

特集

2

統合化と情報共有の先に見える光

データベースが生命科学の未来を変える

生命科学分野の研究で日々生み出される膨大なデータから、必要な情報を活用するためには、これまで研究者ごとに作られてきた多種多様なデータベースを統合化する情報基盤の整備が欠かせない。新たな基礎研究と産業への応用を支える、統合データベース構築の取り組みを紹介する。

Part. 1 疫学とゲノム研究を融合した基盤が新たな医療を導く

究極のヒトバイオロジーを探求したい

これまでの生命科学の研究の多くは、実験動物や細胞を用いて行われてきた。例えば、マウスの体内に人工的にがん細胞を作り、治癒させる実験がある。しかし、人間のがんへの応用においては、生物種の違い以外にも障壁がある。疾患のモデルとなるマウスは純系のため遺伝的に均一な上、同じ環境で飼育するので、ヒトの個人差や多様性を加味した研究が難しいのだ。こうした問題は、多様なヒトの情報を収集し大規模に解析することで解決されていくだろうと、JSTバイオサイエンスデータベースセンター（NBDC）統合化推進プログラム「大規模ゲノム疫学研究の統合情報基盤の構築」の研究代表者で京都大学大学院医学研究科附属ゲノム医学センターの松田文彦センター長は語る。

「最近では1滴の血液や尿から、がんの病状の指標（バイオマーカー）となるさまざまな数値がわかります。そして、病気になりやすさは、以前は『体質』によるものと考えられていましたが、ヒトのゲノム（遺伝子情報）解析により、実は個人の遺伝子配列のわず



統合データベースを開発中の松田さん（写真奥）率いるプロジェクトの研究室風景。写真手前左から、川口喬久さん、若栗浩幸さん。

かな違いと、生活や環境要因によることがわかってきました。分子生物学の研究が進み、大規模なデータを処理するコンピュータの環境が整ってきた今こそ、一人ひとりの遺伝的素因や環境要因に合わせた治療や病気の予防に役立つ、究極の“ヒトバイオロジー”を手がけたいと思います。そのための

基盤として、大人数集団を長期間観察して、生活習慣・環境の違い、罹患の種類、遺伝子情報などを網羅的に収集し、標準化して、互いに関連付けたデータベースの構築が必要なのです」

データ共有のカギは市民の参加と個人情報管理

この臨床・遺伝子情報を統合したデータベースの構築に向けて松田さんらのプロジェクトが現在取り組んでいるのが、本誌特集1で紹介している「ながはま0次予防コホート事業」だ。本事業では、1万人の市民の参加を得て「0次健診」と呼ばれる5年に1度の詳細な健康診断（健診）を行い、20年をかけて市民の健康状態に関するデータを収集する。

生活習慣などの市民への調査項目だけでも742もあるため、大人数の集団から長期間にわたり得られる情報は膨大だ。得られたデータをデータベース化し、研究者がそれぞれの研究に役立てられるよう整備する。デー

医療の進歩に貢献する
質の高いデータベースを
構築します。

松田 文彦 まつだ・ふみひこ

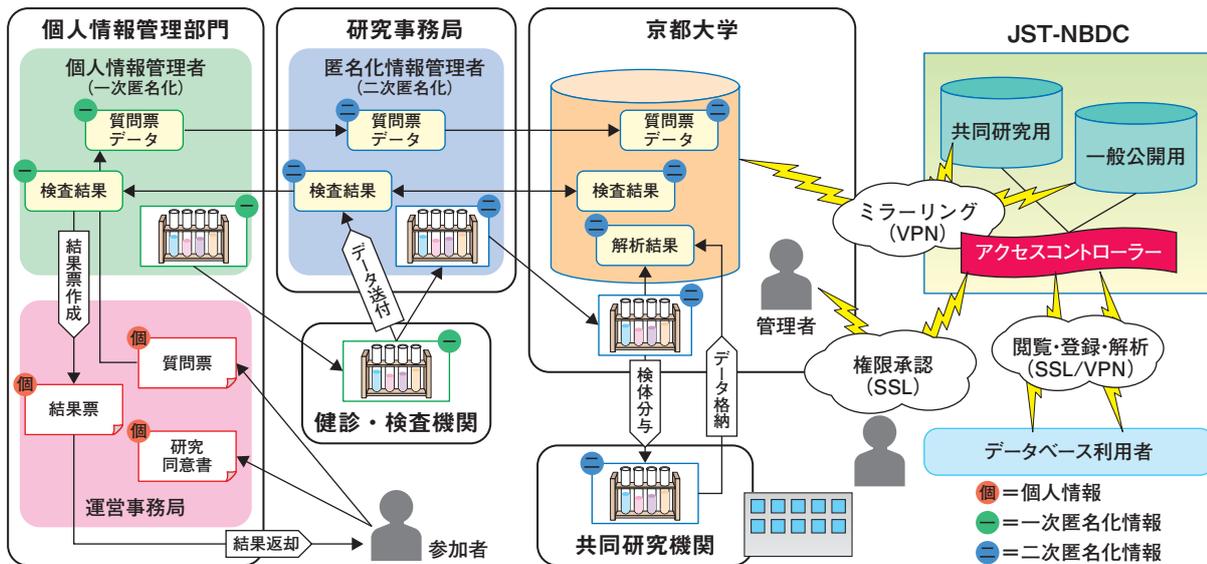
京都大学大学院医学研究科附属ゲノム医学センター センター長

1983年、京都大学理学部数理学系卒業。90年、同大学大学院医学研究科博士課程修了。医学博士。同大学遺伝子実験施設助手、フランス国立ジェノタイプングセンター研究部長などを経て、2007年から現職。





■個人情報を厳格に管理したデータの流れ



「ながはま0次予防コホート事業」における検体収集から解析、管理、提供に至るまでの流れを示したシステム概念図。参加者各人の検査結果や質問票などのデータは長浜市の個人情報管理部門で匿名化（一次匿名化）した上で厳しく管理され、京都大学の研究事務局で再匿名化（二次匿名化）した上でゲノム疫学研究を行う京都大学各教室へと送られる。このように二重の匿名化による、万全の個人情報保護対策が講じられている。

データベースの構築にあたっては、検査項目の設定や検体の収集・管理方法など、さまざまな検討課題をクリアしていく必要があった。

「万一、最初に作成した質問票の内容や手順に間違いがあれば、調査の結果がすべて無駄になりかねません。調査に先立ち、2005年末から1年半をかけて長浜市と質問事項について話し合い、07年夏には273人の参加者を集めてパイロット調査を実施しました」

制度設計時から、個人情報の厳格な管理は徹底している。情報管理者をどこに置くか、検体検査の責任者は誰かなどを明文化して具体的に記述した。これが08年に条例として成立した「ながはまルール」に記されている。市の管理者によって匿名化（一次匿名化）された個人情報には、研究者との間に設けられた匿名化情報管理者により再び匿名化番号を付与される（二次匿名化）。大学は個人識別情報を一切持たない。

市が情報管理に全面的な関与をすることによって、市民に安心を、研究機関に高い研究の自由度をもたらすことができた。同様に、ルールを整備したからこそ、研究者は長期間の追跡と調査分析ができるようになったのだ。

1万人の参加者への調査は、5年ごとに行われる。その間、長浜市は健診の実施や個人情報管理、参加者の健診結果報告などを行い、京都大学では収集した検体の網羅的遺伝子解析、網羅的オミックス情報（生命

分子情報）解析、疾患関連遺伝子同定などのゲノム疫学研究を行う。それらの結果が臨床・遺伝子情報の統合データベースに集約される。

いかにして質の高い疫学情報を集めるか

「ゲノム情報だけのデータベースを作るのはさほど面倒ではありません。遺伝子配列だけなら、約30億個の物質が並んでいて誰が解析してもほぼ同じ結果が出るからです。ところが、“健康”にかかわるデータは、環境の影響や生活習慣、食生活などを時系列で追跡し、数値化しなければいけません。近い将来、質問票による情報収集に代わって微量に存在する代謝物の測定が可能になれば、より正確な診断ができ、より質の高いデータベースを作ることができるようになるでしょう」

これらの測定値に加えて、松田さんは研究に参加する市民の医療機関受診時の情報や薬の処方などの医療情報も加えたいという。現在、医療機関では電子カルテの導入が進んでいる。しかし、電子カルテから機械的に情報を抽出しただけでは、大規模解析に用いる客観的データとしては使えない。電子カルテデータから利用可能な情報を効率よく取り出すには、IT技術をもって情報加工ができる医師の養成や、パターン化することによって、自動的に情報の取捨選択をす

る学習プログラムを開発する人材の育成が欠かせない。今回のプロジェクトでは、こうした人材の育成も含めて、ダイナミックで連続性のあるデータを網羅する学際的な研究を目指している。

データベース構築で可能になる「治す」から「予防する」医療

この研究で期待される成果の一つに、認知症の原因の解明があるという。現在75歳の日本人の認知症発症率は約15%とされる。「ながはまコホート0次予防事業」の参加者は、スタート時30歳以上74歳以下、そのうち、約半数は60歳を超えた人たちだ。60代の参加者を15年間追跡することで、認知症に関するさまざまな素因や体内で起こる兆候が明らかになる可能性がある。

「認知症患者は初期段階で周囲の家族が気づかないことが多いようです。極めて初期的な症状を正しくとらえるには、健康な時から長期間にわたって追跡し、認知度テストを行い、その情報をデータベースに登録することが必要です。仮に対象者が15年後に発症して、体内に特定の物質が増加していれば、データベースを調べ、過去にさかのぼって検証することができます。その原因や兆候がわかれば、治療薬の開発にも役立つでしょう」（松田さん）

「ながはまコホート0次予防事業」は今、1回目の1万人健診が終わったところだ。2回

目の健診結果は、5年後にそろそろだろう。スタート時点で参加した人の健康状態がこの5年間でどう推移したかが把握できれば、自分の状態を照らし合わせて、今後の健康を予測することもできる。例えば、BMI（肥満度を表す体格指数）の高い人は代謝系の病気にかかる確率が高まる、健康づくりをし

た人とそうでない人では健康の度合いに差が生まれるなど、科学的根拠にもとづき、5年先の自分をバーチャル体験し、健康を守るために今後注意すべき点を知ることができる。長期にわたって収集された疫学情報と遺伝子情報が統合されたデータベースを活用することで、個人に最適な生活指導が

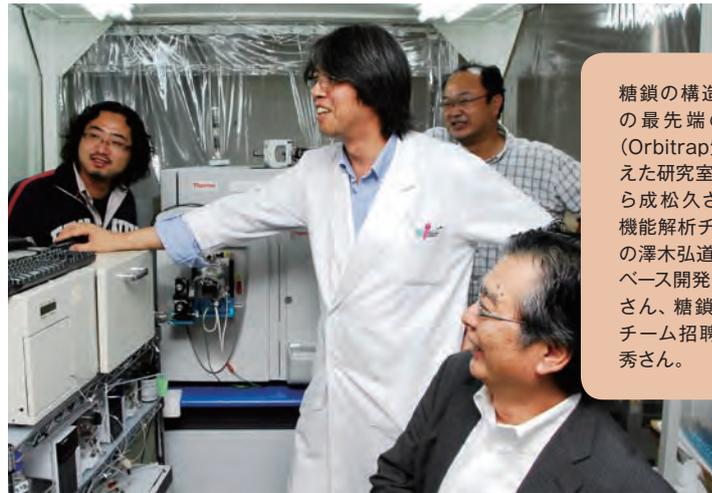
実現し、病気の発症前の高い確度での診断や、発症の防止や遅延を目指す「先制医療」が可能になる。そのための制度づくり、データの取得・解析、複合的な要因をつなげたデータベースを構築することで、「治す」から「予防する」へ、医療の新しい時代が始まろうとしている。

Part. 2 日本の強みを生かし、アジアの拠点を作る

糖鎖が生物の複雑さを決める

ヒトの体の設計図である遺伝子は2万2千個あり、生命活動に必要なタンパク質を作り出している。このタンパク質の多くには、糖が連なって多様な構造を持つ化合物である糖鎖が付いていて、タンパク質に多彩な機能をもたらしている。糖鎖は細胞を保護したりウイルスの受容体となったりと、生体内で重要な生理機能を担っている。血液型を決めているのも糖鎖だ。多様な構造の糖鎖を細胞内で合成するのは、「糖転移酵素」の役割だ。その数は約200種類弱、全遺伝子のうち約1%が糖転移酵素をコードする「糖鎖遺伝子」だ。

「糖鎖の構造はこれら200種の糖転移酵素の働きで決まってきます。タンパク質は基本的にその構造を変えられませんが、糖鎖は変えることができます。糖鎖は生物の種類、個体差、臓器の種類、細胞が成熟する過程などで構造が違い、更に、外界の刺激によっても構造を変化させる柔軟性を持っています。状況に合わせて糖鎖がまるで「洋服を着替えるように」変化できると考えるとわかりやすいかもしれません。その組み合わせはバラエティーに富んでいて、これこそが生物の複雑さを決めているのです」と、産業技術総合研究所（AIST）糖鎖工学研究センターの成松久センター長は語る。



糖鎖の構造を解析するための最先端の質量分析装置（Orbitrap質量分析計）を備えた研究室で。写真手前右から成松久さん、糖鎖遺伝子機能解析チーム特別研、究員の澤木弘道さん、糖鎖データベース開発チーム長の梶裕之さん、糖鎖遺伝子機能解析チーム招聘研究員の鹿内俊秀さん。

糖鎖はその変化の多様さから、病気などの指標になる「バイオマーカー」としての利用価値を見いだすことができる。AISTでは多様な構造の糖鎖生合成のメカニズムを解明するため、すべての糖転移酵素を探索するプロジェクトを進め、2002年度に終了した。AISTを中心に、200種類ある糖転移酵素の約3分の2が日本で発見され、日本はこの分野で世界を大きくリードしてきた。

独自の実験データも無償公開

AISTは10年度まで、文部科学省の「ライフサイエンス統合データベースプロジェク

ト」に参加し、これまでの膨大な糖鎖研究の成果の公開に努めてきた。

「私たちがリードしてきた研究成果を一般公開し、世界の研究者と切磋琢磨しています」（成松さん）

現在その活動はNBDCの統合化推進プログラム「糖鎖統合データベースと研究支援ツールの開発」プロジェクトへと引き継がれ、成松さんはその研究代表者を務めている。プロジェクト内では日本糖鎖科学統合データベース（JCGGDB）の機能強化を行っている。共同研究グループとして糖鎖のデータベースを持つ四つの研究機関、AIST、立命館大学、野口研究所、理化学研究所も参加している。互いに強みを持つ分野のデータを提供して網羅的に糖鎖関連情報データベースを拡大し、公開を行っている。この中には、01年から04年にかけて成松さんが発見した糖転移酵素遺伝子も収録されている「GlycoGene DataBase: GGDB」を始め、糖鎖と特異的に結合するタンパク質「レクチン」と糖鎖の相互作用をデータベース化した「Lectin Frontier Database: LfDB」、糖鎖を質量分析した際に得られるスペクトルから糖鎖構造を推定するための「Glyco Mass Special Database: GMDB」、"N型糖鎖"を持つ糖タンパク質の網羅的同定と糖鎖の結合位置をデータベース化した「Glycoprotein

意欲ある研究者のために世界をリードする日本ならではのデータベースを提供します。

成松 久 なりまつ・ひさし

産業技術総合研究所糖鎖工学研究センターセンター長

1974年、慶應義塾大学医学部卒業。79年、同大学医学研究科大学院微生物学専攻修了。医学博士。同大学医学部助教授、創価大学生命科学研究科教授などを経て、2006年から現職。筑波大学医学医療系連携大学院教授などを兼任。



■日本糖鎖科学統合データベース (JCGGDB) に搭載されたデータベースの一例

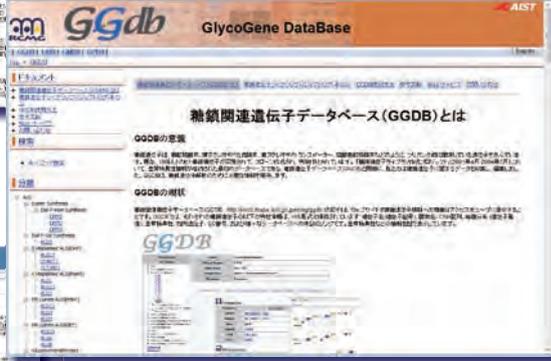
糖タンパク質データベース (GlycoProtDB)



糖鎖のタンデム質量
分析スペクトル
データベース (GMDB)



レクチンフロンティア
データベース (LfDB)



JCGGDBでは、糖鎖に関する多彩なカテゴリーのデータベースが次々と登場している。成松さんらのプロジェクトから、糖鎖関連遺伝子データベース (GGDB) やレクチンフロンティアデータベース (LfDB)、糖タンパク質データベース (GlycoProtDB)、糖鎖のタンデム質量分析スペクトルデータベース (GMDB) などのデータベースをリリース、他のデータベースとの統合を図っている。

糖鎖関連遺伝子データベース (GGDB)

Database : GlycoProtDB」などのAIST独自の実験データもあり、これらも無償で公開されている。

「このようなデータベースの公開により世界中で糖タンパク質の研究が進んだときに、私は最大のテーマであるヒトの糖タンパク質の構造解明に、若手研究者らと挑戦したいと考えています」と、成松さんは今後への意気込みを語る。

誰もが検索しやすい
世界の共通仕様へ

構築された糖鎖統合データベースを国内外の研究者が活用できるように、成松さんらは、海外のコンソーシアムとも連携し、カラフルな記号を使った糖鎖構造図や

キーワードで横断検索できるインターフェイスを開発した。更に、立命館大学の川崎敏祐教授と共同で、これから糖鎖の研究を始める研究者にもわかりやすく実験方法を解説する実験プロトコルのデータベース「GlycoPOD」を構築した。これまでに156名の糖鎖科学研究者が執筆した15分野187のプロトコルが、ウェブ上で公開されている。画面では試薬の情報や実験の手順、サンプルの量などが箇条書きで提示され、図解を用いた解説も掲載されている。また、情報のやり取りができるように、実験者が実験時の留意点を書き込む機能も備わっていて、研究者が使いやすいインターフェイスへと進化を続けている。

現在、糖鎖関連データベースは日本、アメリカ、ヨーロッパに散在するが、遺伝子

やタンパク質データベースのようなハブ機能を持たない。成松さんらの呼び掛けで糖鎖データベース標準化への動きがやっと始まった段階だ。また成松さんらは日本の共同研究体である「日本糖鎖科学コンソーシアム : JCGG」をアジアのコンソーシアム「ACGG」へと発展させ、アジア圏の糖鎖科学統合データベースの構築を目指している。日本が世界にリードする糖鎖研究が、今後発展の期待される中国、台湾、韓国などの国々と手を結ぶことで、国際的な連携・統合はますます強固なものになっていくだろう。そして、さまざまな糖鎖関連のデータベースが有機的に統合されることで見えてくる光が、生命科学諸分野の研究に大きなヒントを与え、生物の持つ複雑さ解明への道しるべになるに違いない。

JST担当者が語る NBDC「統合化推進プログラム」

生命科学分野の進歩に貢献する数々の研究開発が進んでいます。

NBDC (バイオサイエンスデータベースセンター) は昨年4月JST内に発足した組織で、生命科学分野のデータベースを統合し、データの価値を最大化することにより、日本の研究者のみならず、世界のユーザーに貢献することのできる、日本が誇れるデータベースセンターとなることを目指しています。

NBDCは、①戦略の立案、②ポータルサイトの構築・運用、③基盤技術開発プログラム (データベース統合化基盤技術の研究開発)、④統合化推進プログラム (バイオ関連データベース統合化の推進) の四つを活動の柱としており、今回の特集では、④統合化推進プログラムにおける二つの研究事例をご紹介します。

松田先生の研究課題「大規模ゲノム疫学研究の総合情報基盤の構築」は、滋賀県長浜市で行われている1万人規模のコホート研究で得られたゲノム疫学情報をデータベース化して研究者に広く提供することで、予防医学やがん検診などの分野で大きな貢献が期待されるものです。

成松先生の研究課題「糖鎖統合データベースと研究支援ツールの開発」は、糖鎖研究のデータベースの整備とそれらを活用するための検索システムの開発を推進しています。糖鎖研究は日本が世界をリードする分野であり、日本がこの分野のデータを整備することで、今後の糖鎖研究の進展に果たす役割は計り知れません。

現在「統合化推進プログラム」では11の研究開発課題が進行していますが、私たちJST担当者は、研究現場からの要望をプログラム運営に反映するなど、研究を加速させるためのさまざまなサポートを行っています。生命科学の進歩を支える一員として、これからも研究者の方々の声に耳を傾けながら、より良きパートナーとなるよう努めていきたいと思います。



JST 内田信裕