



ゲノムサイエンス
教授 油谷 浩幸
Genome Science
Professor ABURATANI, Hiroyuki

膨大なゲノム情報を解析し 生命現象と疾患のメカニズムを解明

インタビュー：伊福部達教授

ヒトの遺伝子約 3 万個を “個”ではなく“全体”で捉える

—— DNA 解析の分野は今、脚光を浴びています。先生の研究を“予習”していたら、「転写レギュローム」という言葉が出てきて、このあたりがキーワードになる気がします。

レギュローム (regulome) の語尾の「オーム」はゲノム (genome) の「オーム (-ome)」と同じで、ギリシャ語で“全体”という意味です。今まではタンパクや遺伝子を一個一個見ていましたが、先端研に来てからは 3 万個全体を観察する技術を作ったり、その技術で全体がどう変動しているかを研究しています。

—— 3 万個が勝手に変動するのではなくて、有機的な相互作用があるというわけですね。

相互作用もあるし、ヒエラルヒーもあります。池に石を投げた時に出来る波紋の形を見て、石の大きさや飛んできた方向のパターンを推定できるのと同様に、遺伝子の動きから元の変化、癌であれば異常が何であるかという推測がかなり出来るようになっていきます。今までは 3 万個のうちの 1 個の変化を癌の目印にする研究がされていたのですが、やはりそれで全てを説明できるわけではないので、パターンで物事を見ようと。

先端研に来た当初は、バーチャルリアリティ (VR) が専門の廣瀬通孝教授や今は柏に移転された人工物研究センターの岩田修一教授のグループと 3 万個のデータマイニングを行う手法の開発をすすめました。データのパターンをウニのかたちに例えるなど、大量の情報を人に見やすくする方法を共同で開発するなど、先端研らしく非常に学際的な研究を進められたのは幸運でした。

現在は、生命が全体のシステムとしてどう制御されているかという研究と、それとは別に、個人の遺伝情報がどう違うかという研究をしてい

ます。地球上には何十億人という人がいますが、人間はもちろん、同じ日本人でも相当違います。それを調べることで、メリットの方を言えば、どういう病気になりやすいか、どういう薬を飲むと副作用が出やすいか、ということがわかるようになります。

人間の DNA には ACGT の 4 つの塩基しか無いのですが、個体差で配列の違う箇所がゲノムの中に 1000 万くらいあります。1000 万は 10^7 で人間のゲノムのサイズは 3×10^9 ですから、全体に対してごく一部の配列が個体間で違う。一部には繰り返しもある。親から子へ伝わるときに配列が変わってしまったり、一部がひっくり返ったりということが相当の頻度で起きています。起きる方向性はニュートラルで、おそらく生物が進化する過程で多様性を保とうとして生じているのではないかと考えられています。ゲノムの一部分のコピー数の個人差について、国際共同研究で「Nature」などで発表しました

ゲノムから RNA が作られる「転写」のプロセスは どのようにコントロールされているか

—— 二重らせんや ACGT などは止まっている暗号として記録されて不変だと教えられてきた気がするんですけど、実際には変わるということですか。

一人の人間の頭のてっぺんからつま先までは、基本的には変わりません。ただ世代から世代へは、ちょこちょこ変わっていく。さらに、一つの受精卵から神経や脳になる細胞もあれば、肝臓になる細胞もあって、それがどこでコントロールされているかが実は今一番、興味のあるところなんです。 3×10^9 個の ACGT という 4 つのアルファベットの並びは決まりましたが、その意味はまだほとんど解読できていません。たんぱく質はこの領域が作るだろう、そのたんぱく質を作る遺伝子は 25,000 個くらいしかないだろう、ということはわかってきた。さらに細かく言えば、たんぱく質のユニットの使い分けもあります。それでもタンパク質の種類としては 10 万個程度。だとすると、10 万個の部品で人間の体を本当に作れるのかとなると結構、難しいと思うんですね。お酒をつ

くる単細胞生物の酵母でも遺伝子は6000個、ハエでも1万数千個ありますから。

そこで、複雑な生物であるヒトの場合はもっと細やかなレギュレーションがあるというのが「転写レギュローム」というちょっと難しい言葉で表現されています。少しわかりにくいんですが、紐になったDNAあるいは染色体にメチル基やアセチル基、リン酸基といった修飾が入ることで例えば、神経になる細胞はここを使うとか、子どものうちはここを使うけれども大人になったらこっちを使う、あるいは子どもの頃使っていた遺伝子を、大人になって違う目的で使ったり、と細胞ごとに運命を変えているんですね。

——その制御の仕組みが研究の目的なわけですね。

例えば、DNAにメチル基が入ってメチル化された領域はサイレンスされて不活性になってしまっているわけです。メチル化の乱れによって病気がひきおこされることもあります。ヒトのES細胞について聞かれたことがあると思いますが、どんな細胞にでもなれる細胞ですね。一つの細胞から人間の体は、口から腸までの消化管になる内胚葉、筋肉や骨、血液になる中胚葉、そして頭や皮膚になる外胚葉と大きく三つに分かれるわけです。ES細胞はその全てになる能力があるので、実験的に分化していく過程で先ほど言ったようなメチル化、あるいは染色体に加えられるマーキングによってこの領域は使う、使わないといったような制御の仕組みを調べています。

ES細胞は、まだフレキシブルで比較的染色体がルーズな状態にある。それが分化すると固定され、染色体のある部分をコンパクトに折り畳んでわなくしてしまう。使うところだけを開いておく。開いておくにもそのための目印が必要ですし、遺伝子を読み出すために転写因子が関係しています。ゲノムからRNAが作られる反応を「転写」、RNAからたんぱく質を作るところを「翻訳」といって、そ転写反応を直接指示するのが転写因子、これが人間の体には1500以上あります。

最近分かってきたことなんですが、DNAから生まれたRNAの中に、直接またDNAに働いてDNAの読み出しや翻訳反応、作られたRNAをそのまま壊してしまう機能を持ったRNAがある。RNAはおそらく生命の起源として最初に出てきたモレキュール(分子)で、構造的にも一本の鎖ですから形を変えやすく、酵素としての働きや蛋白質を認識する機能を持たせやすい。RNAが作られる時に転写因子がどこにつくかというレギュロームの研究は、国内では先端研がひとつのメッカになっています(参考文献1)。

3万個のパターンをみると、どの転写因子が使われているかの推測はある程度つくのですが、知りたいのは最後に現れた形質の違いを引き起こすためのもともとのDNA配列の違いです。これまでは、最終産物としてのRNAやたんぱく質の違いで病気を説明しようとしたんですが、転写因子が結合する塩基の配列が一つ違っただけでは完全に構造が変わるような0か1かの差にはならず、結果として転写の効率が3倍上がるといった程度です。しかしながら、アルツハイマー病などではその差が60年、70年積み重なって出現するわけです。

あるいはダウン症候群だと染色体の21番が2本か3本かという1.5倍の差にしか過ぎません。でもその差が、先ほどのDNAのコピーの問題もありますけれども人間の全体のシステムに大きな影響を及ぼします。コピー数の違いというゲノムの多様性があることを国際共同研究で明らかにして昨年発表しました(参考文献2)。

「転写」の正常化が 遺伝病の予防・治療のカギに

——ということは、転写が正常に行われるようにコントロールできれば、そういう遺伝病を防ぐことができるわけですか。

そうですね。白血病の場合は転写因子が原因になっていることが多くて、本来であれば静かにしているべきものがONになってしまっている。子どもの頃だけに活動的であるべきものが大人になっても作られ続けたり、誤って作られたりすると癌みたいな状態になってしまうわけですね。染色体の構造が変わってしまうなどで、染色体がスイッチONの状態にしてしまうと、異常な転写が常に起きて細胞は、どんどん増殖を続けてしまいます。

DNAは細胞核にしまわれていますが、ここで何らかの刺激を受け取って核に伝わるとDNAから情報を読み出すので、その経路に異常があると問題が生じますので、その異常回路を遮断すればよいのです。最近、乳がんの治療で使われているハーセプチンという薬は、異常増幅を生じたHER2(ハーツー)という増殖因子受容体を抑える働きによって劇的に近い効果が出ています。

乳がんの話になりましたので女性ホルモンの話をすると、女性ホルモンの受容体はそれ自身が転写因子で女性ホルモンが結合すると、女性の機能を維持するための遺伝子を全てONにするわけです。乳腺の細胞は女性ホルモンが活発な時期に増殖しますから、50歳前後までが乳がんのリスクが高くて、女性ホルモンのレベルが下がるそれ以降は乳がんのリスクも低減します。もっとも年齢とともにリスクが小さくなるのは乳がんくらいですね。体内には新しい細胞を作る際に複製されるDNAの間違いを常に防ぐための監視機構があるんですが、60歳を過ぎる辺りでその機構がサボり始めて、急に変異が増加する。変異があちらこちらに入ってくると監視が追いつかなくなり、積もり積もって癌が発症しやすくなるわけです。

——ということはある人が、いつ頃癌になるかがある程度わかるわけですね。

以前より遺伝性腫瘍の家系があることが知られていました。今、世界中で塩基の配列の違い(多型、SNP)を病気のグループと健康人グループの間で較べて病気に関連する遺伝子領域がわかりつつあります。その中には癌のなりやすさに関する遺伝子があることもわかってきました。といっても、あくまでも因果関係ではなくて相関関係でしかないんですけども。

ヒトゲノム解析とコンピュータサイエンスの融合

——膨大なデータの組み合わせから一部の結論を出すというのは、計算理論そのもののような気がするんですが、例えばコンピュータサイエンスの分野とのタイアップは進んでいるのでしょうか。

確かに出てくる情報量は天文学よりもバイオの方が多いいんじゃないか、というくらい情報が飛躍的に増えていまして、つい2、3週間前に24テラバイトのサーバを注文したところです(笑)。(編集部註:テラは 10^{12} (=1兆)で1000ギガ)

——信じられない…。一人の人間の画像記録が数テラでできるのに。

ゲノムのシークエンスを決定する(DNAの配列を読む)場合に、ゲノムプロジェクトでは何百億円もかけましたが、現在一人当たりだと5千万円くらい、それでできちゃうわけです。それを1000ドルにしようとしてアメリカのNIH(National Institute of Health)がスローガンとして掲げています。実際のところ、5000ドルにするには5年はかかるでしょうが10年はかからない、という感覚です。そうなるシークエンスを進めて、正常な配列と癌の配列の違いを比較することが研究発展の視野に入ってきます。まさに「個のゲノム」の時代になりつつある

と思います。

そこでの問題の一つはデジタルデバインド。バイオの人間は算数が不得意で(苦笑)、ピペットを使って実験するのはいいんですが、出てきたデータを数字にして現象のパターンが違うことを調べるのは苦手です。3、4万の遺伝子レベルはまだよかったです、ゲノム全体の 3×10^9 が相手だとギブアップ、そういう研究が出来る所が国内には少ないのが実態です。近くにVRやコンピュータサイエンスの研究者がいて、特に南谷・中村研究室にかなりハードコアな計算をさせていただいたのは先端研ならではのですね。大規模情報をマイニングするためには、新しい技術や理論が必要になると思います。

——別の先端の技術と融合して本当に最先端をいく、ということがあるんですね。

臨床の現場から、サイエンスの世界へ

——ところで、先生の履歴を拝見したんですが、研究よりも患者さんを診る方が医者らしいし、お金も入ると思うんですが、どうしてこんなことになってしまったんでしょう(笑)。

お金が入るといのは30年前の話でしょう。サイエンスはおもしろいし、自分の好奇心が満たされる感覚があります。医者や弁護士はある意味で受身で、目の前の患者さんを治そうと調べることは出来ても、新たに治療法や診断法を開発することまでは臨床の現場では不可能です。もちろん、いくら研究が進んでも癌を手術したり、骨折を治したりする医者には必要なのですけれど。加えてゲノム科学などの進歩により基礎と臨床の距離が近づいたのも事実です。

臨床医としてのトレーニングを一通り受けて30歳を前にして何をしようか、という時が80年代の中頃。組み換えDNA技術といって、今は分子生物学という分野がヒトで出来るようになりました。それまでは“アナログ”だった生物学が、遺伝子を扱えるようになったことでデジタル化できるようになりました。ジェネレーションギャップがあって、それ以前は生化学、その後の世代はいきなりDNAに入っているの、最初から“デジタル”っぽいんですよ(笑)。最初のトレーニング環境の違いが今も影響している気はしますね。

DNAレベルの研究によって、この遺伝子異常があるかどうかという病気になるとわかっているものが2000くらいあって、それぞれに名前がついています。当時はそれが分かり始めた頃で、自分の名前がついた病気を見つけないという思いもありましたね(笑)。それは結局、かなわなかったですけども。

——こうして遺伝子がある程度解明されて転写制御がわかると、かなりの寿命の延長に貢献すると思うんですが、実際には人間は何歳くらいまで生きられると思いますか。

俗に人間は何もなければ120歳まで生きられるといいますが、個人的には日本人の今の平均寿命は限界に近いところきていのではないかな、という気はしますね。もちろん乳がんなど若い時に罹患する癌は社会にとっての損失も大きいので研究が進んで癌の生存率が高まることを期待しています。

冒頭でゲノム情報が全部わかった時の「メリットは」と申し上げたのには意味があって、例えば「あなたは65歳で認知症になって70歳で癌で死にます」と生まれたときに言われたらどうするんだろう、と。そこは非常にデリケートな問題です。

——先端研の特長である学際性を生かして、VRやコンピュータサイエ

ンスだけでなく文系の研究者とのタイアップも必要になってくるかもしれません。

おっしゃるとおりで、そういつている間にもテクノロジーはどんどん進んでいます。生まれた時にあなたの遺伝子型のタイプはこういう病気になる、と言われて治療が無いのに診断だけされたら自暴自棄になる人も出てきかねません。もっとも、この配列だったらこうなるというのはあくまでも確率論の話で、いくつもの因子が絡み合って最終形質を作るわけです。個別のDNA型と表現型にどれくらいの相関があるかという計算はできませんが、実際は一つではなくて複数の因子が絡んできます。 3×10^9 個のゲノム配列全体のどの組み合わせが絡んでくるかとなると計算できないので本当のところはわからないんですけどね。それにも拘わらず、世の中の人は恣意的にとった10個でメタボリックだ、糖尿病だと診断をして、このサプリを飲んだ方がいいと勧めるような商法が世の中にはいっぱいあります。そのような情報には惑わされないようにしていただきたいですね。

最後に話がそれてしまいましたが、先端研にヒトゲノム解析という新しい分野の研究に相当の支援をさせていただいたお陰で、これまで研究が発展させられたことに心から感謝しています。他の研究所では絶対に実現出来なかったことだと思います。そういう意味では今後もそういう新しい分野に開いた先端研であり続けて欲しいと願っています。

参考文献 1 :

Wendt KS, Yoshida K, Itoh T, Bando M, Koch B, Schirghuber E, Tsutsumi S, Nagae G, Ishihara K, Mishiro T, Yahata K, Imamoto F, Aburatani H, Nakao M, Imamoto N, Maeshima K, Shirahige K, Peters JM. Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature. 2008 Jan 30; [Epub ahead of print]

参考文献 2 :

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, Gonzalez JR, Gratacos M, Huang J, Kalaitzopoulos D, Komura D, Macdonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, Zhang J, Zerjal T, Zhang J, Armengol L, Conrad DF, Estivill X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurler ME. Global variation in copy number in the human genome. Nature. 444(7118):444-454. 2006

(2007年10月24日)

略歴

1980年3月
東京大学医学部医学科卒業

1982年6月
東京大学医学部第三内科入局

1988年8月
米国マサチューセッツ工科大学癌研究センター研究員

1999年3月
東京大学先端科学技術研究センター助教授

2001年9月
東京大学先端科学技術研究センター教授

2002年4月

東京大学国際・産学共同研究センター教授 (-2006年3月末)

関連情報

ゲノムサイエンス部門のサイト

<http://www2.genome.rcast.u-tokyo.ac.jp/>

先端研 システム生物学ラボラトリーのサイト

<http://www.lsbm.org/>

東京大学先端科学技術研究センター

<http://www.rcast.u-tokyo.ac.jp/ja/>