

機能ゲノム科学—生命メカニズムの理解にむけて—

小原 雄治

国立遺伝学研究所

遺伝学研究所からまいりました小原と申します。

環境ホルモンは若干、共同研究はしているのですが、得意ではありませんのでちょっと困惑しておりますが、今日は *C. elegans* のファンクション・ジェノミクスの現状をお話ししたいと思います。

C. elegans は多細胞生物の中で一番早くゲノムが決まりましたが、ずっと前からこういうゲノム情報はコミュニティにシェアされ、それでゲノム研究が進んでまいりました。そのあとショウジョウバエも決まり、概要ですがヒトも決まりました。この間、遺伝子の数が非常に問題になってきたことはご承知のとおりだと思います。例えば線虫よりもハエの方が少なかったという驚くべきことがありましたが、ゲノムは非常に巧妙な仕組みで、いろいろな遺伝子の使い回しをすることによって何とかやっているようです。その仕組みを知りたいというのが、私たちの興味です。

進化を振り返りますと、ゲノムの構造は多細胞生物の原型のようなものから、それがミューテーションを起こし、かつゲノム全体でデュプリケーションが、おそらく 2 回起こり、そのあとさらにチェンジが起こって、現在の我々に至っていると考えられております。したがって今、井上先生がおっしゃったように、いろいろな種間のゲノムの比較をすることが、これからおそらく生物の理解に最も早道だと私自身は思っているのですが、そういうことでゲノム研究が進んでいます。そういう意味ではハエあるいは線虫というのは、ヒトゲノムのプロトタイプともいえると思います。*C. elegans* とはイトミミズのようなもので、バクテリアを食べて生きています。このように S 字型の行動をしますが、長さ約 1 mm 程度です。たくさんの線虫がありますが、その中で *C. elegans* という種を私たちは使っています。この体はたった約 1000 個の体細胞からできており、その分裂パターンは非常に厳密に決まっています。その中で約 3 分の 1 が神経細胞です。非常に簡単ではありますが、動物としての基本的なパターン、三胚葉構造 (triploblastica) があるわけです。ですから、動物のモデルになるわけです。

これは体が透明で、このように顕微鏡の下で発生をずっと見ることができます。今、受精が起こり、卵割が起こって、どんどん細胞分裂が起こっていきます。この間に細胞の分化、およびこの辺に腸ができてきますし、この周りには筋肉が分解してきますし、口ができてきます。こうやって発生が進みます。ここまで、大体 8 時間ぐらいのイベントを速回しでお見せいたしました。

最近のテクノロジーでは、GFP を使って、こちらはヒストン GFP ですが、ゲノムの動きや分配の様子を生きのまま見ることができます。これにミューテーションを入れることにより、いろいろな細胞の分化の過程、あるいは非常に微妙 (subtle) な影響を見ることができるわけです。

ヒトと違い第一分割で、ヒトの場合は両方もし分ければツインになりますが、線虫の場合は前と後ろは全く運命が違います。このあたりがどのように変わっていくかがポイントです。こういうことの全貌を知るために、私たちはゲノムの構造が決まったあと、遺伝子の構造を確定すること、さらにその各遺伝子が、発生のいつ、どこで、どの細胞で発現し、使われるのか、そこで何をするのかをシステムティックに明らかにすることを進めています。これが、ファンクション・ジェノミクスのポイントです。究極の目標は、コンピュータ上でこの発生を再現することですが、このためのデータを取ろうとしているわけです。

このために、最初に遺伝子をどんどん取るために、いわゆる EST プロジェクトで、cDNA をランダムに解析していくことをやり、すでにおよそ 1 万 9000 のプレディクトされた遺伝子の中から、大体 1 万程度の cDNA を同定しています。これを使ってゲノムとの比較から、エキソンイントロン・バウンダリーの決定、さらには遺伝子構造の確定、あるいはその予測の修正がどんどん行われています。

これはカルテックで作られております WormBase というデータベースですが、ここにすべてのゲノム情報、あるいはこういうフェノタイプの情報が集まっています。

発現に関しては、私たちは特に遺伝子からの最初のイベントであります mRNA (メッセンジャー RNA) の合成が、どの細胞でいつ起こるのかを徹底的に明らかにしようということをやっております。このためには、

線虫が非常に小さくて薄いからです、丸ごとの生物を使って、そこに *in situ* hybridization という方法で、mRNA のありかを探ることを非常にシステマティックにやっております。対象が 1 万遺伝子分ありますから、これを全部やるのは、常識的には非常に大変なのですが、簡便な方法を作成し、このようにやっております。

これは典型的なパターンで、一つの視野の中に、このようにいろいろなステージの胚 (embryo) が見えます。この遺伝子の場合、2 細胞からメッセンジャーが見えます。ですからマターナル発現ですが、このように発生が進みますとこれはなくなっていく。こういうパターンを示すわけです。こういうものがたくさんあり、1 万個分あるわけですから、テクニシャンを教育して、どんどんデータを取って、これにアノテーション (注釈) をしていくことを進めております。

ここが非常に大変で、例えばこれはそういう中から胚発生 (embryogenesis) に関して、発生のタイムコースに沿って並べたものです。

これは五つの違った遺伝子をやっている結果です。例えばこの遺伝子がこのようにあり、実は上皮細胞 (hypodermis) にスペシフィックな発現を示します。2 段目は筋肉の細胞 (body of muscle) に非常にスペシフィックな発現を示します。一方、この 3 つは全部が腸を作る系譜の細胞で発現しているわけですが、この遺伝子は最初はなく、この時期に初めて出てきます。ですからこの時期に転写が始まったといえます。一方、この一番下の遺伝子はずっとあとです。実はこの細胞が分裂をしてこれができるわけですが、この時系列の中の、最後の方で出てきたことがわかります。

こういうパターンをずっと並べてみたのがこれで、これまでに約 7500 の遺伝子に関して、実は 7500 行のアノテーションが並んでいるわけですが、このようにたくさんの発現の点を書いてあるわけです。これをいわゆる遺伝子発現クラスタリングという処理をおこない、よく似た発現パターンの遺伝子群を集めてくることをどんどんやります。

そうしますと、例えばこれは腸を作る細胞系譜でスペシフィックに出ている遺伝子が集まってまいります。そしてそれを発現の順にソーティングしますとこのようになっており、この一群の遺伝子は非常に初期に転写が始まっています。それ以前はなかったわけですから、転写の開始の時期を同定 (define) できるということです。

一方、これらの遺伝子群は最初はなく、このあたりから出てきます。しかし、同じ細胞のこの子孫細胞に出てきます。ですから逆に言うと、それぞれ一群の遺伝子が、それぞれの細胞で違ったレギュレーションを受けているわけです。

これをマップしますと、例えばこの時期、この水色で示した遺伝子はゲノム上のこういう場所にあるわけですが、これはこの時期にまさに転写が始まります。ピンクで示した遺伝子はこの時期に始まります。これは実は原腸陥入 (gastrulation) というステージですが、この時期に初めて始まるものがこういうものです。同様に、同じこの細胞の子孫細胞で順番にこの緑で示した遺伝子が発現しますし、このように順番に発火していくことになっています。このことによって、この場合は腸という組織ができていくことがわかります。

問題は、こういう遺伝子がそれぞれ、どういうレギュレーションを受けているのかということです。ここでは、おそらく可能性としては共通のレギュレーションを受けていることが考えられます。そのために、レギュレーションを受けるためのモチーフを検索します。もちろん、実験的にやるのが大事なのですが、情報学的にも研究が進んでおります。

これは腸の系譜で出ているスペシフィックな遺伝子ですが、それをポジティブグループとし、ほかでは出ているけれども、この赤いところでは出ていない、そういう一群の遺伝子をネガティブグループとして集めてきます。それで、こちらの遺伝子の上位にはあるけれども、こちらにはないという、そういうモチーフを探していくことを情報学的にやりました。モチーフというのは非常に短いものですから、7 ベース、8 ベース、9 ベース、いろいろなサイズを調べていきます。こういうものをウインドウ・サーチというのですが、ずっとなめていきます。ゼロは、この場合は転写のほぼ開始線をそろえています。

そしてポジティブにはたくさんあるけれども、ネガティブにはなるべくないというものを探してきます。しかし、短いモチーフですと当然、一定の確率で出てまいります。この 3 つのシークエンスモチーフは、ポジティブグループで必ず出てくるものでした。しかし、ちょっと見にくいですが、下の方のネガティブグループでもある程度出てきます。この 3 つをこういう条件で同時に持つ、「リンクしている」と言っているわけですが、

そういうものをやりますと、下は全部消え、上のポジティブだけになってきます。したがって、これは統計学的に非常に有意であり、この3つのモチーフの組み合わせというのが、この遺伝子発現に非常に重要であるという可能性が非常に高いわけです。これは実験的に確かめないといけないわけですが、こういうものが抽出されてきます。

これは、もう一つ筋肉を作る細胞でのモチーフ検索ですが、やはりこういうものが見つかってきます。当然これを個々でやりますと、ネガティブ・データでもあるわけですが、やはりこれをリンクして共通に持つものとし、ネガティブでは全部消えて、ポジティブだけになります。さらに、このモチーフをデータベース検索しますと、例えばこのピンクのものは、これは筋肉細胞の分化決定因子のバインディング・サイトに非常によく似ていることがありますので、非常にコンシステントな結果であるといえます。こういうことをして、今、各細胞系譜スペシフィックな発現を示す遺伝子群のレギュラトリー・モチーフを体系的に探しているところです。

発現パターンに関しては、やはりマイクロアレイもやっております。これは少し古いですが、マクロアレイといい、ほぼ1万のcDNAがメンブレイン (membrane) に乗っており、これで違った条件のメッセンジャーを使って比較をすることをしています。

これはフランスのジョナサン・ユーバンクという人との共同研究です。線虫に、ある種のバクテリアを感染をさせることができます。感染した結果どうなるかを彼は一生懸命見ているわけです。その感染前後、感染したあと、しばらくたったとき、遺伝子発現がどう変化するであろうかということ調べた結果です。これらの遺伝子が増えてくるのがわかりました。

私たちは遺伝子発現パターンを全部知っておりますから、これが出ると、ただちにこれは何であるか、すぐにわかります。これは口と腸の付け根にある細胞があり、ここで非常に強く発現しています。それからあと腸でも出ているのがわかります。このように、発現アレイと、あとの機能推測が非常にリンクしてできるようになっているわけです。当然DNAチップ、マイクロアレイも作成ができており、使用することが可能です。

最後にファンクションですが、線虫の場合はここにありますRNA i という方法があり、これはエクソンに対応する2本鎖のRNAを注入しますと、対応するメッセンジャーが壊されて、結果的に遺伝子ノックアウトをすることができるわけです。線虫の場合は、このように親虫に打ち込んで、あるいはRNA溶液につけるだけでもいいのですが、打ち込んだあと2~3日後に、そこから出てきた子どものフェノタイプを見ることによって、ノックアウト・フェノタイプを知ることができるという、非常に優れたものであります。

これまでに、これは東大の杉本亜砂子さんとの共同研究ですが、私たちの一連のランダムなcDNAのうち、これはたぶん2600ぐらいのcDNAについて、片っ端からRNA i をやってみた結果です。このように6本の染色体にまんべんなくやっていったわけです。これはフェノタイプの略号ですが、このようにデータがたまっています。

同様のことをイギリス、ドイツの研究者もやっており、線虫においては、すでにもう7000~8000、おそらく1万を超えているのではないかと思います。そういう数の遺伝子のRNA i のフェノタイプがたまっています。このうちかなりのものがWormBaseというデータベースにたまっていて、使うことができます。ただしフェノタイプというのは、よく見ないといけない、あるいは条件を変えないと見えないことがあるので、重複を許して、いろいろな研究者がいろいろな見方でRNA i をやって、フェノタイプを調べ、それを集積しているという状況であります。

私たちは特に発生初期に興味がありましたので、非常に初期に発現している遺伝子群の中から、それを選び出してRNA i をやりました。そうすると、普通は全くランダムにRNA i をやっても、フェノタイプが出るのはせいぜい10~20%なのですが、特に初期に限ってあるクライテリアを設けて選ぶと、このようにノンフェノタイプが35ですから、65%は何らかのフェノタイプが出たことになり、中身を非常に詳細に調べていくことができました。

そうしますと、一番きついフェノタイプは受精以後進行停止 (arrested as fertilized egg) ということで、受精の直後にもう止まってしまうと、そのあとの卵割等に行かない。それ以外のフェノタイプとしては、いろいろな核の動きなどがおかしくなる。そういうさまざまなフェノタイプを、時系列の順番に分類することができました。

例えばこの遺伝子は、これを打ち込みますとこのようになり、見た目は同じように見えるかもしれませんが、明らかに違うところがあります。こちらが野生型で、このように大きい細胞、小さい細胞というように分かります。これが非対称分裂と呼ぶのですが、この非対称になることによって、前・後ろが違う運命をたどることができます。これは中身が変わってくるわけですが、こちらの卵は全く等分裂、イコールの分裂をします。こうしますと前後の違いがなくなり、そのあとも実は全く等分裂、真ん中が割れてこのようになります。一方、こちらは前の方がどんどん割れていくというように違っておりますので、要するにこの遺伝子の場合、非対称性をなくしてしまうという効果があります。したがって、これは発生が進まず、細胞の分化ができません。

この遺伝子をノックアウトしますと、受精直後に全く卵割が起こらなくなってしまいます。ちょっと細かくなりますが、受精したあと、雌核の減数分裂が起こらなくなってしまいます。この遺伝子は何かという、実はプロテソームのホモログであり、プロテソームの一つのサブユニットであり、こういうことが起こることがわかりました。逆にプロテソームの他のサブユニットを全部調べますと、やはり同じフェノタイプを示すことがわかりました。このようにフェノタイプを分類することによって、さらにこういう細胞内小器官のサブユニット構造を、逆に明らかにすることもできるようになっております。

最後は、コンピュータ・モデリングですが、これは世界中の人が、特に最近システムバイオロジーという言葉がされておりますが、たくさんのコンピュータ学者が、一つのモデルケースということでトライしており、これから非常に楽しみな分野です。私たちはこの分裂パターンを、取りあえず座標系をきちんとしようということで、まずコンピュータ・グラフィックスを作りました。これはこのワイルド・タイプの動きをそのまま測定し、このようにコンピュータ・グラフィックスにしたわけです。ですから、これは細胞のコンタクト、あるいは場所などが全部測定されているということです。

それに基づいて、そういうものをミミックする、それを再現するようなモデルを作らないといけないわけですから、それを作っています。これはまだプリミティブですが、膜をこういう青い粒々のメッシュで表し、あとは力学モデルで割れていくものを作っています。まだ4細胞に至るまでですが、ここでぐるっと回って、これが回ることが非常に重要なのですが、そのように上手に回るようなパラメータを設定しているところです。

こういうことで、いろいろなことをやっていますが、こういうデータを全部まとめて、私たちのところではNEXTDBというデータベースにして、このURLから公開をしています。一方、WormBaseとリンクしており、いずれにしても世界中から見えるということです。

以上で終わりですが、アクノレッジメントはしてはいけないという話がありましたが、たくさんの人がかんでおります。これはすべて線虫コミュニティは、データは公開してシェアをすることが伝統であり、そういうことに則って、コミュニティと協力をして進めているということです。以上、線虫のファンクション・ジェノミクスの現状をご紹介します。ご清聴ありがとうございました。

質疑応答

ブルンバーク：小原教授、大変ありがとうございました。1つ、2つ質問していただく時間がありますので、どうぞ。

曾根：国立環境研究所の曾根と申しますが、*C. elegans* のことがよくわからないのですが、発がんとか、*C. elegans* にはがんができるのでしょうか。寿命が短いとか、細胞数が少ないということで、発生や分化にはいいモデル、いい系だとは思いますが。

小原：線虫にがんはあるかというのは、どこでも必ず聞かれる質問なのです。そういう意味ではがんはありませんが、がん遺伝子がありまして、例えばそれがミューテーションを起こしますと、産卵管、産卵器官がたくさんできたり (multi-vulva)、そういう細胞の分裂回数が非常に増えるといった現象でもって、異常が起こることがあります。それは一つのがんです。全部がん遺伝子のキャスケードで説明できますので、それが一つの表れだと思えます。

井上：質問の方はご自由にマイクروفンの前にお立ちください。

ブルンバーク：はい、もう一つだけ質問をどうぞ。

質問：線虫の場合は、何か薬剤に対する反応性というか、毒性は何か？

小原：それは薬剤に対する反応性ですが、線虫は特に企業などでは薬剤のスクリーニングに非常に使われています。神経毒あるいは農薬とか、当然、線虫も駆除される側ということもありますので、そういう研究はたくさん進んでいます。僕はあまりファミリーではありませんが。

最近、今日はお話ししませんでした、たぶんそこにすわっておられる有菌先生を中心にしたグループが、線虫を使って環境ホルモンにさらしたあと、DNAチップを使って、その遺伝子変動をする実験を今やっており、そのうちいろいろな結果が出てくるのではないかと思います。