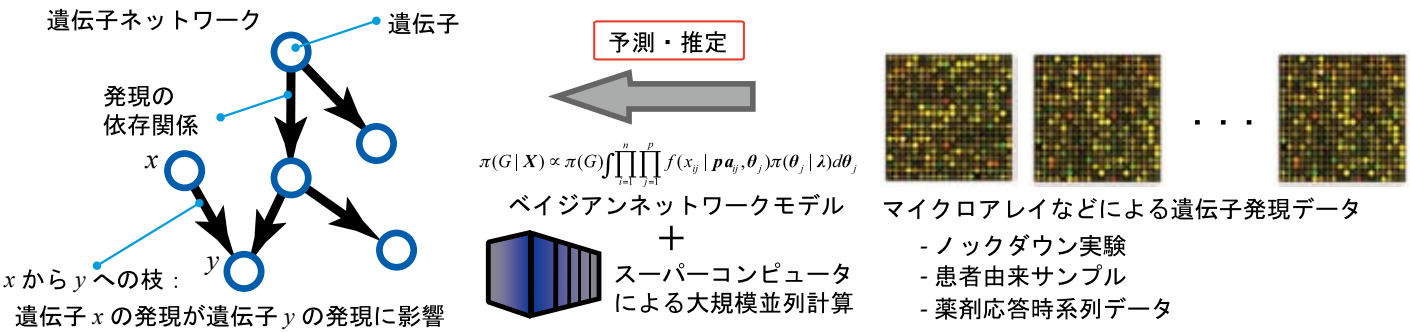
スーパーコンピュータを利用した大規模遺伝子ネットワーク推定アルゴリズムの研究

遺伝子ネットワーク推定問題は生命の基本単位である細胞の中の遺伝子がお互いにどのようにその発現制御を行っているか、実際の細胞から計測した RNA 等の発現データから複雑な計算アルゴリズムにより推定を行う問題です。ゲノム DNA 配列中の遺伝子から産生されるタンパクには他の遺伝子の発現制御を行っているものがあり、それらが協調的に制御し合うことにより発現制御のネットワークを構成し、生命活動を維持しています。マイクロアレイなどによって計測される大量の遺伝子発現データからこのような遺伝子発現の制御関係を推定・予測することができます。遺伝子ネットワークはそのように推定された制御関係を点と矢印（有向グラフ）で表したモデルです（図 1）。遺伝子ネットワークの推定には様々なモデル・計算方法がありますが、当研究室ではノンパラメトリック回帰を用いたベイジアンネットワークによる大規模並列アルゴリズムの研究をしています。ベイジアンネットワークによるネットワーク推定で一般的に用いられるアルゴリズムは並列化が難しいため、スーパーコンピュータによる高速化・大規模化は容易ではありません。従ってスーパーコンピュータの利用を前提とした並列アルゴリズムが必要になります。当研究室で研究しているアルゴリズムは、現在最大で 100,000 並列程度で動作し、さらなる高並列化を目指して現在も研究を続けています。

遺伝子ネットワークの応用とその他の研究テーマ

推定された遺伝子ネットワークを解析し、薬剤作用機序やがんの仕組みを解き明かすことも研究目的の一つです。そのために、既存データベースを活用したアルゴリズムや miRNA などの non-coding RNA の発現データを加えた遺伝子ネットワークの研究も進めています。さらに、今後は発現データだけでなくゲノムデータなどの情報も統合したアルゴリズムの開発にも取り組んでいく予定です。 また最近は GPU やメニーコア CPU などを搭載した、これまでにない新しい計算機が出現しています。当研究室では生命科学分野に限らず、このような新しい計算機の性能を最大限活かすためのアルゴリズム開発など、様々な研究を行っています。

図 1　遺伝子ネットワーク推定



http://www.bs.s.u-tokyo.ac.jp/bioinfo/

東京大学理学部生物情報科学科
〒113-0032 東京都文京区弥生二丁目11番16号