

ライフサイエンスに係る研究開発の将来動向調査

## 生命科学研究への情報科学技術のインパクト

ーゲノム総合科学、システム生物学、情報科学技術ー

平成 28 年 3 月

八尾 徹

理化学研究所 客員研究員

兼 慶応義塾大学 SFC 研究所上席所員

## 概 要

ゲノム・遺伝子・タンパク質・細胞、生命システムの解明と応用に、バイオインフォマティクス（生命情報科学）、計算生命科学、システムバイオロジー、及び Bio-IT（コンピュータ・インターネット）が大きな影響を与えつつあります。

21 世紀に入ってから、次の 3 つの大きなパラダイムシフトがあり、この傾向は今後ますます多面的に生命科学研究のスタイルを変えていくと思われま

- ① **ゲノムワイド解析**（ゲノム科学、オミックス、GWAS, タンパク構造ゲノミックス）
- ② **生命システム解明**（システムバイオロジー、マルチスケール解析、合成バイオロジー）
- ③ **情報科学の新展開**（データサイエンス、ビッグデータ、人工知能）

そして、これらは生命現象解明の基礎研究のみならず、その健康医療分野の研究と応用についても大きな影響を与えるでしょう。

この 3 つの視点に沿って、これらの国際動向を概観し、更に最近の動きを述べます。

### 1. ゲノムワイド解析

ゲノム科学、オミックス、GWAS, タンパク構造ゲノミックス  
21 世紀初頭のヒトゲノム解読完了を契機として、**ゲノムワイド解析**が新しい潮流となってきました。それまでの個々の遺伝子やタンパク質の構造や機能の研究に加えて、網羅的な測定によってそれら**生命部品群を一気に捕捉**しようとするものです。これを加速したのは、DNA シーケンサーや、オミックス解析、タンパク質構造解析など**先端生命計測技術の発達**と、**バイオインフォマティクス及びコンピュータ能力の発達**でした。

その結果、ヒトゲノム中にはタンパク質をコードする領域が予想外に少なく、むしろ多くの non-protein-coding RNA (**RNA 新大陸**) が存在し、それらが**遺伝子制御ネットワーク**を形成し重要な役割を果たしていることが分かってきました。更にはトランスクリプトーム（全転写物）解析等により、**エピジェネティックな遺伝子発現機構**の解明が始まってきています。

また、日米等での**タンパク質構造の網羅的解析**プロジェクトによって、非常に多くの立体構造が解明され、**転写・翻訳・シグナル伝達機構等の分子レベルの解明**に加えて、具体的に**医薬分子設計**に大きな貢献を始めています。

更に並行して展開してきていることが、3 つあります。一つは、**多人数ゲノム解析・多生物ゲノム解析**です。ヒトゲノムについては、一人から特定個人ゲノム (J.Watson, C.Venter, G.Church, M.Tomita ほか) へ、1000 人ゲノムへ、更には 10 万人ゲノムへと規模がどんどん拡大していっています。多生物ゲノムについては、各種の動物・植物がすでに数千種類解読されてきています。2 つ目は、**微生物群のメタゲノム解析**です。これは、地球上に あるいは体内に 存在する膨大な微生物群・細菌群の**集団ゲノム解析**です。例えばヒトの健康には体内細菌の影響が非常に大きいことが再認識されてきています。

3つ目は、**GWAS** (Genome Wide Association Study) のコホート研究の展開です。各国で数多くの大小プロジェクト(数千人~数十万人)が進行してきました。ゲノムと生活習慣データと健康・疾病などとの関連分析をするもので、それらの成果報告は、この10年余りで数百件に上ります(Nature Genetics ほか)。更に最近はこれらをまとめた**100万人規模のゲノムコホート研究プロジェクト**が始まりました。

このような中で、**バイオインフォマティクスは正に必須**のものです。1995年に初の生物ゲノム(インフルエンザ菌)の解読報告(Craig Venter ら)が発表された直後の、Chris Sander らの情報解析結果(多くの新規遺伝子同定)をみて Nature 誌が「バイオインフォマティクス時代の到来」と謳ったこと、及び2001年のヒトゲノム粗配列解読決定をした Craig Venter らの解析手法はコンピュータを駆使した「ホールショットガン」であったこと、いずれも時代を画す出来事でありました。

この間 今に至るまで、**バイオデータベース**(GenBank, PDB ほか)の果たした役割は地道ながら非常に大きく、また2007年以降の**次世代シーケンサー(NGS)**の発展が、上記のような大規模なゲノム解析・オミックス解析に果たしてきている役割も見逃せません。今後の**個人医療(Personalized Medicine--Precision Medicine)**の普及にも、これらのデータベースとシーケンサーは欠かせないものです。

**ゲノムワイド解析**(網羅的解析)は、先端生命計測技術の進歩とコンピュータ及び情報解析技術の進歩によって、今後の生命科学研究にとってますます重要性を増していくでしょう。

## 2. 生命システム解明 システムバイオロジー、合成バイオロジー

上記のような生命部品群(遺伝子・タンパク質ほか)の網羅的解明を受けて、**生命現象を系統的に解明しよう**という動きが、ヒトゲノム解読完了間近の2000年頃から世界的に始まりました。日米に続いて欧州各国・EUも次々とシステムバイオロジーのプロジェクトを発足させ、またセンター・学科の設立が相次ぎました。

ハーバード大に2003年に世界で初めて設立されたシステムバイオロジー学部(学部長 M.Kirschner)も、2005年にプリンストン大に設立された統合生命システム学科・研究所(所長 D.Botstein)も同様の設立理念(目的を共有する異分野の研究者が集まる体制)です。その後 **世界的にシステムバイオロジー関連学科・センター・研究所**が数百か所に作られています。いずれも異分野融合による新しいアプローチ及び人材育成を目指しています。例えば、2006年以降に英国各地に設立された6つの**国立システムバイオロジーセンター**でも各特定ターゲットの研究目的を目指して異分野研究者が同一研究棟で研究する体制(例えばニューカスルセンターでは植物根圏システム解明)を作りました。

また、2000年に米シアトルに設立された NPO の**システムバイオロジー研究所**(Institute for Systems Biology—所長 Leroy Hood)では、当初から生物学・医科学の

専門家と、先端生命計測技術開発グループと情報解析グループとの**3分野融合体制**で発足し、その後数々の研究成果を挙げてきました。

このような中で、2011年には、スタンフォード大の Covert グループが**単細胞丸ごとモデル**を発表しました。また、世界的に多くのコンピュータモデルが作られ、それらを集めたモデルデータベースが、**BioModels** として欧州 EBI に登録されることになっています (2015年現在 583モデル)。また**モデル開発言語 Cell Designer** や**共同開発システム GARUDA** が Hiroaki Kitano 主導で、国際的に開発されています。

このような活発なシステムバイオロジーの進展を受けて、**新しいタイプの合成バイオロジーの展開**が進んでいます。英国では、2013年に国策の一つとして「合成バイオロジー」ロードマップを発表しました。生命システムの解明知見をベースにした合成バイオロジー、あるいは生命システム解明のための合成バイオロジーを目指すとしています。それに基づいた各種の施策の中で、2014年から**国立合成バイオロジーセンター設立の動き**が始まっています。

もう一つ大事な視点として、**マルチスケール解析**があります。生命システムは、分子レベルー細胞レベルー組織・器官レベルー個体レベルー集団レベルと多階層であります。米国 NIH を中心とした多研究所・多省横断的な NIH-MSM グループの動きにも注目すべきでしょう。

システムバイオロジー及び合成バイオロジーは、共に**構成論的アプローチ**であることが特徴です。この傾向は、今後の大きな潮流と言えましょう。

### 3. 情報科学の新展開 データサイエンス、ビッグデータ、人工知能

近年の計測技術と情報技術の発展に伴い、社会のあらゆる側面で情報収集・伝達・活用が急速に進んでいます。いわゆる「**ビッグデータ**」の影響が日に日に強くなっています。このことはビジネスの世界だけでなく、個人の生活にも、また科学技術面でも影響を与えつつあり、その度合いは今後極めて大きくなっていくと思われまます。

もちろん、生命科学研究と応用の面でもその影響は計り知れないものがあります。**例えば健康関連情報**で言えば、個人の行動・体調・環境などのデータが、ブログ・スマホ・装着センサーなどから、刻々収集できる状況になりつつあります (臨床研究・コホート研究・病院カルテ・健診データなど定型データに加えて、新たな非定型データとして例えば感染症発生兆候予測などに)。またそれらの多様なデータの**解析技術 (アナリティクス)** が**人工知能技術**を含めて急速に進んでいます。

これらの大きな動向を踏まえて、**欧米でもわが国でもその研究や施策**が始まっています。そこには**グーグル社、IBM** など巨大情報産業の**先導的な動き**があり、例えば健康

医療産業の今後の展開に大きな影響をもたらすと考えられています。正に「データ駆動型研究」「データサイエンス」が第4のアプローチになろうとしているのです。

科学技術面でのビッグデータの影響については、2008年にNature誌が先駆的な特集号を出しました。そこでは、Google社の動きと次世代シーケンサーNGSの影響が取り上げられました。Google社の膨大な文章情報・画像情報・映像情報の収集・蓄積・提供はその後もとどまるところを知らず、更に新たな解析サービスの提供によって世界の知的活動にますます大きな影響を与えつつあります。また、NGSのその後の更なる性能向上によって、感染症についての国際的大規模疫学のビッグデータによる原因究明が進み(2014年、Science)、更に上述のように2015年から100万人を対象とするゲノムコホート研究によって個人別医療に向けたPrecision Medicineがスタートしました。

その間、第4の科学としてデータサイエンスが提唱され(2009)、それに基づいたNIHの施策が始まり、ビッグデータへの対応が急ピッチで進んでいます。またマッキンゼー報告書健康医療分野でのビッグデータ(2013)には多くの事例が紹介され、この分野の展開に大きな影響を与え始めています。カリフォルニア州でも健康医療へのビッグデータ活用に大きく動き出し、2015年4月には、UCSF Institute for Computational Health Sciencesが設立され、Atul Butteが初代所長に任命されました。

世界的に長年に亘り脳科学の研究が進められてきましたが、複雑な脳機能の解明にはまだまだ年月がかかるでしょう。しかし、その中でもコッホ・トノーニなどによる意識・記憶・学習の研究と理論化・数量化が近年急速に進んできました。これを受けて今後、脳疾患・教育・人工知能への応用が進んで行くと思われれます。特に人工知能の進歩は、上記ビッグデータの解析にとって大きな力になり始めています。

ビッグデータ及び人工知能の、今後の社会への影響、更には科学技術研究への大きなインパクトには注目が必要でしょう。

## 目 次

|  |    |
|--|----|
| まえがき、謝辞                                    | 10 |
| 概論   | 11 |
| 1. 生命科学研究と情報科学                             | 11 |
| 2. 21世紀初頭の生命科学の特徴と変革                       | 13 |
| 各論   |    |
| 1. ゲノム総合科学の展開と情報科学                         | 15 |
| 1-1 ヒトゲノム解読計画から完了まで 1990-2003年             | 15 |
| その展開（シーケンサー、HGP、セセラ社）と成果                   |    |
| 1-2 本格的バイオインフォマティクスの興隆 1995年以降             | 17 |
| ① バイオインフォマティクス時代の到来                        |    |
| ② ホモロジーサーチ                                 |    |
| ③ 比較ゲノミクス                                  |    |
| ④ ホールショットガン法                               |    |
| ⑤ 2000年頃までのゲノム解読                           |    |
| 微生物, モデル生物（酵母・線虫・ハエ・シロイヌナズナ）               |    |
| 1-3 ゲノム科学の発展 2000年以降                       | 18 |
| ① 多人数ゲノム、個人ゲノム、がんゲノム、多生物ゲノムの解析             |    |
| ② 発現解析、オミックス（トランスクリプトーム、プロテオーム、メタボローム）     |    |
| 1-4 ヒトゲノム機能解析プロジェクト ENCODE 計画 2003年～       | 19 |
| 1-5 FANTOM コンソーシアム 2004年～                  | 20 |
| RNA新大陸発見、遺伝子制御ネットワークの重要性                   |    |
| 1-6 ゲノムから表現型へ GWAS、エピジェネティクス、オミックス         | 21 |
| トランスオミックス解析（オミックス平面解析からオミックス空間解析へ）         |    |
| 1-7 植物分野でのゲノム総合科学                          | 24 |
| 1-8 微生物ゲノム・メタゲノム解析                         | 25 |
| 1-9 次世代シーケンサーの登場（2007年～）とそのインパクト           | 26 |
| 1-10 生命科学研究を支えるデータベース群の発展 1970年代～, 2000年代～ | 28 |
| 1-11 ポストヒトゲノムの10数年間のインパクト                  | 30 |
| 2. タンパク質構造ゲノミクス・構造生物学の進展と情報科学              | 31 |
| 2-1 タンパク質立体構造解析の進展と成果                      | 31 |
| ① その重要性と解析技術                               |    |
| ② 構造ゲノミクスプロジェクト（日本、米国）                     |    |

|     |  |    |
|-----|--|----|
| ③   | 立体構造解析による作用機序の解明—生命現象解明への貢献<br>(ATP合成酵素、リボゾーム、チトクローム C, タンパク質合成、アディポネクチン他) |    |
| 2-2 | 構造バイオインフォマティクスの進歩 データベース、分子計算、医薬分子設計                                       | 32 |
| ①   | 立体構造データベースとその解析 PDB, SCOP  |    |
| ②   | 立体構造予測・モデリング CASP (構造予測コンテスト)<br>インターネットコンテスト利用による技術開発の加速 (事例)             |    |
| ③   | 分子シミュレーション MD (Molecular Dynamics)<br>マルチスケールモデリング、スーパーコンピューティング           |    |
| ④   | 構造ベース医薬分子設計 SBDD (Structure Based Drug Design)                             |    |
| 2-3 | 構造解析・設計技術の最近の進歩  | 35 |
| ①   | X線自由電子レーザー法による動的状態測定   |    |
| ②   | 低温電子顕微鏡 Cryo-EM の新たな展開   |    |
| ③   | 複合法による構造モデリング  |    |
| ④   | タンパク質構造ダイナミックスの解析  |    |
| ⑤   | 人工アミノ酸、人工塩基対による新規タンパク質   |    |
| 3.  | システムバイオロジー、合成バイオロジーの進展   | 38 |
| 3-1 | システムバイオロジーの展開 15年  | 38 |
| 3-2 | システムバイオロジー元年(2000 ICSB, NIH, ISB), 特集 (2002 Nature, Science)               | 38 |
| 3-3 | 米国のシステムバイオロジー展開  | 39 |
| ①   | システムバイオロジープロジェクト(2000年～)   |    |
| ②   | 米国国立システムバイオロジーセンター群 設立 (2007年～)  |    |
| ③   | ハーバード大 システムバイオロジー学部 設立 (2003年)   |    |
| ④   | 民間システムバイオロジー研究所 ISB 設立 (2000年)   |    |
| ⑤   | がんシステムバイオロジーセンター群 設立(2008年～)   |    |
| ⑥   | システムバイオロジーベンチャー企業 (2000年～)   |    |
| ⑦   | Systems Medicine, Personalized Medicine, Precision Medicine へ              |    |
| 3-4 | 欧州システムバイオロジーの展開  | 40 |
| ①   | ドイツのシステムバイオロジー(2003年～) HepatoSys project                                   |    |
| ②   | スイスは 2007年に国家プロジェクト SystemsX.ch を発足。                                       |    |
| ③   | イギリスは 2006年から国立システムバイオロジーセンターを6か所に設立。<br>2012年から合成バイオロジーを国家的重要施策に          |    |
| ④   | EUのシステムバオロジー推進策 (2005年～)   |    |
| 3-5 | 植物システムバイオロジー ベルギー、カナダ、日本(篠崎・斉藤・平井他)  | 41 |
| 3-6 | 細胞システムバイオロジー   | 41 |
|     | 単細胞丸ごとコンピュータモデル (2012年スタンフォード大 Covert)                                     |    |

|  |    |
|--|----|
| 細胞間変換(Direct Reprogramming)  | 42 |
| 3-7 モデルデータベース BioModels  | 42 |
| 3-8 日本の活躍 (上田、北野、富田、金久、末松・倉智・野間・久田・岡野・松野・<br>宮野・高木・黒田・岡田・大浪・金子・古澤・合原・田中・杉山・石川)   | 43 |
| 3-9 合成バイオロジーの動き  | 44 |
| ① 概要   |    |
| ② 具体的な動きの例   |    |
| ③ 合成バイオロジー国際コンテスト iGEM   |    |
| ④ 合成バイオロジーベンチャー会社  |    |
| ⑤ Nature Biotechnology 誌 (2009年12月号に Synthetic Biology の特集)  |    |
| ⑥ 英国の合成バイオロジー展開  |    |
| ⑦ その他注目事項  |    |
| 参考 <a href="#">システムバイオロジー内外動向 2011</a><br><a href="#">合成バイオロジーの動向 2013</a><br><a href="#">システムバイオロジーと創薬技術 HS 報告 2008, 2014</a><br><a href="#">日本のシステムバイオロジー JST 2010</a><br><a href="#">米国システムバイオロジー調査団 2004</a> |    |
| 4. 脳科学の進歩と情報科学、そして人工知能   | 48 |
| 4-1 はじめに   | 48 |
| 4-2 脳科学研究 EU, 米国, 日本   | 48 |
| ① EU の Brain Project<br>Horizon 2020 –神経科学・情報科学・医科学の統合展開   |    |
| ② 米国 BRAIN Initiative<br>オバマイニシヤティブー精神疾患<br>参考: コッホの「意識」研究   |    |
| ③ 日本の脳科学総合研究<br>理研の総合研究センター 及び脳機能ネットワークの全容解明プロジェクト   |    |
| 4-3 意識の統合情報理論  | 50 |
| 意識はいつ生まれるのか、トノーニの著書紹介、意識の統合情報理論<br>覚醒・意識・記憶 オプトジェネティクス (井ノ口・利根川)<br>脳とグリア細胞 (工藤)   |    |
| 4-4 人工知能研究   | 51 |



|   |                              |
|---|------------------------------|
| その歴史・現況・展望                                    | 52                           |
| 参考：「人工知能研究半世紀の歩みと今後の課題」 西田豊明 情報管理 No.7, 2012  |                              |
| 「米国における人工知能研究に関する取り組みの現状」 八山幸司 JETRO-IPA 2015 |                              |
| 「研究開発の俯瞰報告書－情報科学技術分野 人工知能」 JST/CRDS 2015      |                              |
| 日本の近況   | 53                           |
| 人工知能研究センター設立、産業界の動き、AI 利用成功事例 など              |                              |
| 辻井、山川・一杉・高橋・松尾、トヨタ・IBM・楽天・リクルート、奥野・三宅         |                              |
| 計算能力・記憶能力代替から思考能力・発見へー北野提唱                    |                              |
| 5. 情報科学の新たな展開ーデータサイエンス、ビッグデータ                 | 55                           |
| 5-1 概要  |                              |
| 5-2 科学技術でのビッグデータの影響                           | Nature 2008 (Google, EBI) 55 |
| 5-3 科学における第4のパラダイム データサイエンス                   | Microsoft Research 2009 56   |
| 5-4 ビッグデータに対する考え方 ビッグデータの正体                   | Kenneth Cukier 2013 57       |
| 5-5 健康医療分野でのビッグデータ                            | Mckinsey Report 2013 58      |
| 5-6 欧米の取り組み 米国、NIH, 米州、英国ほか                   | 58                           |
| ① 米国のビッグデータへの取り組み                             |                              |
| ② NIH のデータサイエンスに対する取り組みと BD2K センターの設立         |                              |
| ③ 米国州政府の健康医療へのビッグデータ活用、計算健康科学研究センター           |                              |
| ④ 欧州 英国、スイス                                   |                              |
| ⑤ 参考文献  |                              |
| 5-7 日本の取り組み                                   | 60                           |
| ① 情報通信政策 「データが切り拓く未来社会」ー社会全般                  |                              |
| ② 関係者 (バイオインフォマティクスリーダー)ー論文・総説                |                              |
| 田中博、松田、高木、宮野、田中謙                              |                              |
| 5-8 今後の展望 ビッグデータと人工知能                         | 61                           |
| データ・情報の発生源の変化 (IoT 社会)                        |                              |
| 終わりに  | 62                           |
| 課題・関心事など                                      |                              |

## まえがき、謝辞

今回、ライフサイエンス振興財団の依頼によって、「生命科学研究への情報科学技術のインパクト」について、動向調査をまとめることになりました。このテーマはあまりにも対象が広く大きく短期間にまとめることは難しいのですが、著者が長年担当してきたゲノム・遺伝子・タンパク質を起点とする「生命情報科学」の動向を中心に、歴史的・国際的な視点からまとめてみました。

その内容は、概要・各論の目次以下にある通りですが、それぞれに詳しい内容を記述出来ませんでしたので、参考文献・資料及びオンラインアクセスサイトをご参照して頂ければ幸いです。

理化学研究所ゲノム科学総合研究センターの前所長和田昭允氏及び榊佳之氏には長年のご指導を厚く感謝し、特にお礼を申し上げます。そしてこれまでにお世話・ご指導頂いた内外の研究者・関係者にも厚くお礼を申し上げます。

## 概観

### 1. 生命科学研究と情報科学 (生命科学与情報科学の相補関係—別図)

#### 1) 情報科学的に見た生命体の特徴

生命体は無機的な物質・物体ではなく、以下のような特徴を持ちます。

- ・常時動いている。変化していく。個別である。再現性がない。
- ・相互作用をしている。集団行動をする。
- ・環境の影響を受ける・避ける・働きかける。
- ・自己複製・自己修復する。遺伝する——生命体自身が情報を持っている。
- ・高等生物・特に人間においては、精神的な情報の扱いが必要である。

#### 2) 生命科学研究にとって情報科学は必須

上記のような特徴を持つ生命体の研究には、まず最初に総合的な**観察**記述情報が必要で、次に対象を絞った個々の**測定**データの収集・処理・解析・蓄積が必要になります。そしてそれらをまとめて**総合的な理解**に至るのです。

これらのすべての情報を扱い、生命現象の解明を目的とする生命情報科学 (バイオインフォマティクス) は、生命科学研究にとって当たり前の存在であり、すべての対象・段階で必須のものとなっています。

そこには、日常的に生命科学研究者が誰でも使う**データベース・ソフトウェア**から、非常に大量で複雑な情報を扱う高度な**数理解析技術** (シミュレーション・人工知能など) まで、各種の段階があり、当然、それらの**整備や普及促進と先端的な研究開発推進**とが並行して進められなければなりません。

その背景には、**生命計測技術の進歩**によるデータの急増、**ゲノム解析・ゲノム総合科学の進展**、**生体部品の解明から生命システムの理解**への指向、**個人・社会・環境からのデータ収集** (自動、高速、大量) に伴う**ビッグデータの急増**など様々な要因があり、またそれらを支えるコンピュータの発達・インターネットの普及・オープンサイエンス、クラウドソーシング、データサイエンスの発展があります。これらを踏まえた総合的な施策と人材育成が今後の急務です。

C.Ouzounis 15 years of Bioinformatics “Rise and Demise of Bioinformatics”

PLoS Computational Biology 8(4), Apr.26, 2012

「生命とは？ 物質か！」 和田昭允著 オーム社 2008

「生命とは、情報！」 豊田哲郎 理研講演会 2010

「オミック スペース」—「概念」和田昭允、「図示」豊田哲郎 2002

「トランスオミックス アナリシス」黒田真也 東大黒田研 HP2012

「生命のストラテジー」 松原謙一・中村桂子 岩波書店 1990

「生物と無生物のあいだ」、「動的平衡」 福岡伸一 講談社 2007, 2009

「したたかな生命—ロバストネス」 北野宏明 ダイヤモンド 2006

# 生命科学と情報科学

## 21世紀 生命科学と情報科学

### バイオインフォマティクスへの期待

生命情報科学、生命情報学  
計算生物学、計算情報生物学  
システム生物学、生命情報システム学

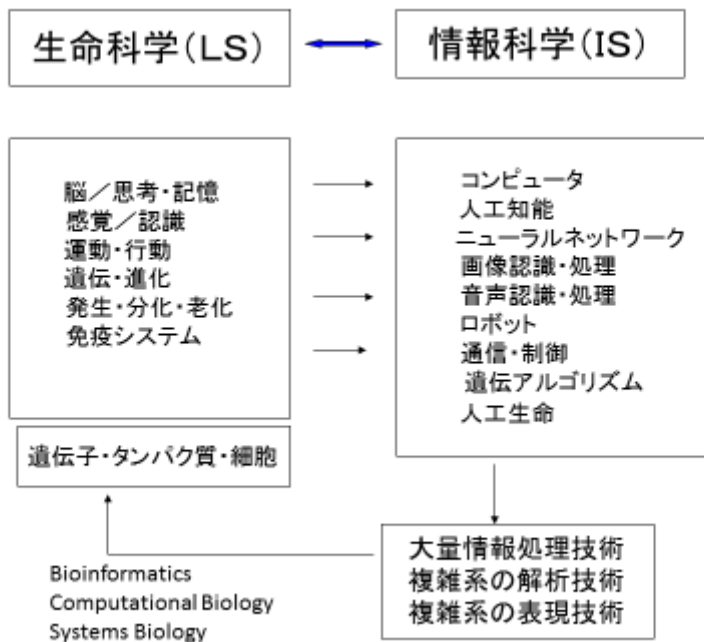
- 「生命現象の解明に情報科学技術を活用する」
- 「生命現象を情報科学的に理解する」
- 「生命現象を情報システムとして理解する」
- 「生命現象から情報科学技術を学ぶ」
- 「生命体となじむシステムを開発する」

### 生命現象の

個別的・部分的な解析から総合的・包括的な解析へ

17

概念図 ・生命科学・現象から学ぶ情報科学  
・情報科学・技術が牽引する生命科学研究



18

## 2. 21世紀初頭の生命科学の特徴と変革

- 1) ゲノム研究の大きな展開
- 2) パラダイム変革ーゲノムワイド解析
- 3) ゲノム総合科学へ 新たな発見と展開
- 4) 計測・実験・情報解析技術の進展
- 5) ゲノム情報の応用と社会、総合的人材の育成

21世紀に入って、生命科学研究のスタイルは大幅に変わりました。1990年初頭から始まったヒトゲノム計画が2000年初頭に完了した時、それまで半世紀に亘って数々の成果を挙げて発展してきた「分子生物学」は正に一つの踊り場に達しました。次の大きな発展段階を迎えたのです。それまでの個別解析から網羅的解析ーゲノムワイド解析ーの時代に突入しました。

微生物から植物・動物、そしてヒトに至るまで網羅的なゲノム・各種オームの解析(ゲノミクス・オミクス)が始まり、一挙に多数の遺伝子が同定されるようになってきました。このような大きな変化を踏まえて、各国ではポストゲノム時代のプロジェクトが次々と発足し、またそれに呼応した学科やセンターが相次いで設立され、新たな人材育成が進んできました。正にゲノム総合科学の発展の時代を迎えました。

そこでは多様で大量な生命情報を扱うバイオインフォマティクスが必須な科学技術として位置づけられてきました。この傾向は最近の次世代シーケンサーの発達や先端生命計測技術(イメージング等)の進歩により益々加速されてきており、更に急速に拡大しつつあるビッグデータの扱いと複雑な現象の情報解析技術の向上が喫緊の課題となってきました。

このような流れの中で、生命現象をシステムとしてとらえようとする「システムバイオロジー」が、2000年頃から提唱され、その後各国が強力な施策を続け、この10数年あまりで急速な進展をしています。各種オーム(ゲノム、トランスクリプトーム、プロテオーム、メタボローム、フェノーム等)データすなわち生命部品群データが網羅的に測定される時代に入って、それらを組み合わせた生命システムの理解は必然の方向であり、正に21世紀はシステム生物学の時代と言われ始めています。

更にその知識を基にした合成生物学への展開も始まっています。複雑な生命システムを扱うバイオシミュレーションが必須の技術となり、生命現象基本システム(シグナル伝達・代謝・細胞周期ほか)のモデル化と共に、Whole Cell Modeling, Virtual Cell, Virtual Organ, Virtual Physiology, Virtual Patientなどのモデルが開発されてきております。

更に近年の急速なインターネットの発展とオープンサイエンスの傾向から、個人・社会・環境データを含むビッグデータの活用と、その解析技術(特にベイズ統計と人工知能)に期待が高まっています。

そして、これらゲノム総合科学とシステム生物学、更に情報科学は、その応用面でも広く健康・医療分野から環境・エネルギー・材料分野まで、大きな期待を集めています。特に、医療面では個別化医療・予防医療のような具体的な応用が始まっています。これらは先端計測技術・コンピュータ技術の進歩と組み合わせて、今後 20 年～50 年以上に亘る社会へのインパクトは計り知れないものがあります。

また生命科学研究成果の応用に当たっては、ELSI (Ethical, Legal and Social Issues) の議論が重要になっています。21 世紀の生命科学・バイオテクノロジーの発展には、総合的な視野を持つ新しい人材の育成が今後の大きな課題です。

## 1. ゲノム総合科学の展開と情報科学

21世紀初めのヒトゲノム解読計画の完了は正に科学史を飾る偉業でした。生命科学研究に、「ゲノムワイド解析」（網羅的解析）というパラダイムシフトをもたらしたことと、国際的共同研究によるオープンサイエンスを実現したことは、画期的なことでありました。

以下に、その経過とその後の展開を記し、それらの中での情報科学・バイオインフォマティクスの動きを、年代を追って振り返ってみたいと思います。

### 1-1. ヒトゲノム解読計画から完了まで（大規模国際コンソーシアムで）1990-2003年

1987年日本で開かれた「DNA塩基配列高速解析ワークショップ」は、その後のヒトゲノム解読計画の発足に向けて大きなきっかけとなりました。岡山県林原研究所で開かれたこのワークショップには、欧米日から約100人のこの分野の専門家が集まりました。特にここで紹介された日本の「DNA塩基配列高速解析システム」プロトタイプの完成報告（Nature 1987）は、参加者に衝撃を与えました。1982年から和田昭允主導の下、大学・企業の研究者が共同で開発してきた文部省プロジェクトの成果でした。当時ヒトゲノムDNAの全解読は夢物語と思われていましたがこのような自動化・高速化を進めれば数十年で可能になりそうだとの認識が高まりました。

このワークショップを契機として、ヒトゲノム解読計画の具体化構想が一気に進み1990年に米国を中心とした「ヒトゲノム計画」HGP 国際コンソーシアムが発足しました。米国は15年間で\$3B(約3000億円)を拠出するとしました。ここには英国・ドイツ・日本など数カ国が参加しました。その進行の途中1996年には、関係者が集まったバーミュダ会議が開かれ（榊佳之参加）、読み取られた塩基配列データは24時間以内に世界に公開するという取り決めがなされました。ゲノム情報は人類共有の財産であるとされました。正にオープンデータの先駆的な動きでした。

この時期の世界のヒトゲノム解析センターは10カ所程度で、1台数千万円するシーケンサーをそれぞれ多数台導入しました。アメリカには4カ所の大センター(ボストン、セントルイス、ヒューストン、バークレイに各数十台)があり、イギリスケンブリッジに大規模なサンガーセンター(約100台)があり、その他はドイツ・日本ほか数カ所に中規模センターで、これらが分担する染色体番号を明示してシーケンス解読を進めていました。

ところが、1998年に **Craig Venter** が率いるセレラ社が発足し、3年以内にヒトゲノム全解読を済ませ、主要な遺伝子情報の特許を取ると発表しました。当時の最速DNA

シーケンサーABI370を230台導入し、最新鋭コンピュータを使ったホールショットガン法で短期間に全配列を決めてしまう、そのために**情報解析研究者・技術者を250名採用し、最先端コンピュータ導入を含めた全費用を500億円で収めるとの計画**でした。この発表は上記国際コンソーシアムに大きな影響を与え、HGPはヒトゲノム粗配列を早期に公表し、精密配列完了時期も早めると計画変更をしました。

結果的に、粗配列は2000年6月に、国際コンソーシアムとセセラ社が同時発表し、精密配列は2003年4月に(当初計画より**2年短縮**、DNA二重らせん構造決定から**50周年目**に)国際コンソーシアムが発表しました。結局ヒトゲノム解読の国際的な達成比率は、アメリカ60%、イギリス30%、日本6%、その他4%ということになりました。この国際コンソーシアムの成果は、科学史上に残る偉業であったと言っても過言ではありません。

尚、セセラ社は、上記の粗配列解読完了前後の数年間にヒトゲノム配列データの販売をしていましたが、国際コンソーシアムの精密データが無料公開された後は、営業方針を変更してしまいました。データそのものが売り物になるというビジネスモデルは成功しませんでした。しかしこの時のセセラ社の動きは、正にコンピュータをフルに駆使したバイオインフォマティクス力の強烈なインパクトを世に示しました。

### ヒトゲノム解読計画完成の成果

ヒトゲノム解読計画完成の意義と成果については、当時非常に多くの論説が出ましたが、その要点をまとめると次のようになるでしょう。(Francis Collins ほか)

- ・ライフサイエンスにおける大事業を国際共同で達成したこと
- ・オープンデータを貫き、オープンサイエンスの道を拓いたこと
- ・その後のヒトゲノム解読の基盤(参照配列)を提供したこと
- ・その後のヒトゲノム研究の方向を示したこと(遺伝子数の少なさ、機能未知遺伝子)
- ・ゲノム情報の医療・社会への応用の方向を示したこと(ゲノム医療ほか)
- ・ゲノム研究におけるELSI問題(個人情報、倫理、権益ほか)を提起したこと

### 1-2. 本格的バイオインフォマティクスの興隆(初の微生物ゲノムが契機) 1995年以降



上記のヒトゲノム解読完了に数年さかのぼって、ゲノム時代にはバイオインフォマティクスが必須のものとなるとの認識が世界的に高まって来ました。

#### ① バイオインフォマティクス時代の到来(1995年～)

その契機となったのは、1995年6月に発表された史上初の微生物ゲノム-インフルエンザ菌ゲノム(1.8Mbp)の発表でした。TIGR社の**Craig Venter**以下の研究者約20名が数年かけて解読したゲノム報告が**Science**誌に発表されるや否や、英国ケブリッジにある欧州バイオインフォマティクス研究所(European Bioinformatics Institute)の **Chris Sander**以下5名がそのゲノムデータの解析を数日で終え、10数個の重要遺伝子を検出したと翌7月の**Nature**誌に発表しました。この報告はネイチャー編集者によって「バイオインフォマティクス時代の到来」**“Bioinformatics: Comes of Age”**と評され、その後の世界的なバイオインフォマティクス興隆へとつながって行きました。

#### ②ゲノム塩基配列相同性解析—ホモロジーサーチ

ただそれ以前(1990年頃)から、NCBI/NIH の**David Lipman**等によるDNA情報解析ソフトウェア**BLAST** がすでに存在し日常的に使われていたことを見逃すわけにはいきません。この **BLAST** はその後もDNA情報解析の中心的存在で、ゲノム時代に入って高速化の努力が継続してなされて来ました。

#### ③比較ゲノミクス—進化、保存遺伝子

この間(1995-1999) のもう一つの重要な動きは、「比較ゲノミクス」の始まりです。当時大腸菌ほか多くの微生物ゲノムが解かれていく以外に、いわゆるモデル生物といわれる酵母・線虫・シュジョウバエ等のゲノム解読が進展していました。1997年米国NCBI/NIH の **Eugen Koonin**等は、真正細菌・古細菌と酵母(最小の真核生物)のゲノム(数Mbp-数百Mbp)を用いて、これら3王国で進化上保存されている遺伝子の同定・分類をし(Nature 1997)、そのデータベース**COG (Conserved Orthologous Genes)** を発表しました。その後、哺乳類ゲノム(~数Gbp)が多く解読されている今、それらの比較ゲノミクスは想像を絶する規模になっております。そしてこの**COG**は更に拡張されて現在でも比較ゲノミクスの基本データベースになっています。

#### ④ ホールショットガン法      ゲノム解析技術—BAC法との優劣

TIGR の微生物ゲノム(1.8Mbp)解析には、当初からホールショットガン法が使われました。セセラ社はこの方法をヒトゲノム(約3000Mbp)解析にも採用しようとしたのですが、イントロンの多い高等生物には使えないという見方が当時は一般的でした。ヒトゲノム国際コンソーシアムの方はBAC法で進めていました。その論争に応えるためにセセラ社は、1998年にショウジョウバエゲノム(約100Mbp)解読にホールショットガン法を使い十分な精度が出ることを示しました(Science 2000)。これには数理解析の力とコンピュータ活用の力が必要で、このことで貢献した **Gene Myer** は**バイオインフォマティクス技術者**として世界的に有名になりました。結局セセラ社のヒトゲノム解析はホールショットガン法でなされました。

⑤ **2000年頃までの各種ゲノム解析** ヒトゲノム解読に並行して、1995年以降各種生物のゲノム解読が、ヒトゲノム解読に並行して進みました。主なものは、大腸菌・酵母・線虫・ショウジョウバエ、シロイヌナズナなどのモデル生物と、多くの微生物でありました。

### 1-3. ゲノム科学の発展—多ゲノムの解読、遺伝子発見、ゲノム機能解析 2000年以降

ヒトゲノムの粗配列決定(2000年)、精密配列決定(2003年)を契機として、一気にポストゲノムの時代に入ったとの認識が高まりました。ただゲノム解析が終わったのではなく、むしろ始まったとの認識が大切でした。

2つあります。1つは多人数のゲノム解読・多生物のゲノム解読が必要との認識でその後10年あまりで非常に**多くのゲノム解読**が進んできております。もう1つはゲノム上のDNAの意味の解明—**ゲノム機能解析**が重要になって来ました。どの部分が重要な遺伝子であるかそれ以外の部分はどのような働きをしているかが研究課題として大きく浮かび上がってきました。そこには従来想像もつかなかった複雑な機構が隠されていることがようやく分かり始めてきました。

#### ① 多人数ゲノムの解析

国際ヒトゲノム計画は、ヒトゲノムの基本参照配列を決めることに大きな目標がありました。実際は人種も違う複数のヒトの部分ゲノムを継ぎ合わせた全体ゲノムの配列が一本分として提示されたのです。それを受けて、次の2つの動きがありました。

1つは、ある個人の全ゲノムを解読しようとするものです。真っ先に着手したのが、

**James Watson** と **Craig Venter** です。それぞれ別々に1億円以上かけて自分のゲノムを解読し公表しました。**Craig Venter** の場合は「ヒトゲノムを解読した男—自伝」(2008)に自分のゲノムと自分の能力・性格・体質などの関係について興味深いコメントを書いています。

その後個人で自分のゲノム全解読をした人は、**George Church, Stephen Quake, Michel Snyder, 富田勝**などがあり、氏名が公表されていない人は数百人に上ったと思われまます。個人ゲノムの解読サービスをする会社として **23andMe** が有名になりました。

もう1つは、プロジェクトとして、**1000人ゲノム、がんゲノム**が始まりました。これは多人数のゲノムを比較することによって新たな知見を得ようとするものです。

更に英国は、2010年から1万人ゲノムプロジェクトを推進しています。これまで各種コホート研究を進めてきた強みを活かすことが出来るとしています。

一方米国も2014年からこれまでの数多くのコホート研究をまとめて**100万人ゲノムプロジェクト**を始めました。いずれも大規模コンピュータ解析が必要で、以前には想像もつかなかった規模です。

## ② 発現解析     マイクロアレイ、DNAチップ (2000年前後)

ヒトゲノム解読が終了する前後から、マイクロアレイ・DNAチップ等を用いて、ゲノム上の各遺伝子の各細胞での具体的な発現状況を調べる研究が活発化しました。この研究は、個々の研究室レベルでできることで、世界中で広く研究されるようになっていきました。しかしヒトゲノム解読が終わった後からは、それらを系統的に調べようとするゲノムワイド解析が始まりました。その代表的なプロジェクトが次の2つでそれぞれ大きな成果を挙げて来ました。

### 1-4 ヒトゲノム機能解析プロジェクト     米ENCODEプロジェクトの発足と進展

米国は、**2003年10月**ヒトゲノム機能解析プロジェクトENCODEをスタートさせました。当初4年間でヒトゲノム上の1% (断片40カ所、合計30Mbp)を解明するパイロットプログラムとして国際的にも呼びかけをし、結果的に米英スほか10カ国が参加しましたが、日本は後述のFANTOM をベースにしてゲノムネットワークプロジェクトを始めることにしていたので参加を見送りました。

このプロジェクトは4年間での解析技術の成果（ChIP-Chip法他、コストダウン）を受けて、**2007年から**ヒトゲノム全領域を対象とした**本格プロジェクト**に発展しました。この時には、日本の独自技術(CAGE-Chip)がすでに世界の注目を集めており、日本は要請を受けてプロジェクトに参加しました。この時期、次世代シーケンサー（後述）が急速に発展し、ChIP-Seq, CAGE-Seq, RNA-Seq などの新しい解析技術が主流となりました。

この**ENCODEプロジェクト(2007-2012)**では、2つの新しい試みが始まりました。1つは**DCC**データコーディネーションセンターの併設です。ENCODEプロジェクト内の各グループが独自にデータを保管・解析するのではなく、DCC がデータ収集・整理・蓄積・流通を当初から全グループに公開し更に外部公開時期も早めようとするものです。このシステムはその後の他のライフサイエンスプロジェクトにも採用されることになりました。日本でも近年このシステムが採られていることは たいへん良いことだと思います。

もう1つは、**modENCODE** の発足です(2008-2012)。ヒトゲノムの機能解析を助けるためにモデル生物ゲノムの機能解析を並行して進めることになり、線虫とショウジョウバエがモデル生物として選ばれました（リーダー **Robert Waterstone**）。実験サイクルの短いこれらの生物の機能解析研究成果は、ヒト・マウスの機能解明に大いに役立ちました（Nature 2012）。

**ENCODEプロジェクト全体の成果は2012年に大きく報告されました(Nature 2012)**。主な成果は次のとおりです。

- ・クロマチンの役割の詳細－接近可能性、メチル化、進化保存性ほか
- ・遺伝子プロモータの長距離作用
- ・転写因子群と調節ネットワーク

特に、ヒトゲノム DNA の **80%以上が転写され機能を持つ**ことを示したことは大きな成果でした。(理研プレスリリース 2012年9月)

## 1-5 RNA新大陸の存在と遺伝子制御ネットワーク

### **FANTOM (Functional Annotation of Mammalian Genome) の成果と貢献**

日本が2004年から主導してきた FANTOM 国際コンソーシアム(リーダー理研**林崎良英**)は、当初の目標であったマウス cDNA の機能データベース構築の範囲を超えて、大きく発展し数々の成果を挙げて来ました。これまでに FANTOM-1 から FANTOM-5 まで、

毎回大目標を掲げ、研究と集会と成果論文発表を続けて来ました。毎回一週間前後続く集会（ジャンボリー）には世界から 100-200 名の代表研究者が集まり討議し、その後研究成果が Nature, Science, Cell 誌等に主論文と関係論文多数が発表されるというパターンを繰り返してきました。これまでの関係研究者は数千人を超え、この分野の世界最大の研究集団となっています。（Piero Carninci, 鈴木治和、河合純ほか）

その経過と成果は、理研の **FANTOM** ホームページ([fantom.gsc.riken.jp](http://fantom.gsc.riken.jp)) に詳しく載っていますが、その中でも次のような成果は注目に値するでしょう。

- ・各種細胞における遺伝子発現状況のデータベース提供（2002 年）
- ・RNA 新大陸の発見—多数の non-coding RNA が転写（2005 年）
- ・転写制御ネットワーク解析の重要性指摘・例示（2010 年）
- ・プロモーター・エンハンサーのデータベース提供（2015 年）

このような網羅的データベースの提供は、世界中でそれを利用している研究の成果が直接見えないことが多く、その影響の大きさを測ることはできません。しかし、京大山中グループが **iPS 細胞** 研究の初期(2003 年)に **FANTOM** データベースを見て多機能性細胞に戻すために加えるべき転写因子の候補を絞り込めたという事実は、ノーベル賞受賞に至る研究の一里塚としてたいへん重要なことで、日本としてもたいへん嬉しいことでした。

## 1-6.ゲノムから表現型へ (Genome to Phenome) の解析、オミックス解析

ゲノム配列解析に続く研究として、上記の ENCODE プロジェクト、FANTOM コンソーシアム以外に下記の 2 つが大きく浮上して来ました。GWAS とオミックス解析です。

### ① GWAS の成果

**GWAS**(Genome-Wide Association Study) は、遺伝子型と表現型を従来の部分的な解析（いわゆる連関解析や SNP 解析）から一気にゲノム全領域に広げたもので、2003 年に日本の理研グループが先鞭をつけて以降に世界的に広まり、その解析成果が次々と発表されて来ました。1 つの GWAS で、ある目的(例えば、糖尿病関連遺伝子群の同定—例. 2016 年 1 月前田・今村・門脇発表)のために、数千人～数万人のゲノムデータを集め、健常者群と患者群との間に明らかな有意差を示す遺伝子群を列挙するというステップを取ることで、従来と桁の違う数の関連遺伝子が見つかるのです。このような GWAS が各国で行われ、それらの成果が、次々と発表されてきました。最近の 10 年間で百数十報が Nature Genetics 等に載っています。非常に大きな成果を挙げてきたと言えるでしょう。ここで

は高度な統計解析手法が使われ、バイオインフォマティクスの一分野を形成しています。

## ② オミックス解析

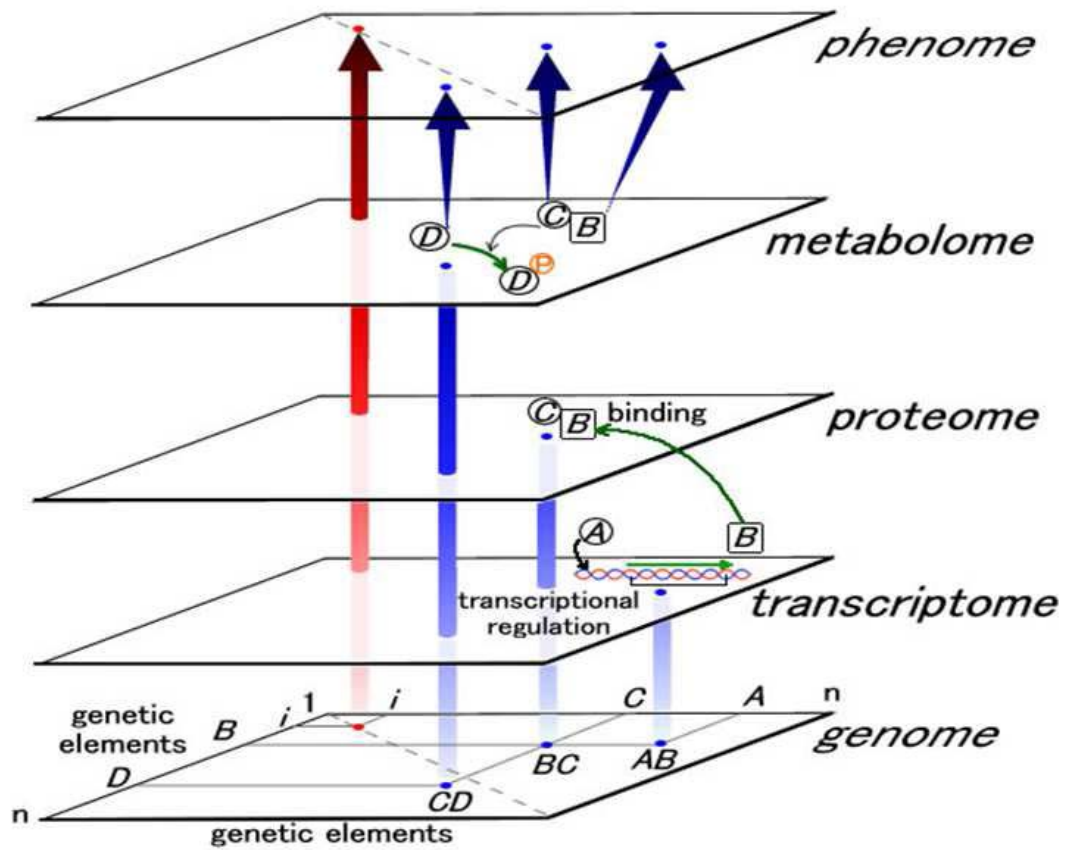
もう1つは、オミックス解析です。DNAからRNAに転写されたトランスクリプトーム解析、RNAからタンパク質に翻訳されたプロテオーム解析、更にはそれらによって代謝された産物のメタボローム解析と、最終表現型に至る途中の表現型(Intermediate Phenome)の解析が盛んに行なわれるようになって来ました。ここでは、DNAチップ・マイクロアレイ、質量分析計などが使われます。最近はRNA-Seq, CAGE-Seqなどの手法も活躍しています。このような中間表現型は、病気の早期診断のためのバイオマーカーとして有力なものとして活発な研究が進みました。更にこれらの過程で、遺伝子発現機構・タンパク質生成機構の研究が進み、**non-coding RNA**を中心とする**RNAの役割**の理解、**エピジェネティクス**や**オルタネーティブスプライシング機構**の理解が、急速に進んで来ました。2000年初頭までのゲノム理解が最近10年間に大幅に書き換えられたと言えます。(1-11参照)

平面解析から立体解析へ      オミックス空間解析、トランスオミックス解析

このオミックス解析について、重要な概念が2002年頃に理研G S Cの和田所長によって提唱され、その具体的図示化が豊田リーダーによってなされました。各オミックス平面の詳細な解析から更に進んで、各平面を縦に貫通するオミックス空間解析が重要というものです。2012年以降に東大黒田真也が **Trans Omics Analysis** という言葉で上記の概念を具体的に発展させ始めています。

オミックススペース解析

(トランスオミックス)



理研ホームページ及び東大黒田資料より

### 1-7. 植物分野でのゲノム総合科学

長い植物研究の中に、10 数年前に始まった「アラビドプシス研究国際コンソーシア

ム」MASC は、大きな一石を投じました。多様な植物の中からモデル植物として「シロイヌナズナ-Arabidopsis」を取り上げ、**MASC (Multinational Arabidopsis Steering Committee)**の下に国際共同研究体制を作り上げたことは、特筆に値します（主要メンバー約 30 名、リーダー理研篠崎一雄、研究者全世界 9000 研究室、23000 研究者）。

ちょうどゲノム解読の時代を迎え、植物の基準データベースを作り上げることに成功しました。モデル植物としてのアラビドプシスのデータベースは **TAIR** として公表され、イネほかの作物研究に役立っているようですし、開花・開茎に関係する分子群の発見や細胞分裂・細胞輸送のメカニズムの解明や更には植物の酸素センサーの発見など科学的な進歩もまとめられています。この活動の詳しい報告は、**MASC Annual Report** として 2002 年から毎年刊行されています。また塚谷裕一著「[変わる植物学](#)」（東大出版会 2006）にも経緯が詳しく載っています。

尚、植物研究の中で、他の動物モデル（線虫・ショウジョウバエなど）と並行して RNAi や siRNA の研究や、ゲノム編集技術 CRISPR/cas9 の応用の進展に相乗効果を発揮していることも見逃せないことでしょう。

これらの研究には、植物の環境ストレス耐性の研究で多くの研究成果を挙げてきた篠崎一雄グループや、植物メタボロミックスから植物生理活性物質の研究 Phytochemical Genomics をリードしている斎藤和季や、植物成長に関わる植物ホルモンオーキシンほかの研究で世界に貢献してきた神谷勇治など多くの活躍があり、最近アジア地区との共同研究でも関開明ほか日本は指導的な役割を演じてきています。

日本の誇るイネ研究でも、イネゲノムプロジェクトで世界をリードしてきました。1990 年代初頭から始めたイネゲノム解読計画（～1999）は、その後更にイネ遺伝子機能解析に進み、2005 年にはその成果を **Nature** 誌に発表しております。常にヒトゲノム計画と並行して一歩先を歩いてきたと言えましょう。最近の中国の圧倒的なゲノム解析センター **BGI** の動きで量的には太刀打ちできませんが、品種改良面での研究蓄積は今後とも世界に貢献できる領域でしょう。

この間、植物特有の葉緑体のゲノム解析に取り組んでこられた杉浦昌弘の貢献も見逃すわけには行きません。今後の光合成研究、動物のミトコンドリア研究に影響を与えるでしょう。

---

MASC Arabidopsis Annual Reports    TAIR

K.Saito “Phytochemical Genomics – a new trend “ Current Opinion in Plant Biology 2013

M.Sugiura; Plant Biology 2014, PNAS 2013 ほか

---

## 1-8. 微生物ゲノムーメタゲノム解析

微生物はゲノムサイズが小さいので、早くからゲノム解析が進められていました。



1995年にTIGR社が世界初めての生物ゲノム解読結果を発表したことは、上述のとおりですが、その後も非常に多くの種類の微生物ゲノム(大腸菌ほか)が解読されました。

さらに、その後のシーケンサーの急速な進歩と呼応して、メタゲノム解析(集団ゲノム解析)が行われるようになってきました。1999年サンジェゴ海洋研究所が中心になって世界の主要地域の海洋微生物群を網羅的に解読するプロジェクトがスタートしました。

その後に、腸内細菌群の重要性に着眼し、**Human Metagenome Project**がスタートしました。それまで培養できる細菌のみを研究していた世界に、体内に常在するすべての細菌群のゲノムデータが提供されるようになったのです。それまで50-100種類位の腸内細菌を研究していたところに、いきなり1000種を超える細菌のデータが提供され始めました。健康・疾患に関する大きな影響が注目されています。

ここでも大きなパラダイムシフトが起こりました。

このことについても数々の総説が出ています。

---

Metagenomics: An Application Based Perspective by Y.Bashir Chinese J.Biology 2014  
A Bioinformatician's Guide to Metagenomics by V.Kunin Microbiol.Mol.Bio. 2008

---

一方、今回の2015年ノーベル賞受賞の大村智が、その研究の過程で、抗生物質生産菌放線菌 *S.avemectinius* の全ゲノムの情報解析から、この放線菌が avermectin の他に37種の有機化合物を作る遺伝子を持つことを明らかにしています。

Nature Biotechnology 21, 526 (2003)

Satoshi Ōmura, et al, Yoshiyuki Sakaki, and Masahira Hattori;

Genome sequence of an industrial microorganism *Streptomyces avermitilis*: Deducing the ability of producing secondary metabolites

PNAS 98 (21) 12215-12220 Oct.9 2001

また微生物の遺伝子改変で新しい生理活性物質を創る技術が進行しています。

Jay Keasling UCBK, LBNL, BEI 所長 Metabolic Engineering, Synthetic  
Biology Antimalarial drug artemisinin engineered to produce it by E-coli &  
Yeast May 30, 2006 Science

## 1-9 次世代シーケンサーの登場とそのインパクト

2002年秋にボストンで行われた「1000ドルゲノム」旗揚げのワークショップは、最近のゲノムシーケンサーの革命的な変化の先駆けでありました。**Craig Venter** ほか野心的なメンバーがヒトゲノムの解読を1000ドル程度で出来るDNAシーケンサーを開発しようと呼びかけました。当時数千億円かけて一人分のヒトゲノムの精密解読が間もなく完了すると期待されていた時期で、この提案は遊び心半分のワークショップでした。しかしこの高い目標設定は多くの研究者や企業の関心を引き、具体的な開発が始まりました。それを受けて2004年にNIHは「1000ドルゲノム」シーケンサーの開発のためのグラントを出し始めました。短期目標(数年内に数万ドルで)と長期目標(10年後に1000ドルで)を明示し、以後毎年募集を続けました。その影響もあり2007年頃から高速・廉価なシーケンサーが次々と市場に出始めました。

筆者は、毎年初にフロリダで開かれるAGBT (Advanced Genome Biology and Technology)の会に隔年に出ておりましたが、2006-2008年頃の会ではこのシーケンサーの話題が各参加者から発表され熱気に溢れていました。そこでは、シーケンサーの性能・コストだけでなく、その応用分野のこと、ソフトウェアのこと、膨大なデータ(ビッグデータ)の扱いのことなどが熱く語られていました。測定したDNA配列生データは、収集・解析後、破棄するとの考えまで示され驚いたものでした。[\(参加報告書 2007年3月\)](#) 三原・竹田氏と共にAGBT参加 p.8, p.18参照

2007年12月岩波「科学」に、林崎・五條堀と八尾の共著で、「次世代シーケンサーは生命科学研究に革命をもたらす」という刺激的な題の対談記録と総説を出しました。その後たしかにこの影響は衝撃的なものが続いています。

2010年のNature誌には、DNAシーケンシングコストが7年間で14000分の1まで下がったと報告され(図)、その後も更に下がり続け、最近ついに1000ドルゲノムを達成しました。この間に、これらのシーケンサーはDNAだけでなくRNAも読めるようになって来ましたので、オミックス解析にも使えるということで、適用範囲が非常に広まり、ゲノム解析は以前のような大センター集中型だけでなく、各研究室・各病院で行われるようになって来ました。個人ゲノム診断・治療(Pharmacogenomics)の時代を迎えようとしております。

J.Black et al; "The Impact of NGS on Haematological Research" Pathogenesis 2, 9, 2015

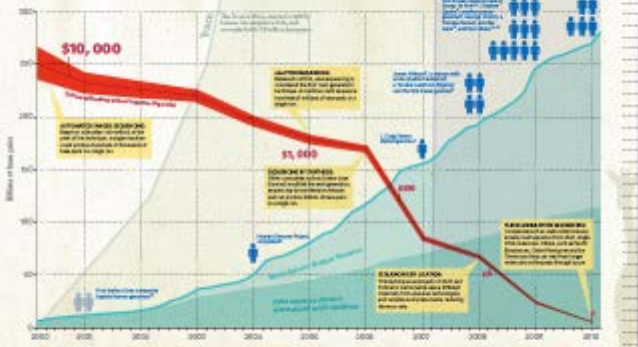


# THE SEQUENCE EXPLOSION

A few days after the announcement of the draft of the human genome in 2001, there was a realization: there would be a revolution in the way we do science. The Human Genome Project had just begun to sequence the 3 billion base pairs of DNA that make up the human genome. The technology to do this was rapidly improving, and the cost was falling. In just a few years, the cost of sequencing a human genome would drop from \$100 million to \$10,000. This would allow scientists to study the genomes of many other species, and to look for variations in the human genome that are associated with disease. The Human Genome Project was just the beginning of a new era in genetics.

## DNA SEQUENCES BY TAXONOMY

Approximately 100,000 species are known to exist on Earth. The number of DNA sequences generated for each species is increasing rapidly. This is due to the fact that the cost of sequencing DNA is falling, and the number of species being sequenced is increasing.



## HOW MANY HUMAN GENOMES?

The number of human genomes sequenced is increasing rapidly. This is due to the fact that the cost of sequencing a human genome is falling, and the number of people being sequenced is increasing. As of 2010, approximately 100 human genomes have been sequenced.

- 1. The first human genome was sequenced in 2001.
- 2. The first mouse genome was sequenced in 2002.
- 3. The first Arabidopsis genome was sequenced in 2004.
- 4. The first chicken genome was sequenced in 2004.
- 5. The first dog genome was sequenced in 2005.
- 6. The first pig genome was sequenced in 2007.
- 7. The first cow genome was sequenced in 2009.
- 8. The first rice genome was sequenced in 2005.
- 9. The first yeast genome was sequenced in 1996.
- 10. The first fruit fly genome was sequenced in 2003.

Sequencing Cost \$10K/Mb (2000) – \$1/Mb (2010)

Nature 464, 670-671, Apr. 1st, 2010

## 1-10. 生命科学研究を支えるデータベース群の発展、活動

生命科学関連の基幹データベースとしては、1980年以前から次の3つがありました。

|                       |         |   |               |
|-----------------------|---------|---|---------------|
| DNA塩基配列データベース GenBank | 1970年創設 | → | 1995年 400Mbp  |
| タンパク質アミノ酸配列データベース PIR | 1975年創設 | → | 2000年 150KSeq |
| タンパク質立体構造データベース PDB   | 1980年創設 | → | 2005年 35千件    |

これらは分子生物学研究の基礎的必須な手段として発展してきました。

ゲノム時代に入ってデータ量が急増し、2015年には GenBank が 200Gbp, PDB が 10万件の規模にまで、急増してきました。

この間、多様なデータベースが構築されてきて、**NAR (Nucleic Acid Research)** 誌が2000年から**データベース特集号**を毎年刊行するようになり、分子生物学関連データベースをまとめた総覧となっています。2015年版によれば、数百種類のデータベースが8つのカテゴリー（核酸、タンパク質、代謝・シグナル伝達パスウェイ、ウイルス・微生物、ヒト・モデル生物、変異と疾病・医薬、植物など）に分類されています。

さらに最近は、エピジェネティクス関連のデータベースと解析ツールも提供され始めました。Epigenie team が、NCBI, Ensemble, WashU などと連携して総合的なサイトを提供しています。

尚、これまでのデータベースの解析技術については、下記3件が参考になります。

**DATABASE The Journal of Biological Databases Curation**

「バイオデータベースとソフトウェア最前線」特集 実験医学増刊 26-7 (羊土社)

NBDC (日本バイオデータベースセンター) ゲノム解析ツールリンク集

このように生命科学関連のデータベースは多様化し、それぞれ急速に増加しており、生命科学研究を支える必須技術となっています。それらのデータベースをきちんと管理して行くには下記のような縁の下の仕事が必要です。

データベース管理 (受付、検証、登録、蓄積、構築、連携、保守、利用体制、日  
常サービス、普及・広報・教育)

更にそれらが活用されるには解析ソフトウェアが整備されている必要があります。

(配列解析、構造解析、機能解析、画像解析、ネットワーク解析、分子シミュレーション、システムシミュレーション、マルチスケールモデリングなど)

参照サイト

CBI 研究会—学会 ホームページ

日本バイオインフォマティクス学会

バイオインフォマティクスに関する学問・技術・事業・人材・研究会

データベース、解析技術、スパコン、ネットワーク解析

このような生命科学関連データベースの健全な維持・発展のためには、次のような施策が必要です。世界的に協調して支援していく必要があります。

1. データベースセンターの存在（継続性） NCBI, EBI, PDB など
2. データベースの標準化・関連技術開発  
オントロジーの整備、解析ソフトウェアの開発、ウェブ対応（今後重要）
3. 人材の育成 学科、センターほか

尚、新たなデータベース対応の雑誌や総説も発刊されています。

新雑誌刊行 Scientific Data—Nature 2014 から

GigaScience –GBI/Springer 2009 から（ビッグデータ対応）

バイオインフォマティクスの 15 年を総括している次の論文も注目に値します。

総説 Rise and Demise of Bioinformatics ? Promise and Progress

Christos A. Ouzounis EBI PLoS Apr.26, 2012

#### Abstract

The field of bioinformatics and computational biology has gone through a number of transformations during the past 15 years, establishing itself as a key component of new biology. This spectacular growth has been challenged by a number of disruptive changes in science and technology. Despite the apparent fatigue of the linguistic use of the term itself, bioinformatics has grown perhaps to a point beyond recognition. We explore both historical aspects and future trends and argue that as the field expands, key questions remain unanswered and acquire new meaning while at the same time the range of applications is widening to cover an ever increasing number of biological disciplines. These trends appear to be pointing to a redefinition of certain objectives, milestones, and possibly the field itself.

## 1-11 ポストヒトゲノムの10数年間のインパクト

以上、ヒトゲノム解読完了(2000,2003年)前後からの10数年間の経過を述べましたが、この間、ポストヒトゲノムの様々なインパクトについて多くの論説が出ています。以下に代表的なものを記します。

榭佳之著 ゲノムサイエンス 講談社 2007

### 第2部 ポストシーケンスの時代

疾患遺伝子を追いつめる、転写制御の秘密に迫る、ヒトの進化をたどる

F.Collins; “Has the Revolution Arrived?” Nature Apr. 1<sup>st</sup>, 2010

様々な視点 シーケンスコスト 1999—2009 までに 14000 分の 1 に低下

個人ゲノム情報が手軽に得られる 2010 年すでに 13 人

哺乳類ゲノム 2010 年 14 種完了

大規模比較ゲノムが進んでいる。

遺伝子診断が始まっている。SNP, Haplo Type

遺伝と環境の疾病影響の解析が進んでいる。GWAS Cohort

Study

ゲノム機能解析プロジェクトをスタートさせた 2003--。

エピゲノミクス研究に突入した 2008--。

1000 人ゲノムプロジェクトスタート

---

Eric Lander; Initial Impact of the Sequencing of the Human Genome

Nature Feb.10, 2011

Eric Green; Charting a Course for Genomic Medicine from Base Pairs to Bedside

Nature Feb.10, 2011 別添図—10年間のゲノム[関連成果年表](#)

---

John Mattick The Central Role of Regulatory RNA Jan.2016 at RIKEN

あらためて、non-protein-coding DNA の重要性を強調し、数万~数十万に及ぶ short and long non-coding RNA が発現・翻訳・エピゲネティクにおいて、様々な未知の制御機能を持つ可能性が高いことを警告しております。

“The Rise of Regulatory RNA” Nature Reviews Genetics Apr.29, 2014

---

## 2. タンパク質構造ゲノミクス・構造生物学の進展と情報科学

### 2-1. タンパク質立体構造解析の進展と成果

#### ① 立体構造解析技術と立体構造データベース

タンパク質が生体内で機能を発揮するには、そのアミノ酸配列特有の立体構造をとる必要があります、立体構造解析の重要性は早くから言われてきました。しかし立体構造を解析するには十分な試料と高度な機器と高度な解析ソフトウェアが必要で、非常に専門的な領域となっています。これまで主な手法は、X線結晶構造解析、NMR溶液構造解析法および極低温電子顕微鏡でありました。それらの成果である構造情報を登録する制度として1971年からPDBが発足し、1998年と2003年の大幅な体制変革によって、今では世界のデータベースとして wwPDB が米欧日3極で支えられています。このタンパク質立体構造データベースの登録件数は最近(2015年) 10万件を超えました。

このPDBには実測データのみが登録され、コンピュータによるモデリング構造は含まれません。大まかには、X線結晶解析によるもの90%弱、NMRによるもの10%弱、その他が数%です。

1990年代前半までの時代には、日本はタンパク質構造解析の分野で世界から大きく取り残されていました。

#### ② 構造解析プロジェクト – 日本 及び 米国 (2000年以降)

2000年以降のゲノム科学総合研究センターの発足を契機として、日本はタンパク質構造解析分野で世界をリードする立場に躍り出ました。すなわち、理研横浜のGSCに設置されたNMR群と理研播磨にある放射光Spring-8 を駆使した構造解析プロジェクト「タンパク3000」が2002年から始まり数多くの成果を挙げました。米国はその後を追って2003年からPSI(Protein Structure Initiative)をスタートさせました。更に日本は、先の成果を踏まえて2007年から重要タンパク質の構造を解くための「ターゲットタンパク質」プロジェクトをオールジャパンでスタートさせ次々と成果を挙げて来ました。米国はやはり1年遅れで、PSI-2 をスタートさせました。

またこの間藤吉好則率いる極低温電子顕微鏡(Cryo-EM)で世界をリードし続け、アクア



ポリン水チャンネル構造ほか数々の成果を挙げて来ました。

日本が構造解析分野で世界の中で後塵を拝していた20年位前のことを考えると正に感慨深いものがあります。

これらの構造解析の過程では、いずれも測定値から立体構造までもっていく理論計算ソフトウェアが決定的に重要で、X線解析ではCCPソフトウェアパッケージ、NMRではDGソフト、最近のCryo-EMではELUDAソフトの進歩がキーポイントになっています。

### ③ 立体構造解析による作用機序の解明－生命現象解明への貢献

タンパク質の立体構造解析は、1990年代以降、個々の生命現象の詳細な解明に非常に重要な役割を果たし続けてきました。例えば、転写・翻訳の分子メカニズムの解明、シグナル伝達機構の解明などです。これらの成果は、主要雑誌で紹介されるだけでなく、ノーベル賞を与えられるレベルでした。そのいくつかを挙げて見ましょう。

① **ATP合成酵素F1サブユニット**の立体構造決定による反応素過程の解明の功績によってBoyer とWalker らは1997年のノーベル化学賞を受賞しました。

② **リボソーム大サブユニット**の詳細な構造および翻訳プロセスの解明によってV.Ramakrishnan, A.Yonathらは2009年のノーベル化学賞を受賞しました。

③ **Gリボソーム大サブユニット**の詳細な構造および翻訳プロセスの解明 によって、R.Lefkovitz, B.Koblika らは2012年のノーベル化学賞を受賞しました。

そのほかにも、**バクテリオロドプシン**立体構造の決定が光駆動プロセスの解明に、また**チトクロームC酸化酵素**の立体構造決定がプロトンポンプ機構の解明に役立ってきました。また、Tom Steitz らは、**哺乳類タンパク質合成プロセス**の解明を(Nature Aug.15, 2013)、横山茂樹・門脇孝らは、**糖尿病関連のアディポネクチン受容体**の構造・機能の解明 (Nature 24.Sep, 2014) を、それぞれ立体構造解析によって致しました。

## 2-2 構造バイオインフォマティクスの進歩 データベース、分子計算、医薬設計

### ① 立体構造データベースとその解析

このような流れの中で、**Structural Bioinformatics** という分野は地味ながら多面



的な活動を続けてきております。まずその共通基盤として上述の通り、タンパク質立体構造データベース**PDB** があります。生命科学研究を支える大きな存在です。世界の**wwPDB**は、米国の**RCSB/PDB(Helen Berman)** と日本の**PDBj(中村春木)** 及び欧州の**PDBe** で構成されており、その内容がホームページに載っています。また、このような構造基盤データを分析して、構造分類のデータベース**SCOP** (**Structure Classification of Proteins**) のような有用なデータベースが開発されて来ました。

その上で次の3つの研究が進展しています。

## ② 立体構造予測、モデリング

アミノ酸配列から立体構造を予測することはこの分野の長年の悲願であり、その技術を競う**CASP (Critical Assessment of Protein Structure Prediction Methods)** という国際コンテストが1995年以来隔年開かれてきました。英国サンガーセンターの**Tim Hubbard** (以前日本の蛋白工学研究所ポスドク) 及び米国CARBの**John Moult** (米国の蛋白工学研究所研究員) らが共同で1994年に始めた方式で、インターネット上に提示されたアミノ酸配列に対しそのタンパク質の立体構造を予測し、その成果を競うものです。

その中でも特記的なことは、2005年にワシントン大の**David Baker** のグループが出した**フラグメント法による予測**が最優秀の成績を収め、その後も首位を守り続けています。**Baker** グループは、この方法を**人工タンパク質設計**に用いて次々と成果を挙げています (Koga, Baker 2015) ,(Brunette, Baker 2015), (Doyle, Baker 2015) 。今後の合成バイオロジーの展開に大きな影響をもたらすと期待されます。

これとは別に**ホモロジーモデリング法**が実際的に多く使われています。当該タンパク質と配列ホモロジーが高い(30%以上)タンパク質の立体構造が既知であれば、その構造を鋳型として推定構造を作り上げるという手法で、登録データが増えるにしたがってこの方法でモデリングできるタンパク質が増えてきております。この方法のコンピュータソフトウェアも多く出回っており、医薬開発等の研究現場で日常的に使われるようになってきました。

---

## ◎ インターネット上で技術開発を競い合う方法

このCASP のようにインターネット上で技術開発を競い合う方法は、正にバイオインフォマティクス分野の研究開発に適した方法で、その後類似のコンテストが多く行われるようになってきました。最近の例 (IPAB研究会主催—関嶋征和)では、インターネット上に立体構造が提示された酵素 (ヒトc-Yes Kinase) の阻害剤候補を240万化合物から選ぶ手法のコンテストが行われた。11グループが参加し様々な方法で各上位20候補を提示し、それらの阻害活性測定を事務局側で外注して結果を競いました。その結果は近く発表されますが、このような進め方は正にオープンイノベーションの典型で、**短期間に衆知を集める方法**として世界各国で行われています(例. IBM DREAM チャレンジ、UCSF CAGI-Critical Assesment of Genome Interpretation コンテスト)。

-----

### ③ 分子シミュレーション – 巨大分子・マルチスケールシミュレーションへ

タンパク質分子動力学シミュレーション**MD (Molecular Dynamics)** が、タンパク質の詳細な動きを見るためにX線結晶構造だけでは分からない作用機序が理解できます。この方法は最近のコンピュータ能力の進歩により大きく複雑な系に対しても使われるようになってきました。Matin Karplus, Michael Levitt らはこの原理と方法の研究でノーベル化学賞 2013を受賞しています。

最近、**光合成関連の膜タンパク質**のような大きなタンパク質分子の解析も行われています(K.Ogata; JACS Oct.23,2013)。更に**細胞内分子混雑環境**でのタンパク質・核酸・代謝物・溶媒を含んだ全原子レベルの**世界最大級の生体分子シミュレーション**が、スパコン「京」用に開発されたソフトウェア**GENESIS(Generalized Ensemble Simulation System)**を用いて達成されました(Yuji Sugita; PNAS Jul. 2015、理研ニュース2015年10月号杉田勇治)。同様の指向の細胞モデリングは米国Georgia 大のJef. Skolnick, Ed.Chowらによって進められており、その成果が最近発表されました。(Large-Scale Modeling of Intracellular Macro Molecular Dynamics PNAS Nov.2015)

更に粗視化モデルと組み合わせて大きな系を解析するようになってきて、**シグナル伝達経路上のリン酸化酵素複合体の構造の動的モデリング**や、**クロマチン構造とエピジェネティクス制御の動態解析**に成功しています。(Shoji Takada, Accounts, Chem. Research 2015)。

もっと大きな器官や臓器の長い時間変化を扱うには、**Multi-Scale Simulation** が使わ

れるようになってきました。東大の久田俊明・杉浦清了・岡田純一らが開発してきた世界最大の心臓シミュレーションモデル“UT Heart”も「京」を使った、マルチスケールシミュレーションです。このように分子レベルシミュレーションから臓器・個体レベルのシミュレーションまでをつなぐマルチスケールシミュレーションは世界的な方向で、現在米国NIHでも多研究機関共同でMSM (Multiscale Modeling)コンソーシアムが走っています。

特に日本では早くからタンパク質分子シミュレーションの分野で多くの人材が育ってきており(中村,福西,木寺,池口,岡本,杉田,高田,河野,白井,緒方ほか)、最近のスーパーコンピュータ「京」の牽引力もあり、ますます活発化してきております。

Intl. Workshop –SCLS (Super Computational Life Science) in Tokyo, Oct.2015

Structural Biological Dynamics JST Intl. Symposium in Tokyo, Nov.2015

BioSuperComputing Newsletter 2014—2015

#### ④ 構造ベース医薬分子設計(Structure based Drug Design)

—CADD Computer Aided Drug Design—

阻害あるいは活性化対象のタンパク質立体構造に対し医薬候補がどのように作用するかを解析し、効果の高い化合物を探し出す方法です。SBDD(Structure Based Drug Design) という手法が1990年頃から始まりましたが、この方法は最近15年で大きな進歩を遂げておりFBDD (Fragment Based Drug Design) やLBDD (Ligand Based Drug Design)法も含めて、すでに医薬開発の現場で日常的に使われるレベルまで一般化しました。立体構造が解かれたタンパク質数が増えてきたのと ホモロジーモデリング手法の発達に依ることです。

2015年11月に理研横浜研究所で開かれたスウェーデンとの合同シンポジウムでもこの分野の先端的研究状況が紹介されました。(本間、広川 ほか)

Joint Symposium Structural Biology for Drug Discovery RIKEN-KLI 2015

### 2-3. 構造解析技術の最近の大きな変化

構造解析技術の分野の最近の大きな変化は、次の3つでしょう。

① 1つは、**X線自由電子レーザー(XFEL)法**によるタンパク質構造の動的状態測定法の開発です。ハードウェアの発達だけでなく、そのデータの解析ソフトの優劣が勝負になるようになっています。Stanford大、RIKEN 等が競っています。

② 2つ目は、新たな極低温電子顕微鏡 **Cryo-EM** の出現です。ウイルス・リボゾーム (~80S) ・ガラクトシダーゼ・イオンチャンネルのような大きなタンパク質の立体構造が、3.5~2.2Åの精度まで次々と解かれ始めています。このブレークスルーをもたらしたものは検出器の革新と解析ソフトウェアの進歩です。

Ewen Callaway; “The Revolution will not be crystallized”

Cryo-Electron Microscopy is kicking up a storm in structural Biology.

Nature 10. Sept 2015

Ray Wang, Yifan Cheng, David Baker et al. “De novo protein structure

Determination from near-atomic resolution Cryo-EM maps”

Nature Methods April, 2015

Frank DiMaio, Yifan Cheng, David Baker et al; “Atomic-accuracy models

from 4.5Å Cryo-Electron Microscopy with density-guided iterative

local refinement”

Nature Methods April, 2015

Yifan Cheng et al; “A Primer to Single-Particle Cryo-Electron

Microscopy”

Cell April 23, 2015

Yifan Cheng; “Single-Particle Cryo-EM at Crystallographic Resolution”

Cell April 23, 2015

X. Bai, G.McMullan and S.Sceres “How Cryo-EM is revolutionizing

Structural Biology” s in Biochem Sciences” Jan. 2015

③ 3つ目は、立体構造**モデリングアンサンブル法**の浮上です。複数の方法(X線、NMR, モデリングほか)を併用して、立体構造を提示するもので、2010年頃から UCSF の

A.Sali が提唱したものです。精度の検証は難しいのですが、生命科学研究者には大きな支援になります。

以上、いずれも**解析ソフトウェアの進歩**によるものです。これらは今後の発展を見守るべきでしょう。これら以外に下記2件にも注目したいと思います。

#### ④ 構造ダイナミクスの解析

2015年11月5日～6日東京で開かれた、JST-CREST/PREST 国際シンポジウムでは、“**Biological Structural Dynamics**” というテーマで、内外の指導的・先端的研究者が集まって、この分野の最近の進歩について発表・討議が行われました。

特に、次のような新しい技術開発・応用について注目すべき発表がありました。

- ・ スクリプス研の **Ian Wilson** が、長年のインフルエンザ菌の構造解析から近く汎用ワクチンの開発が可能になりつつあること
- ・ ブロード研の **Feng Zhang** が、ゲノム編集技術 CRISPER/Cas-9 のための新しい酵素群を開発中であること、
- ・ UCSF の **Yifan Cheng** が、最近の Cryo-EM により、TRP イオンチャンネルの構造と機能の解析に成功したこと
- ・ 分子研の **Nobuyasu Koga** が、D.Baker らと共同で、タンパク質構造設計( $\alpha\beta$  型、 $\beta\beta$  型他)に成功し、構造形成原理の解明に役立っていること 等。

#### ⑤ 非天然型アミノ酸、新規塩基対による人工タンパク質の創製

天然のアミノ酸 20 種類、塩基対 2 種類を拡張して、それらを基に新しいタンパク質を創りだそうという研究が進んでいます。これらの近況については、理研のホームページをご覧ください。理研ニュース 2015 年 12 月号 非天然アミノ酸 坂本健作  
理研ニュース 2013 年 8 月号 人工 DNA(第三の塩基対) 平尾一郎

### 3. システムバイオロジー、合成バイオロジーの進展

#### 3-1 システムバイオロジーの展開 15年

システムバイオロジー ([生命現象をシステムとして捉えることを目的](#)) は、1998年 ERATO 北野共生システムプロジェクト立ち上げに際し北野宏明によって世界的に初めて提唱されました。ちょうどヒトゲノムプロジェクトの完了直前の時期であり、網羅的に解明されつつある多数の遺伝子の個々の機能解析と共に、それら遺伝子・蛋白質など生体分子間のつながり・ネットワークの解明が重要になるとの認識が始まる時期でありました。

#### 3-2 システムバイオロジー元年(2000年)

2000年9月、世界初の ICSB (International Conference Systems Biology) が、北野の提唱で JST/ERATO 支援の下、東京で開催されました。この記念すべき第1回以降、毎年 米国西岸／東岸・欧州・豪州・日本他 世界各地で開催が続いています。第10回は2009年スタンフォード大で開かれ、北野はその席上この分野の推進功労者として顕彰されました。2015年12月には第16回がシンガポールで開催されました。HP 参照

この2000年は次の3点でその後のシステムバイオロジーの展開に重要な年でありました。正に「システムバイオロジー元年」と言うべきでしょう。

① 日本東京で第1回システムバイオロジー国際会議が開催されました。

ERATO 北野宏明リーダーが提唱、システムバイオロジーの理念・進め方・課題を共有。

② 米国 NIH が、システムバイオロジーを強力に推進し始めました。

NIGMS の M.Cassman 所長が、「次はシステムバイオロジーの時代」と推進策を旗揚げ。

③ 米国にシステムバイオロジー民間研究所(ISB)が設立されました。

L.Hood 所長が、生物学と計測・情報の三者融合体制によるシステム理解・応用を強調。

これ以前に日本は、金久實の代謝パスウェイデータベース KEGG の公開(1995年から)、ERATO 北野共生システムプロジェクト(1998年発足) 及び、富田勝の細胞シミュレーションモデル E-cell の構築(1999年から)を始めており、世界をリードしてきていました。そして2000年以降の米国のシステムバイオロジー促進策や、2003年以降の欧州 EU やドイツのシステムバイオロジー施策により、急速に活発化してきました。2002年に Nature 誌、Science 誌がそれぞれシステムバイオロジー特集号を刊行したこともその動きを促進しました。

その後の世界的な展開について、下記をご覧ください。

[システムバイオロジー内外動向 2011](#)

[システムバイオロジーと創薬技術 2008, 2014](#)

### 3-3 米国のシステムバイオロジー展開

#### ①システムバイオロジープロジェクト(2000年～)

米国 NIH 主導の 2000 年開始のシステムバイオロジープロジェクトの最初のものは AFCS (細胞シグナル伝達システム解明アライアンス) で、全米の数十の関係研究機関を巻き込んだ年間数十億円規模の大きなものでありました。これは 7 年間の分散研究であり、対象細胞の標準化が十分ではなくて成果については賛否が分かれましたが、米国は 2001 年以降も次々と新しいテーマで大型・中型プロジェクトを発足させ、システムバイオロジー研究を強力で推進し続けました。その結果、実験と計算の両面が共同して生命システムの解明を進めていくシステムバイオロジーのアプローチが多くの機関に浸透して行き、人材も多く育っていったのです。

この間の注目すべき事項のみを以下に列挙します。

#### ②米国国立システムバイオロジーセンター設立 (2007年～)

2000 年以降～2006 年までに、NIH は、システムバイオロジープロジェクトの大型・中型グラントを出し続けました。その間に、2004 年米国はシステムバイオロジー調査団を結成し、米国内だけでなく、欧州および日本の関係機関を訪問し、総合的なシステムバイオロジーの実態調査をし、総合報告会および報告書を公表しました(米国システムバイオロジー調査団 2004)。そしてその後の施策は分散研究から集中研究へと移行しました。

2007 年以降、米国は国立システムバイオロジーセンターを逐年設立し始め、2015 年現在 全米に 15 センターが出来ています。

#### ③ハーバード大システムバイオロジー学部設立 (2003年)

2003 年ハーバード大学に世界初のシステムバイオロジー学部が発足 (学部長は、免疫研究で著名な Mark.Kirshner 教授) し、新しいタイプの研究方式と人材育成に乗り出しました。この動きは世界中に大きな影響を与え、このあと各国にシステムバイオロジー専門の学科・学部・センターが数多くできてきました。

#### ④民間システムバイオロジー研究所 ISB 設立 (2000年)

2000 年にシアトルに、民間研究機関 ISB(Institute for Systems Biology)が発足 (所長は、DNA 解析・合成等の機器開発で著名な Leroy Hood 教授—のちに京都賞受賞) し、生物学者・情報科学者・先端機器開発技術者の 3 者融合の体制をとりました。その後数々の研究成果を挙げ、また多くの人材を世界に輩出してきています。

#### ⑤がんシステムバイオロジーセンター設立(2008年～)

L.Hood 博士は、2008 年米国がん学会(AACR) 総会で基調講演をし、システムバイオロジーの有用性と、P4 Medicine (Personalized, Predictive, Preventive and Participatory) の方向性を示し、がん研究関係者に多大な影響を与えました。その前後から、米国では Cancer Systems Biology センターが次々と設立され、現在 10 センターになっています。

#### ⑥システムバイオロジーベンチャー企業 (2000年～)

米国では、2000年以降、システムバイオロジーベンチャー企業が設立され、それぞれ特徴を持った技術で、大手医薬会社等にサービスを提供しています(例、**ENTELOS**社—肥満病・糖尿病システムモデル—糖尿病学協会 AADB と共同開発)。大手医薬会社はシステムバイオロジーの有用性に注目し、自社内に専門組織を持つ以外に外部のベンチャー会社を利用しています。

#### ⑦Systems Medicine, Personalized Medicine, Precision Medicine へ

米国はシステムバイオロジーの健康医療分野への応用を強く意識し、Systems Medicine, Personalized Medicine,そして Precision Medicine と、ますます総合的な視野に基づく医療の実現に向けて展開しています。

### 3-4 欧州システムバイオロジーの展開

#### ①ドイツのシステムバイオロジー(2003年～)

ドイツが2003年に肝臓細胞のシステムバイオロジープロジェクト **HepatoSys** を立ち上げたことは大きく注目されました。数多い候補課題の中から特定対象に絞ったことが珍しいことでした。3年間のパイロットプロジェクトのあとの本格プロジェクトで数々の成果を挙げ(薬物代謝メカニズム、細胞再生プロセスなど)、そのあと2011年からはその応用を目指すために **Virtual Liver Network** を展開しています。

ドイツはその間各種のシステムバイオロジープロジェクト・人材育成に取り組んでおり、その**国際版レポート**が毎年発行されています。2012年のそのレポートでドイツはシステムバイオロジーで世界トップになると豪語したことは注目されます。

また、2015年11月に JST/CREST 主催の国際会議(**Structural Biological Dynamics--From Molecules to Life with 60 Trillion Cells**)の招待講演者として来日した独 DKFZ の Ursula Klingmuller 博士(HepatoSys プロジェクトの主リーダーの一人)によれば、更に次のレベルの Systems Medicine プロジェクトが立ち上がろうとしているとのこと。

#### ② スイスの国家システムバイオロジープロジェクト **SystemsX.ch**

スイスは2007年に国家プロジェクト **SystemsX.ch** を発足させ、基盤技術センターを国が整備し、多様な研究提案をボトムアップで受け付けています。2008年には8提案を採択、以降逐年採択提案が増え続け、2015年には27提案が採用されました。これらの毎年の報告会は若手中心のお祭り騒ぎのように活発です。

#### ③ イギリスの国立システムバイオロジーセンターと合成バイオロジーセンター

イギリスは2006年から国立システムバイオロジーセンターを6か所に立て、5年経過後の2012年から国立合成バイオロジーセンターの設立に切り替えています。そのロードマップでは**合成バイオロジー**を国家的重要施策の一つとしています。BBSRC 長官の **Ken Douglas** がこの施策をとったことは興味深いことです。彼はマンチェスターの



国立システムバイオロジーセンター長から BBSRC 長官に着任した後の重要施策としてシステム思考を盛り込んだ**新しいタイプの合成バイオロジー**を展開しようとしています。2013 年以降合成バイオロジーセンターが全英で 6 か所に逐次設立され始めています。

#### ④ EU のシステムバオロジー推進策 (2005 年～)

EU は早くから (科学技術計画 FP-6 の後半 2005 年頃から)、システムバイオロジーの促進策をとり始め、FP-7 では系統的な施策を展開しました。その詳しくは、八尾報告 (2003-2012)を参照されたい。**Hans Westerhoff** (アムステルダム自由大学教授(兼)英マンチェスターシステムバイオロジーセンター長) および **Fred Marcus** (EC/EU 科学技術推進機構 ST 主管)がその推進の中心人物として活躍しました。**Fred Marcus** はのちに大著 **Cancer Systems Biology** (Springer2011) の編著をしました (北野・八尾一部分担執筆)。

### 3-5 植物システムバイオロジー

植物研究分野でもシステムバイオロジーの動きは活発になってきています。上記の英国システムバイオロジーセンターの一つニューカスルセンターは植物の根圏系のシステムバイオロジーを展開しました。

また、2005 年にベルギーゲント大に**植物システムバイオロジー研究所**が **Inze** 所長の下に発足しました。総勢 200 人、各種の機器・実験室を備え、実験研究者と情報解析研究者が一つ屋根のもとに研究する理想的な研究所です。

また日本の理研植物研究センターでも、**篠崎一雄**センター長以下が早くからシステムバイオロジーの重要性に注目し、**斎藤和季**リーダーの下でメタボローム解析を中心としたシステムバイオロジーが展開され成果を挙げています (**Yumi Hirai Nature** 2012)。

また 2008 年に **Joshua Yuan** が **Cell** 誌に **"Plant Systems Biology comes of age"** とレビューを書いたことは、この分野に影響が大きかったと思われます。

### 3-6 細胞システムバイオロジー 単細胞丸ごとモデル、細胞間変換

2011年3月の**Cell**誌が**細胞システムバイオロジーの特集号**を出したことは、21世紀の細胞研究の方向性にかなりの影響を与えるものと思われます。そこでは、16編に及ぶ総説・個別論文によって細胞研究におけるシステムバイオロジーの有用性・課題・展望などが述べられています。

翌2012年7月の**Cell**誌には、スタンフォード大の**MW Covert**グループが**単細胞丸ごとコンピュータモデルの完成**を報告し、更に2013年12月の**Nature Methods**誌で、このようなモデルが科学上の発見をも加速すると述べ、合わせて**モデルの公開**を宣言しています。

---

#### 単細胞丸ごとコンピュータモデル

スタンフォード大Markus Covertグループが、ヒト病原菌Mycoplasma Genitaliumの「細胞丸ごとコンピュータモデル」を作り上げこれを使って表現型の予測が出来るようにしました。この細菌群は、ゲノムサイズが小さく(55~140万bp)、遺伝子数も少なく(400~600個)、以前からミニマムゲノムプロジェクトの対象になってきていましたが(C.Venter et al; PNAS 2006),その理論モデルが作り上げられたことは画期的なことです。

525遺伝子を持つM. genitaliumの細胞内分子プロセスを16種変数、28サブモデル(DNA複製・転写制御・修復、RNA処理・修飾・翻訳、タンパク質処理・修飾・フォールディング、染色体形成・分解、代謝パスウェイ等)に分けて、それぞれに適切なモデル型を作り上げ、それらを全体としてまとめるという手法を取っています。そのモデル作成には900文献と1900の測定パラメータが使われ、更に独自の実験でモデルの検証をしています。単細胞だけでなく細胞集団としての検証もしています。細胞周期の変化による細胞行動の変化にも注目しています。

そして最後にこのモデルを使って、遺伝子ノックアウトによる各種表現型の変化の予測結果を示し、実験で検証しています。その中には実験に合わない結果がいくつかあり、それが新しい発見につながった例(lpdA geneの役割)も示されています。

---

細胞システムバイオロジーの中で、細胞自身のモデル化に加えて、もう一つ重要なことは、**細胞間の変換 (Reprogramming)**であります。このことについて最近、バイオインフォマティクスの観点から重要な進展がありました。**CellNet** は細胞変換を支援するプラットフォームであり、**Mogrify** は細胞変換を予測する計算ソフトウェアです。詳細は、下記文献をご覧ください。

P.Cahan, J.Collins et al; “CellNet: Network Biology applied to Stem Cell Engineering”  
Cell Aug.14, 2014 <http://cellnet.hms.harvard.edu>

O.Rackham, Y.Hayashizaki, J.Gough, FANTOM et al: “A Predictive Computational Frame-work for Direct Reprogramming between Human Cell Types” Nature Genetics  
Jan.18, 2016

### 3-7 モデルデータベース BioModels

以上のように、システムバイオロジー研究が活発化する中(世界の研究機関・学部 200以上、論文年間 5000 件以上)で、実際に研究成果に使われたコンピュータモデルを登録するデータベース “BioModels” が 2005 年に EBI (European Bioinformatics Institute of EMBL) に開設されました。当初から各種の生物についての細胞周期・代謝・シグナル伝達モデルなどが数多く登録され、新たにモデルを作成しようとする研究者に提供されます。発足から 10 周年を迎えた 2015 年現在、BioModels には人手による検証済みモデルが 583 件、未検証分が 796 件登録されています。ほかにパスウェイデータベースからコンピュータによって自動的に作成されたモデルが約 14 万件公開されています。

またシステムバオロジー関連のジャーナルはすでに多種類刊行されていますが、

Nature Molecular Systems Biology, BMC Systems Biology,  
Systems and Synthetic Biology (Springer) ほか

2015年今回新しく Nature 社から **Systems Biology and Applications** が発刊されました。Hiroaki Kitano が Chief Editor をしています。いよいよ多くの応用事例が発表されるでしょう。

### 3-8 日本の活躍

---

参照：JST-SPC レポート 2010 年 11 月

[システムバイオロジーの国際動向と日本の展開](#) (八尾著)

---

2000 年前後に日本がシステムバイオロジーを国際的に先導してきたことは、上述の通りですが、その後のめばしい動きを述べてみましょう。

2002 年に当時東京大学大学院生だった上田泰己がサーカディアンリズムに関係する遺伝子群を同定し、Nature 誌に発表したことは、世界的に先駆的なシステムバイオロジーの成果と評価されてきました。(Nature 2002, Aug.1)

また 2002 年の Nature 誌、Science 誌のシステムバイオロジー特集号に北野宏明が総説を出したことは、世界を主導するものでありました。この 2 編はその後の各国のシステムバイオロジー研究成果論文に必ず引用されていると言っても過言ではありません。(Nature 2002, Nov.14), (Science 2002, Mar.1)

更に 2004 年に北野宏明が、生命システムの「ロバストネス」理論を発表したことは、特筆に値します。(Nature Review Genetics 2004, Nov.5) この理論はその後世界的に多くの生物系・医学系の研究者に注目され、各種の発表が続いていますが、その概要の分かりやすい解説は、次の本でしょう。「したたかな生命」(北野宏明・竹内薫共著、ダイヤモンド社 2007)。がん・糖尿病・免疫・進化・共生など具体的な例で説明されています。

更に、北野主導で、大きなソフトウェア開発が国際的に進められてきています。Systems Biology Markup Language (SBML), Cell Designer 及び GARUDA です。特に GARUDA はシステムバイオロジーの共通基盤プラットフォーム作りを目指す野心的な試みです。

慶応大学の富田勝グループは早くから細胞モデル E-Cell の展開を進め、世界をリードしてきました。モデルの改良・拡張はその後このグループで進められています(高橋恒一ほか)。これに関連し慶応大先端生命科学研究所 IAB では、システムバイオロジーの数々の研究成果や技術を公開しています。(KEIO IAB ホームページ参照)

また、金久實主導の代謝パスウェイデータベース KEGG は世界中から利用が続き、一方では機能の拡充が続けられています。日本の誇れるデータベースの一つです。

この間 2006-2010 年の間、5 大学での「LP 細胞生体システムモデルソフトウェア

開発プロジェクト」(野間、清野、倉智、末松、富田他)が行われました。その成果ソフトの一部はスパコン「京」へ移植されています。

一方で、東大で久田・岡野らで開発されてきた**心臓シミュレータ UT-Heart** は、スーパーコンピュータ「京」に載せられ世界トップレベルの性能を発揮し注目を集めました。更に現在その実応用に向けて動き始めています。

また東大宮野悟を中心とする「**がんシステムバイオロジー**」の研究グループが発足し、関連ビッグデータの活用を含めた解析と応用に成果を挙げつつあります。

東大黒田真也が、トランスオミックスのアプローチでインスリンの時間的作用機序の解析で成果を挙げています (Mol.Sys.Biol.2013, JST/CREST 2015)。トランスオミックスは、各オミックスを縦断的につないで生命現象を系統的に理解しようとするアプローチで、2002年ごろ和田・豊田が提唱した「オミックススペース解析」の考え方の実践版です。

理研岡田真理子が、2014年 Science 誌(May 16) に、がんに関する EGFR シグナル伝達系の理論と実験に基づく細胞のスイッチ応答のメカニズムを発表しました。

日本は、生命システムの理論解析に元々強く、最近では、東大金子邦彦・古澤力が、2012年 Science 誌に、幹細胞の動的システムの理論を発表し、また東大合原一幸が FIRST プロジェクトにおける複雑系数理モデルを生命科学分野に適用した例を主要国際誌に次々と発表しています。

また、2012年に理研に生命システム研究センター**QBiC**(Quantitative Biology Center 柳田敏雄所長)が発足したことも国際的に注目を集めており、成果が相次いでいます。

尚、医科歯科大の田中博が、長年医療分野の情報科学推進をリードし、最近数年間にゲノム医療・オミックス医療・システム医療に関する理論と世界の潮流について、多くの学会・研究会で、指導的・啓蒙的な講演を続けられていることも見逃せません。

また前東大・現理研の杉山雄一が、薬物動態の長年の研究を発展させて、薬物システム動態学(Systems Pharmacology)の分野で世界を牽引する活動を続けられていることも特筆に値します。元 Pfizer・東工大、前理研、現 NPO 理事長の石川智久は、薬物トランスポートの研究で世界をリードし FDA の Pharmacogenomics のアドバイザーを務め、また関連国際学会を主宰しています。

このように、日本には、この分野の国際的な指導的人材が多くいることを誇るべきでしょう。

### 3-9 合成バイオロジーの動き

#### ① 合成バイオロジーの動き 概要

上述のシステムバイオロジーの進展と並行して「合成バイオロジー」の展開が活発です。最近はこれら二つが独立ではなく協調して進展していることが大きな特徴です。上述のイギリスの国立システムバイオロジー6センターから国立合成バイオロジー6センターへの大変換は正にその象徴的な動きです。イギリスはシステムバイオロジーの成

果を取り入れた合成バイオロジーの展開を図り始めました。詳細は別資料を参照して頂きたい、ここでは要点のみを記述します。(合成バイオロジーの展開 2013)

これまでも、**遺伝子工学**によって、一部の遺伝子の改変によって新規な機能を持つ微生物や植物を作ることや、**蛋白工学**によって、新規の機能を持つタンパク質を、天然部品のみならず人工アミノ酸や人工塩基対も利用して作り出すことが可能になってきておりますが、その上で、最近では**ゲノム工学**という段階になりました。**CRISPR/Cas9** 等による**ゲノム編集技術**の大きな進展が見られます。

## ②具体的な動きの例

以下に、これまでのいくつかのトピックスを挙げて見ます。

- ・2011年5月にクレイグベンターのグループ (**JCVI**)が、人工合成ゲノムを導入した微生物を生きて増殖させることに成功したと発表しました。
- ・2012年 新学術領域研究「合成生物学の基盤構築」が、九州大学岡本正宏をリーダーに発足しました。
- ・また同年、新学術領域研究「分子ロボティクス」(萩谷昌己代表)が始まりました。
- ・2013年6月に慶応義塾大学先端生命科学研究所(鶴岡)からスピリアウトしたベンチャー会社**スパイダー社(関山社長)**が、微生物で人工クモ糸を量産することに成功したと発表しました。ゲノム-遺伝子-タンパク質-代謝パスウェイの解析をベースに設計を行った新しいタイプの合成バイオロジーです。
- ・ゲノム工学は日本の**板谷光泰**(現慶應義塾大学先端生命科学研究所、前三菱化成生命科学研究所)が1990年台に提唱し、枯草菌と納豆菌の二つの生物ゲノムの入れ替えによる二つの機能を持つキメラ生物を成功させました。

## ③合成バイオロジー国際コンテスト iGEM

ポストングループが、2002年にゲノム設計のための生物回路群(スイッチ、振動、複製機能など)を数多く集めこれらを組み合わせた新機能生物を創るためのコンテスト iGEM を開始しました。これは毎年行われ国際的に大きな会に発展しています。

- ④ **ポストングループ**は、MIT Tom Knight や Harvard 大 George Church らが **Ginkgo Bioworks** という会社を作って、基礎研究と有用物質創りを進めています。
- ⑤ **Nature Biotechnology** 誌が2009年12月号に **Synthetic Biology** の特集をしました。全部で15報あって、20人の専門家による15報で **Synthetic Biology** の定義から始まって、基礎技術としてのDNA合成・ゲノム工学、応用としての微生物・植物などによる物質・エネルギー生産・緑化のこと、更には、特許やビジネス、悪用や社会的問題など広範囲に議論されています。(この内容については、私が「実験医学」2010年4月号に「合成バイオロジーの新たな展開と課題-ゲノムデザイン、システムバイオロジー、そして社会-」という題で解説を書きましたので、そちらをご覧ください。**Nature Biotechnology** Dec. 2009 他) このあと、**Nature** 本誌が2010年1月21日号で **"Synthetic Systems Biology"** の特集をしました。これには、最新研究論文6件

と過去 10 年の主要論文 13 件が載っている以外に、合成バイオロジーの難しさ 5 項目を指摘しているのが注目されます。

## ⑥ 英国の合成バイオロジー展開

2012 年にイギリスが、「合成バイオロジーのロードマップ」を発表、それ以降次々と具体的施策を展開しています。

---

### 英国の合成バイオロジー推進施策

2012 年 7 月、英国 Technology Strategy Board から「英国合成バイオロジーロードマップ」が公表されました。

短期（～3 年）、中期（～8 年）、長期（～20 年）について提示しています。

このような政策を強く打ち出している背景には、合成バイオロジーの世界市場規模が現在(2011 年)の\$1.6B から 2016 年に\$10.8B に急速に拡大する見通しであること (BBC リサーチ) と、英国がこの分野で強い立場にあることとの認識があります(\*)。

合成バイオロジーの応用分野・市場については、次の 6 分野を挙げています。

1. 医療・健康、
2. 精密化学・特殊化学品（化粧品ほか）
3. エネルギー、
4. 環境、
5. センサー、
6. 農業・食糧

加えて、基盤技術の整備を重要視しています。

合成バイオロジーは、生物学・化学・工学・IT など多様な分野の融合によって進める必要があるため、知識・技術の融合センターと人材育成の必要性を強調しています。

一方では社会の受容や規制の整備を重視し、チェック&バランスが大切と論じています。

(\*)英国のこの分野における強さについて;下記の点を強調。

- ・画期的なブレイクスルーの実績（特に生物科学、工学、IT 分野）とそのイノベーションを支えてきた機関（HEFCE, EPSRC, BBSRC 他）の存在。
- ・合成バイオロジーの世界の研究者数約 3000 人の中で、研究論文数では英国は米国に次ぎ第 2 位。
- ・更に英国における強力な化学会社、製薬・バイオ企業の存在。

イギリスはこのロードマップに沿って、2012 年以降次々と具体的施策を実行してきています。(新規研究プロジェクト認可、ベンチャー企業支援、「知識・イノベーションセンター設立」、欧州 16 か国との合成バイオロジーネットワークの促進、英国立ゴセンバイオロジー 6 センターの設立など)。このための総合司令塔の設立やコミュニティー形成にも力を入れています。今後の動きに注目すべきでしょう。

## ⑦ その他注目事項

尚、下記 3 件は、合成バイオロジーの大きな成果でしょう。

- ・パークレイの Jay Keasling ; 大腸菌の代謝パスウェイ改変による抗マラリア薬 artemisinin の産生効率向上
- ・北里大の大村智 ; 天然微生物の抗生物質生産菌の発見とゲノム情報解析

・ユーグレナ社の出雲充・鈴木健吾：ミドリムシによる栄養食品・バイオエネルギー生産

最後に私が最近注目した3つの動きをご紹介します。

1. 2014年5月にボストンで、“**Bits vs Biology**” というワークショップが開かれました。デジタルな世界（コンピュータ）とアナログな世界（生物）をつなぐ様々な課題を自由に提起し討議するために生命科学・バイオテクノロジー・情報科学などの指導的研究者や行政・社会分野の関係者が集いました。[参加報告](#)
2. 2015年11月に東京で、「**細胞機能の構成的な理解と制御**」報告会が開かれました。若手による「さきがけ研究」で、様々なユニークな発想が示されました。
3. 2011年から始まっているHPCI「**計算生命科学**」推進プログラムの成果が2015年12月(東京)、2016年1月(大阪)と次々に発表されています。予測可能性と制御可能性を目指すとしています。2016年3月号SCLSに、計算生命科学の役割について、柳田・木寺・江口による座談会と、各課題の成果・今後の課題について詳しく載っていますのでご参照ください。

以上様々な活発な動きをご紹介しましたが、これらに共通する考え方は上述のように、要素還元的なアプローチではなくて、構成的アプローチによって生命システムを理解しようとするものです。21世紀の生命科学研究における大きなパラダイムとなっていくでしょう。



## 4. 脳科学の進歩と情報科学、そして人工知能

### 4-1 はじめに

脳科学は生命科学の中でも際立って広範な研究分野であります。分子・細胞レベルから精神・意識などに及びます。脳の研究は通常の科学における要素還元主義のアプローチだけでは到達できそうもない対象を含むことを銘記しなければなりません。

ここでは、次の3点に絞って調査結果をまとめてみます。1つは、脳科学研究に関する各国プロジェクトの概況と、2つ目は情報科学的にみて注目すべき「意識の情報統合理論」を取り上げます。3つ目は、人工知能の研究開発の近況についてです。

### 4-2 脳科学研究

#### ① EU の Human Brain Project

2010年から始まったEUの次期大規模研究プロジェクト Horizon 2020 の募集において、数十件あった提案の中から最終選定2件の一つとして Human Brain Project が選ばれたことは国際的にみても2013年のビッグニュースでありました。約12億ユーロの大プロジェクトです。このHBPは、**神経科学**と**計算科学**と**医科学**とをまとめて発展させる狙いがあり、それらの相乗効果を期待しているようです。ただ Henry Markram 主導の非常にチャレンジングな計画に対し、プロジェクト外の批判勢力からの大きな論争があったようで、このようなビッグサイエンスの順調な船出には大きな難しさがあることが示されました。(Stefan Theil: Scientific American Oct.1<sup>st</sup>, 2015)

しかし、2013年の選定の後の最終審査結果が2015年3月に発表され、2015年10月30日によりやくHBPグループとECの正式調印がすみ、2016年から具体的に研究が始まることになりました。HBPの正式ホームページには、研究目的・研究体制や最近のイベントや研究成果が次々と発表されています。具体的にはこのHBPの中に13のサブグループ(マウス脳データ、ヒト脳データ、理論神経科学、神経情報学、脳シミュレーション、高性能コンピューティング、医療情報学、ニューロロボットなどの他、倫理・社会のグループ)があり、ここには、欧州各国及び国際研究機関が参画し、米国・日本からの参加も予定されています。

研究成果論文も2014~2015年に数百件でており、最近ヒト大脳皮質のブロードマン地図を書き換える新たな地図が出来たと報告されました。3次元画像処理・ビッグデータ解析ほか最新の技術が駆使されています。脳地図作製法に新たな道を開いたと言えましょう。

K.Zillen, K.Amunts: Architectonic Mapping on the Human Brain beyond Brodmann  
Neuron 12-001, 2015

上記3つの科学分野の相乗的発展を目指す野心的な本プロジェクトの進展には目が離せません。



---

## ② 米国 BRAIN Initiative

米オバマ大統領は、2013年4月に BRAIN Initiative (Brain Research through Advancing Innovative Neuro-Technologies) の発足を認可しました。

「人の心の理解を通じて、精神疾患の予防・処置・治療を目指す」として、総額 800M ドルを拠出するとしました。そして 2014 年度\$100M, 2015 年度 \$ 200M, 2016 年度 \$300M と逐年増額してきました。それらを NIH, NSF, DARPA などに配分し、Allen Brain Institute, HHMI ほかの脳研究を加速しようとしています。さらに政府は当初から、このプロジェクトに民間支援の呼びかけを行っていましたが、その結果、更に 300M ドルを超す支援が 2014 年 9 月までに集まったと発表しました。米国らしい展開です。

尚、長年脳科学研究をリードしてきた Caltech 大の **Christof Koch** 教授が、「脳アトラス」で有名なアレン脳科学研究所の所長として迎えられたことは注目すべきことでしょう。下記 4.3 の Tononi の「意識の統合情報理論」との関係で、その展開に興味がありません。

ここではまず、C. Koch 自身の「意識」研究についての近著をご紹介します。

## ③ 日本の脳科学総合研究

日本では、脳の総合的な研究を推進するために 1997 年に理化学研究所に「脳科学総合研究センター」が設立されました。世界的にみて非常に先駆的な研究センターで、「脳を知る」「脳を守る」「脳を創る」の 3 領域を設定してスタートし、のちに「脳を育む」の領域を加え、目標達成型の研究を統合的に進める体制を築き上げました。そして 2007 年には 10 周年記念として「**脳研究の最前線**」という記念出版をしました。その本では、脳の「認知」「進化」「疾患」および「数理」の 4 つの側面から 12 名の研究者が、それぞれの分野の先端研究内容を分かりやすく解説しました。特に、「まえがき」で伊藤正男が述べているように、全体として「脳とこころ」の謎に挑む姿勢でまとめられました。

詳しくはその本を見て頂きたいのですが、「心」「知性」「言語」「認知」「情報伝達」の進化を調べる基礎研究と、「精神疾患」「脳疾患」「快樂」などの理解と治療、更には「ロボット」「人工知能」「理論」にまで言及しています。

また最近の研究動向については、理化学研究所 BSI のホームページをご覧ください。近年の大きな変化としては神経回路ネットワークの研究が増えてきたことでしょう。2004 年の嗅覚神経回路の配線図の解明も一つの成果です。更に生体透明化技術やライブイメージングなど計測技術の開発も進んでいます。2014 年度スタートのオールジャパン「革新的技術による**脳機能ネットワークの全容解明プロジェクト**」にも、多くの大学・研究機関と共に取り組み始めました。

### 4-3 意識の情報統合理論

意識と睡眠について研究してきた精神科医・神経科学者である米国の **Guirio Tononi** (ジュリオ・トノーニ) が、2004 年に発表した意識の情報統合理論は、脳研究における情報科学的な側面として大きな注目を集めました。(著書:意識はいつ生まれるのか―下記)

近年の脳研究において、認知・意識・記憶その他の脳活動の理解には、個々の神経細胞というよりは、神経細胞ネットワークが重要ということは、すでに議論されてきたところですが、トノーニは最新の脳活動測定手段を駆使して、自分の専門分野について、この理論を打ち立てました。この理論の更なる発展・展開が期待されます。

G.Tononi; “Consciousness as Integrated Information” *Biological Bulletin*. 2008

M.Massimini, G.Tononi et al: “Breakdown of Cortical Effective Connectivity during Sleep”  
*Science* 2005 睡眠中の状態

M.Massimini, G.Tononi et al: “Local Sleep and Learning” *Nature* 2004 睡眠と学習

M.Massimini et al; “Recovery of Cortical Effective Connectivity and Consciousness in Vegetative Patients”  
*Brain* 2012 植物状態からの回復

G.Tononi, M.Massimini et al; “A Theoretically based Index of Consciousness independent of Sensory Processing and Behavior” *Science Translational Medicine* 2013 意識測定

IMAGEN Consortium; “Correlated Gene Expression supports Synchronous Activity in Brain Network”  
*Science* June 12, 2015 休息時のネットワーク

---

ここに、トノーニ自身による「[意識の統合情報理論](#)」についての単行本の読後感を添付します。

---

#### 統合情報理論

基本的命題—ある身体システムは、情報を統合する能力があれば、意識がある。

第1公理—意識の経験は、豊富な情報量に支えられている。つまり、ある意識の経験というのは、無数の他の可能性を、独特の方法で排除した上で、成り立っている。いいかえれば、意識は無数の可能性のレパートリーに支えられている、ということだ。

第2公理—意識の経験は、統合されたものである。意識のどの状態も、単一のものとして感じられる、ということだ。ゆえに、意識の基盤も、統一されたものでなければならない。

最終命題—意識を生み出す基盤は、おびただしい数の異なる状態を区別できる、統合された存在である。

つまり、ある身体システムが情報を統合できるなら、そのシステムには意識がある。

意識の単位  $\Phi$

1. システムの中で、ある部分に伝わる情報がほかの部分にも届くかどうか、またその逆も成り立つかどうかを確認する。
2. システムに揺さぶりをかけ、情報がいろいろな構成要素によってどの程度共有されているかを調べる。システムを分割し、他方への伝わり方を調べる。  
システムの構成要素がそれぞれ情報を発している場合に、その上のレベルで生み出された情報を、統合された情報という。 以下別添資料参照

-----

以上ここでは、脳科学の一分野である「意識」研究についてのみ記述しました。その理由は、今後の健康長寿社会の実現に向けて「脳とこころ」の問題が重要視されており、「意識の数量化」は認知症対策等に何らかの貢献が期待されると考えたからであります。前述の通り、脳科学全般の研究範囲は広大であり、未知の領域が広いようです。

ここでは、その一端を述べたに過ぎないことをお断り申し上げます。

最後に、もう2つ追加します。

- ①最近スイスのグループが、視床下部の神経ネットワークを、光遺伝子(Optogenetic)技術によって操作することによって、「覚醒」「意識」がコントロールできることを報告しています。(C.Herrera, A.Adamantidis et al; Nature Neuroscience 21 Dec.2015)
- ② 脳機能に対しニューロンだけでなくグリア細胞が大きな影響を持つことを強調している下記の著書にも注目したいと思います。

「脳とグリア細胞」工藤佳久著 技術評論社 2011年1月刊

工藤氏は三菱化成生命科学研究部脳神経科学研究部門在籍中(1978-1995)に、グリア細胞の特異な働きを発見し、その後長年に亘りその機能を研究し、上記の結論に達したとのことでした。

#### 4-4 人工知能研究

元々、コンピュータ開発の目的は、人間の計算能力・記憶能力・思考能力の模擬と代替にあり、人間の脳のメカニズムの解明はコンピュータ開発と不可分のことでした。しかし、人間の脳はあまりにも高度で、まだ解明できていないことが山積しています。

その中であって、**計算能力**および**記憶能力**はコンピュータがほとんど代替可能なレベルに達し、世の中の様々な局面に充分利用されるようになってきました。最近の人工知能と呼ばれるものは、この範囲のものが多いのです。計算能力や記憶能力を組み合わせれば、かなりのレベルの能力を持つ「人工知能」は作れますし、これからは多種多様な人工知能製品が出回るようになるでしょう。

例. ・ チェス・将棋などのようなゲーム対局ロボット

- ・ 質問応答システムによる医師支援・弁護士支援・教師支援・老人相手ロボット  
(例. IBM ワトソンコンピュータ)

詳細は、下記参考書等をご覧ください。

本当の意味の「思考能力を持つ」人工知能の開発には、人間の学習機能・認知機能を十分に理解しなければなりません。そのための最近の画期的な進歩は、機械学習の中でも下記②のディープラーニング（深層学習）でしょう。

#### ① 機械学習

教師あり学習－正解・不正解データセットから学習 従来からある方法  
教師なし学習－多数のデータの区別を自ら学びとっていく方法  
(ベイズ法、サポートベクター法、ニューラルネットワーク法など)

#### ② ディープラーニング

膨大なデータの中から「特徴表現」を自ら獲得して行く新しい機械学習の方法「ディープラーニング（深層学習）」が、2012年にトロント大のジェフリー・ヒントンらによって開発され、世界コンペティションで圧勝しました。

1000万枚の画像データからコンピュータが自動的に特徴を抽出して分類(概念形成)をしていき、ネコの特徴とイヌの特徴を自動的につかんでいきます。人間があらかじめ特徴箇所を全く教えない画期的な方法です。正に子供が自分で学習するのと同じです。ネコ・イヌなどの名前は後から親や教師が教えることとなります。長年の課題であった「学習するコンピュータ」の道を拓いた出来事でありました。

---

このような最近の画期的な進歩に至る過程で、人工知能研究は50年以上に及ぶ長い歴史があります。その概要は西川豊明の下記総説をご参照ください。

「人工知能研究半世紀の歩みと今後の課題」 西田豊明 情報管理 No.7, 2012

---

その間いくつかの冬の時代をくぐって、最近急速に迎えている「人工知能の春」に至る過程を、最近のホットニュースを交えながら分かりやすく松尾豊が解説しております。

① 探索・推論、②知識ベース、③機械学習から、ディープラーニングに至り、その上で、今後どうなっていくかを論じています。

「人工知能は人間を超えるか」ディープラーニングの先にあるもの

松尾 豊 角川選書 2015年3月

---

更に、人工知能先進国である米国の最近の急速な動きをJETROの八山幸司が報告しています。特に、Google, IBM, Microsoft, Amazon, Facebookなどの企業の動きの詳しい報告が注目されます。

「米国における人工知能に関する取り組みの現状」 八山幸司 JETRO 2015年2月

---

尚、網羅的な調査報告としては、下記をご覧くださいのが良いと思います。

「情報科学技術分野 2015」の「人工知能」欄 JST/CRDS 2015年報告

---

また、2015 年に出た次の 2 編は、生命科学研究分野への人工知能応用について言及しています。

**Nature Insight “Machine Intelligence” Nature May 28, 2015** 人工知能の最も新しい総説集であるが、その中で “Robots that can adapt like Animals” では、環境変化に対応する自律的なロボットの可能性を示しています。

**M. Libbrecht & W. Noble: “Machine Learning Applications in Genetics and Genomics” Nature Review Genetics June 2015**

膨大に産生されるゲノム・エピゲノム・オミクスデータの解析に対し、各種の機械学習法をどう活用すべきかについて解説しています。

---

### 最近の日本の人工知能関連の動き

次に、日本における「人工知能」関連の最近の急速な動きを挙げて見ます。

- 1) 2015 年 5 月に、産業技術総合研究所に AI センターが設立されました。辻井潤一所長以下、所員 65 名、客員 45 名。ポスドク 15 名の予定。
- 2) 2016 年度に、理化学研究所に AIIP(Advanced Integrated Intelligence Platform Project)が設立される予定。
- 3) 2015 年 8 月に、NPO 法人全脳アーキテクチャーイニシャティブ(WBAI) が設立されました(世話人: 山川宏・一杉裕志・高橋恒一・松尾豊ほか)。脳全体のアーキテクチャーに学んだ汎用人工知能の研究を目指すとしています。毎回の勉強会(2014 年から 18 回開催)には、産学の多くの研究者が集まり(約 200 名)、熱気のある議論が展開されています。
- 4) 2016 年 1 月、トヨタ自動車(株)が米国シリコンバレーに人工知能研究所 (Toyota Research Inst.)を設立。ギル・ブラット所長 200 名、5 年 10 億ドル MIT・スタンフォード大と共同研究。クルマ・介護ロボットなど人と協調できる人工知能技術開発。

更に、楽天はボストンとシンガポールに人工知能の研究拠点を設立。

リクルートは、研究所を新たにして総合人材サービス領域での人工知能研究を加速。

- 5) 2015 年 12 月 14 日に、日本オミックス医療学会が「AI 創薬シンポジウム」を開催され、産学から多くの研究者・技術者が集まり、関心の高さが伺われました。そこではビッグデータと AI 技術(スパース推定・ディープラーニング)を使った創薬標的探索・バーチャルスクリーニング等になどの事例(田中博)や、疾病の背後にある観察できない体内メカニズムを膨大な文献データ(200 万文献、800 万ファクト)を駆使して推定した DARPA の事例が紹介されました(辻井潤一)。

また、「ケミカルゲノミクス情報に基づくインシリコ創薬システム(CGBVS 法)を駆使して活性化化合物構造を生成する AI の成功例が示されました(奥野恭史)。

更に、IBM から Watson システムの医療分野への世界的な展開状況が示されました（元木剛）。また、ディープラーニングによる生体内の細胞分子誘導メカニズムや抗体の抗原認識機構の解明に至った例も示されました（三宅淳）。

このように実際に AI 手法がメカニズム解明と創薬に役立ち始めたことは注目に値します。

以上のように、産官学こぞって人工知能研究に殺到している状況です。このような人工知能の発展によって、将来かなりの分野の知的労働が代替されると予想されています。最近の野村総研の調査報告 2015 によれば 49%の仕事が置き換わるだろうとされています。人間はどのような役割を持つようにしていくべきかについて様々な意見が出ています。

その上で、独創性を持ち新たな発見・発明が出来るような人工知能の開発が、次の大きな目標です。今後の人工知能が、アイデア創出・発明発見などをどこまで出来るようになるかは、注目に値します。

北野宏明は、2015 年の米国人工知能学会で国際的にそのことを呼びかけました。

**H.Kitano; “Artificial Intelligence to Win the Nobel Prize and Beyond: Creating the Engine of Scientific Discovery” AI Magazine, 2015**

グローバルなオピニオンリーダーの一人である北野の提言は、今後に大きな影響力を持つことでしょう。（北野は、1995 年頃「ロボカップ」の提唱、2000 年頃から「システムバイオロジー」の提唱・先導、2004 年に生命体の「ロバストネス理論」の論文発表、更には 2010 年頃「クラウドソーシング」（ネットワーク DB 構築）の実践 などで、次々と世界をリードしてきました。）

## 5. 情報科学の新たな展開ーデータサイエンス、ビッグデータ

### 5-1. 概要

近年の計測技術と情報技術の発展に伴い、社会のあらゆる側面で情報収集・伝達・活用が急速に進んでいます。いわゆる「ビッグデータ」の影響が日に日に強くなっていきます。このことはビジネスの世界だけでなく、個人の生活にも、また科学技術面でも影響を与えつつあり、その度合いは今後 5~10 年の範囲でも極めて大きくなっていくと思われれます。

もちろん、生命科学研究と応用の面でもその影響は計り知れないものがあります。例えば健康情報で言えば、個人の行動・体調・環境などのデータが、ブログ・スマホ・装着センサーなどから、刻々収集できる状況になりつつあります（臨床研究・コホート研究・病院カルテ・健診データなど定型データに加えて、新たな非定型データとして例えば感染症発生兆候予測などに）。またそれらの多様なデータの解析技術（アナリティクス）も人工知能技術を含めて急速に進んでいます。

これらの大きな動向を踏まえて、欧米でもわが国でもその研究や施策が始まっています。そこにはグーグル社、IBM など巨大情報産業の先導的な動きがあり、例えば健康医療産業の今後の展開に大きな影響をもたらすと考えられています。正に「データ駆動型研究」「データサイエンス」が第4のアプローチになろうとしているのです。

これらの全般的な動きと先駆的な事例をいくつか紹介し、今後を展望いたします。

### 5-2. 科学技術面でのビッグデータの影響

Nature 誌が「ビッグデータ」特集号を出したのは、2008年9月4日号でありました。当時すでにインターネットの拡がりやグーグル社の動きなどが背景にあり、ライフサイエンス分野では、欧州サングァーセンターのゲノム情報の巨大化の例が取り上げられました。日本から五條堀孝もコメントに加わっていました。その後の NGS (次世代 DNA シーケンサー) の急速な発達によって、ライフサイエンス分野のデータ急増は衆知の通りであります。



2008年9月4日号の Nature 誌は Big Data 特集号でした。世界的に非常に早い時期の特集であり、その後のサイエンス社会の意識変革に一石を投じました。Science in the Petabyte Era という副題がついていて大きな二つの流れが紹介されていました。

一つは、設立 10 周年を迎えた Google 社の動きでした。世界中の膨大な文章情報・画像情報さらには映像情報を蓄え、人工知能を

利用した検索サービスをして世界のあらゆる階層の人の知的レベルの向上に役立っていることを示した上で、2018年までにさらに多様な情報の統合サービスが進むことを多くの識者が述べていました。Google社は、その後画像情報のビッグデータ解析によって「ネコ」の顔認識が「教師なし学習法」によってできたことを示しました。ビッグデータと人工知能技術の融合による成果例でありました。

もう一つは、ゲノム・DNAの解析技術いわゆる次世代シーケンサーNGSの異常とも言える発展を迎えて世界最大のシーケンスセンター Wellcome Trust Sanger Instituteがペタセンターになったことを紹介していました。しかし、その後のNGSの更なる性能向上によって、自身で20ペタバイトを保有するだけでなく、ついに外部のクラウドコンピューティングを利用することになっていったこと(Nature June 13, 2013)や、同一感染症の国際的大規模疫学のビッグデータによって原因究明の精度が格段に向上した例が紹介されています(Science Nov. 28, 2014)。NGSの進歩はその後も続き、ヒトゲノム解読計画がほぼ終了した2000年にヒト一人のシーケンシングコストが\$400Mであったものが、2015年現在\$1000に低下しました。

これを受けて米国NIHは2015年1月に、100万人以上を対象とする”Precision Medicine Initiative”(精密医療または適確医療)を立ち上げました(Obama \$215M)。これはPrecision Medicineの実現を目指すためにガンをはじめとする各種疾病の患者の医療記録・遺伝子型・細菌叢・生活環境・生活スタイルのビッグデータを総合的に集めて分析するコホートスタディで世界最大規模です。これまでの中小コホートスタディ200件(数千~1万人対象)を一気に上回るもので、個人別医療の実現に向けて大きく期待されています。

もう一つ、生命科学研究分野での重要な変化は、「生き物の動的な状態の3次元計測とその表示・解析」の急速な進展です。静的・平面的な顕微鏡データではなく、刻々の変化を捉え、それを直接コンピュータに送ります。その最も基本になるのが、「細胞」観察であります。単細胞である微生物から多細胞である高等生物まで、細胞研究は大きく変わり始めています。ここでは、時間変化を含む膨大な4次元データが昼夜を問わず集積され、この中から特徴ある変化を捉えなければならないのです。研究スタイルの大きな変化がおこりつつあります。このことは研究発表あるいは教育にも影響を与えています。(YouTubeによる発表が急増)

### 5-3 科学における第4のパラダイム データサイエンス

上記Nature2008の特集に次いでもう一つインパクトがあったのは、”The Fourth Paradigm -- Data Intensive Scientific Discovery”という報告書が、2009年にMicrosoft Research Inc.のTony Heyらによって公表されたことです。科学の発達、「実験」・「理論」・「シミュレーション」と共に、「データ」によってもたらされる時代になったとの認識です。この概念は、科学研究の世界に大きな一石を投じたと言えます。 (Science 6 Mar. 2009)



- 第1世代 「経験科学」 数千年
- 第2世代 「理論的科学」 数百年
- 第3世代 「コンピュータシミュレーション」 数十年
- 第4世代 「データ科学」

と位置付けています。

この報告書第2章 Health and Wellbeing の中では、“A Unified Modeling Approach to Data-Intensive Healthcare”を提唱し、診断・治療に当たって従来の専門分野を超えた総合的なアプローチが可能なことを示しています。(事例-後述)

#### 5-4 ビッグデータ時代における考え方

ビッグデータ時代にはこれまでの科学技術におけるデータに対する考え方を変える必要があります。ビッグデータを加速する3要因と扱いの変革3点は下記の通り。

- ・ビッグデータがこのように加速してきた大きな要因は、次の3つ。
  - (1) インターネットの発展による人・物・情報のつながりの普及 (ICT)
  - (2) 各種センサーの発達と発信元の多様化 (IoT)
  - (3) コンピュータ能力 (記憶容量・スピード) の増大と、解析技術 (アナリティクス、特に人工知能) の進歩
- ・その上に、データの扱いが大きく変わってきました。
  - 第1 データを選別・抽出するのではなく、全データを扱う。
  - 第2 仮説ドリブンでなくデータドリブンのアプローチをとる。
  - 第3 因果関係でなく、相関関係を重視する。

---

このような考え方に大きな影響を与えつつある下記の著書をご参照ください。

「ビッグデータの正体」 V・M ショーンベルガー& ケネス・クキエ著 2013

齋藤栄一郎訳 講談社 2013年5月刊

原題 “A Revolution that will transform how we live, work and think” にあるように“**Data is New Oil**” との考え方で、情報の産業革命が世界のすべてを変えるとしています。

冒頭から、世界を変えるビッグデータの一例として、グーグル社によるインフルエンザ流行の予測成功の紹介をしています。(2009年の新型インフルエンザ流行の数週間前に、グーグル社がインターネット上の5000万件の個人検索情報からの予測結果を Nature 誌に発表した。)

このような具体例を、旅行・娯楽・金融・物流・製造・教育・科学技術など社会の多様な局面で示しながら、以下のような重要な考え方の変化 パラダイムシフトを提示しています。

- ・量が変われば質が変わる。
- ・すべてのデータを扱う。
- ・因果よりは相関。
- ・あらかじめ仮説を立てないデータ解析結果から考える。

この中でも、「因果よりも相関」という考え方は、これまでの科学技術の進歩の方向に逆行する考え方です

が、医療研究分野・コホート研究分野では統計解析が一般的であったので大きな違和感はないようです。ただ、近年のアナリティクス（データ解析技術）の進歩は見逃せません。特に Deep Learning（深層学習）の人工知能技術が、大きな牽引力になっています。

データについても、最近では従来とは異なるデータソース（ブログ、ケータイ、スマホ、ウェアラブル端末）からの不定型、動的なデータを扱うようになってきて、ますますビッグデータ化してきています。

---

## 5-5 健康医療分野でのビッグデータ マッキンゼー報告書

2013 年に出されたマッキンゼー社の調査報告書 では多様なソースからの患者データを疾病間にまたがって総合解析した結果の診断治療の成功例が数多く紹介されており、さらに今後の展望が述べられています。

**The Big Data Revolution in Healthcare” A Study by Mckinsey & Co.**

by Drs. B.Kayyali, D.Knott and S.Kuiken April 2013

最近の情報技術の進歩によって、従来分散していた患者データを多様なソース（病院・製薬企業・保険機関ほか）からまとめることによって、総合的な分析ができるようになった。特に疾病ごとに分かれていたデータを連携することによって新たな発見につながる例もみられ、大きな価値を見出している。

- ・ Kaiser Permanent 社は、HealthConnect というコンピュータシステムに、多くの医療機関からのデータを集約している。その分析によって、心臓病関連で開発費および診療費の約 10 億ドル削減に成功した。
- ・ Blue Shield 社は、NantHealth 社と共同で、統合システムによって医師・病院・健康プランの個人別対応ができるようになり、多くの疾病について効果を上げている。
- ・ AstraZeneca 社は、HealthCore 社(WellPoint’s data and analytics の子会社)と共同で、独自の臨床試験データを含む実世界データを用いて慢性疾患・生活習慣病に最も効果的な治療法を開発している。

McKinsey の調査では、このような統合ビッグデータ解析によって 2010 年以降でも 200 以上の新規ビジネスが生まれているとしています。

## 5-6 欧米の取り組み 米国、NIH、米州、英国ほか

### 1) 米国のビッグデータへの取り組み

オバマ大統領就任以来の大きな施策の一つとして、「オープンデータ」がありました。それは政府行政機関の保有する各種データを市民に開放しようとするもので、2010 年には 40 万件のデータが公開されました。その上で、米国は、ビッグデータの活用を促進するために、“**Big Data Research and Development Initiative**” を 2012 年 5 月に発足させました。「ビッグデータを科学研究、気象及び医療研究・教育・国家安全保障のために有効に活用する能力を高めるためとして、White House Office of Science and Technology Policy (OSTP) 主導の下、多省庁共同 (NIH, NSF, DOD, DOE ほか)で進める施策で、その技術開発に総額 200M\$以上を投じると発表しました。これは大統領科学

技術アドバイザリーボード(PCAST) の推奨によるもので、その共同議長には J.Holdren 博士と共に、Broad Institute 所長の **Eric Lander** (ゲノム医療の推進者)が加わっています。

6つの政府機関が\$200M を投じて技術開発を促進するとしていますが、全部で約 90 あるプログラムの中 23 プログラムが **NIH 関連の研究テーマ**に向けられることになっています。ゲノム情報だけでなく、各種の生命現象・疾病の研究が、この恩恵を受けていくと思われます。

## 2) NIH のデータサイエンスに対する取り組みと **BD2K センター**の設立

NIH は今後の健康医療分野におけるビッグデータとデータサイエンスの重要性に鑑み、12か所に **BD2K (Big Data to Knowledge)**センターを設立しました。その中の一つ **LINCS-BD2K (at Mount Sinai)** は全体の統括・調整を担当し、他の 11 センターはそれぞれ特有の役割を持つとしています。ゲノム・タンパク質・オミックスデータなどにとどまらず、むしろ個人の移動情報、モバイルセンサー情報を含む広い意味の表現型の扱いなどを実地で研究するような新しいタイプのセンターが多くなっています。(NIH-HP 参照)

## 3) 米国州政府の健康医療へのビッグデータ活用

**米カリフォルニア州政府**が **UCSF** と共同で健康医療で世界一を目指すとしており、サンフランシスコのミッションベイに大きな投資を始めています。2010 年発表された大拡張計画(総額\$1.52B、多くの病院、研究者・医師など 5000 人規模)の実現に向けて 2015 年から具体的に動き始めました。この新 **UCSF Institute for Computational Health Sciences** の所長として **Stanford** 大の **Atul Butte** 教授がリクルートされたことは大きなニュースとして注目を集めました。

UCSF Taps Atul Butte to Lead Big Data Center, by Kristen Bole on January 15, 2015

**Atul Butte** は、ビッグデータを駆使したシステム医療・ゲノム医療のリーダーであり、**Stanford** 大の **Translational Bioinformatics** の教授として過去 5 年間に多くの成果を挙げてきました。特に、多分野にまたがる遺伝子・疾病間のビッグデータの解析によって**既存薬の適用範囲の拡張**の道を示しました。また、旧来の**疾病分類法にゲノムに基づく新たな視点**を提唱しました。彼は新たな立場で、ビッグデータを活用した総合的診断・治療の道を拓くものと大きく期待されています。

## 4) 欧州

① 英国では、2014 年に **Big Data Roadmap in UK** が出されており、それに基づく公募が始まっています。

② スイスでも、システムバイオロジー国家プロジェクト **Systems X** の中でビッグデータを重要課題として取り上げています。

“Big Data in Systems Biology” X-Letter 30, May 2015

## 5) 参考文献

更に健康医療分野でのビッグデータの活用について、下記のような文献が出ています。

Big Biological Impacts from Big Data Science Sciencemag.org p.1298-1300, 2015

Life Sciences: Preparing for Big Data and Analytics EY Report 2015

International Workshop on Big Data in Life Sciences Atlanta, GA., Sep.9, 2015

How Big Data is transforming The Fight Against Cancer Forbes Tech. 2015

Bio and Health Informatics meets Cloud: BioLab as an example

Review ; BioMed Central Health Information Science and Systems Jan.6, 2013

Big Data for Personalized Medicine and Biomarker Discovery

Nicholas Dracopoli May 5, 2015 at 11<sup>th</sup> Annual Biomarker and Diagnostics

Big Data meets Public Health M.Khouny et al. Science Nov.28, 2014

BigDataViewer: Visualization and Processing for Large Image DataSets

Nature Methods June 2015

Big Data – Proposed study to track 10,000 New Yorkers

Gather Data first, Produce Hypotheses later Science Oct 30, 2015

## 5-7 日本の取り組み

### 1) 情報通信政策

2014 年版「情報通信白書」の第3章で「データが切り拓く未来社会」を論じています。ビッグデータ・オープンデータ・パーソナルデータの日本全体の総合的な現況と予測について社会の動きと情報通信政策が分かります。たいへん大事な資料で、その社会的環境の中で、科学技術・健康医療分野の展開を考えるべきでしょう。

### 2) 関係者－論文・総説

更に、下記のような論文・総説が、日本のバイオインフォマティクス指導者から、次々と出されるようになってきました。

「ビッグデータのゲノム・オミックス医療への応用」 田中 博 2014年12月5日

「ライフサイエンスにおけるビッグデータの解析」 松田秀雄 2014年9月12日

「ますます多様化・大規模化する生命情報に求められる情報技術とは？」

高木利久 2012年11月9日

“Tackling Biomedical Big Data –Cancer System” 宮野 悟 2013年5月31日

「ビッグデータ応用：ブームから挑戦へ」 JST/CREST 田中 譲 2015年10月28日

## 5-8 今後の展望 ビッグデータと人工知能の活用、データ・情報の発生源の変化

以上のようなビッグデータの傾向は、今後ますます大きくなっていくものと思われます。Google 会長の Eric Schmidt が述べているように (The New Digital Age 2013)、現在世界の 20%以上の人々が利用しているインターネットを、5~10 年後 80 億人の大部分が使うようになり、それぞれが受信をするようになると予想されています。更に、最近急速に喧伝され始めている IoT (Internet of Things) の影響は、計り知れないものです。すべての物体・動物・人間につけられた各種の機器・センサーから発せられるデータがクラウドに送られる時代を迎えようとしています。

個人レベルでも社会レベルでも様々な個別情報と周辺情報が、日々・刻々と捕捉・集積され転送される時代になってきました。これらの中には科学技術研究にとって貴重なデータが含まれている可能性が高いのです。これらは、データ収集の概念を大幅に変えていくでしょう。またそれらを解析・整理することによって得られる知識・論文なども膨大になってきました。このようなデータ及び知識は、人間の解析・思考レベルをはるかに超す勢いにあります。

ここに「人工知能」の力を借りなければならない局面が出てきております。現在の人工知能はまだ限界がありますが、うまく利用すれば人間にとって助かるレベルにはなってきました。

- ・ IBM ディープブルーのように、チェス・将棋などで人間に勝てるレベルに達したこと
- ・ Google サービスのように、膨大な情報の中から必要な情報を探し出してくること
- ・ IBM ワトソンのように、専門的な知識を組み合わせる質問への回答候補を示すこと
- ・ Amazon のように、膨大な情報の中から概念抽出ができること (猫の顔の認識)

など、近年の成果は 10 年前には考えられないレベルになってきました。

これらを上手に利用すれば、ビジネスでも研究開発でも非常に効率を高めることができることは確かでしょう。

科学技術、健康医療に関する重要な情報・データが、実験室・研究所の外に、個人・社会・自然の側に、多く発生・存在する時代になろうとしています。よく考えて行かなければならないでしょう。

## 終わりに

以上、

ゲノムワイドな大きなパラダイムシフト（網羅的生命部品解析）

システムバイオロジーの視点に基づく構成的アプローチへの変化 及び

ビッグデータ革命のこと

について述べてきましたが、時代の潮流をよく見れば、これらの大きな変化は必ず起こり、

そして続くと考えられます。

情報技術のあまりにも急速な動きと、底に流れる大きなパラダイムシフトとを、どのように捉えて行くかが人類的なレベルで問われているように感じています。

更に、私が感じている課題や視点を下記に項目のみを列挙します。

多様なお考えを持っておられる多くの皆様と幅広い議論をさせて頂ければ幸いです。

平成 28 年 3 月 八尾 徹

---

課題・視点の項目列挙（本報告と一部重複）

1. 今後の生命科学研究に大きな影響を与えそうな技術

ゲノム編集技術

オプトゲネティクス

生き物透明化技術

一細胞内解析技術

細胞間ネットワーク測定技術、操作技術

生物・生体情報の連続測定技術、可視化技術

2. 今後の生命科学研究に大きな影響を与えそうな考え方・進め方

オープンサイエンス（知識の共有）

ネットワーク上のコンテスト（世界から知・アイデアの効率的収集）

人間・ロボット共存システム

実験室以外からの表現型データの急増とその扱い

物理化学的アプローチとマクロ的アプローチ（要素還元論型とシステム構成論型）

3. 生命科学技術の発展と、人間・社会・自然とのかかわり

ELSI (Ethical, Legal, Society Issues) についての検討グループの強化

・生命倫理について

・プライバシー・知財について

- ・ 社会との交流について、自然・環境・動植物への影響について
- ・ 人材育成について

## 八尾 徹

- ◎ 理化学研究所 イノベーション推進センター  
中村特別研究室 客員研究員 (アドバイザー)
- ◎ 慶應義塾大学 S F C 研究所 上席所員

### 略歴

- 1935年7月 福島県 生れ
- 1954年3月 福島県立会津高校卒業
- 1958年3月 東京大学工学部応用物理学科卒業
- 1958年4月 三菱化成工業(株)入社 (現三菱化学) (1998年3月退職、40年間勤務)
  - 1982年 三菱化成総合研究所システムセンター所長 (4年間)
  - 1986年 蛋白工学研究所研究部長 (ナショプロへ出向5年間)
  - 1991年 三菱化学横浜総合研究所技師長 (7年間)
- 1998年10月 理化学研究所ゲノム科学総合研究センターアドバイザー(10年間)
  - 2004年 産業技術総合研究所 兼務 (~2013年)
  - 2007年 科学技術振興機構 研究開発戦略センター兼務 (~2011年)
  - 2007年 慶應義塾大学 S F C 上席研究員 兼務 (継続中)
  - 2007年 理化学研究所横浜研究所研究推進部 嘱託
  - 2013年 理研和光・横浜の3センター客員研究員
- 2015年 4月以降 現職 (上記) (理研勤務18年目, 慶応大兼務10年目)

### 大学等での非常勤講師

東大、東工大、東京農工大、東京医科歯科大、慶応大、学習院大、桐蔭横浜大、  
東京理科大、名古屋市大、神奈川科学技術アカデミー、東京テクニカルカレッジ、大  
宮スーパーサイエンスハイスクールほか

### 公的機関・プロジェクト等の評価・推進委員・アドバイザー

通産省-- 生物分子工学研究所、ゲノムインフォマティクス、アトムテクノロジー、  
生物機能ミミック、糖鎖合成遺伝子  
農水省--イネゲノムシミュレータ、イネ・植物 データベース  
文科省--細胞・生体機能シミュレータ、J S T北野共生システムプロジェクト  
新学術領域 岡本合成生物学、先端機器共用施設 瀬藤質量分析計  
横浜市技術リエゾン委員  
シンガポールライフサイエンス教育プロジェクトアドバイザー その他