第 28 回埼玉医科大学医学会総会

第 4 回公開シンポジウム

ゲノムからみた医学

－埼玉医科大学ゲノム医学研究センター開設を記念して－

平成 13 年 12 月 14 日（金）16:00–19:30

埼玉医科大学第三講堂

総合司会　禾　泰壽（埼玉医科大学分子生物学）

ゲノム医学研究の目指すもの

村松 正實（埼玉医科大学ゲノム医学研究センター所長）

マウスゲノムエンサイクロペディア

林﨑 良英（理化学研究所ゲノム科学総合研究センター）

多因子疾患における疾患感受性遺伝子のポジショナルクローニング
－NIDDM 1 を例にして

堀川 幸男（群馬大学生体調節研究所）

癌の機能ゲノミクス解析

油谷 浩幸（東京大学先端科学技術研究センター）

ゲノム研究の倫理的側面
－埼玉医科大学倫理委員会における審議の実情と今後の問題点

山内 俊雄（埼玉医科大学倫理委員会委員長・精神医学）

司　会（禾）

今回は，本学の日高キャンパスにありますゲノム医学研究センターの開設を記念して，「ゲノムからみた医学」というタイトルのシンポジウムを企画しました．5名の先生方を講演者としてお招きしました．

皆様ご存じのように，本年2月にはヒトゲノムのドラフト配列が発表され，2003年にはヒトゲノムの完全配列が決まる予定になっています．このヒトゲノム解析結果は，すでに日進月歩の勢いで医療に応用されつつあります．

本シンポジウムでは，現在行われている最先端ヒトゲノムの研究の医療への応用例を講演していただこうと思います．同時に，そのような進歩あるゲノム医学の中で，本医学部はゲノム医学研究センターを中心に，どのようにゲノム医学研究を展開していくかを講演していただきたいと思います．

本記録は，平成13年12月14日に行われた埼玉医科大学医学会総会第4回公開シンポジウム「ゲノムからみた医学　－埼玉医科大学ゲノム医学研究センター開設を記念して－」を採録したものである．

シンポジウム

ゲノム医学研究の目指すもの

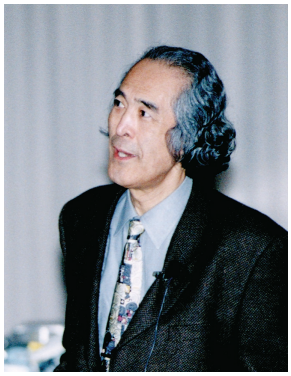
村松　正實
（埼玉医科大学ゲノム医学研究センター所長）

座長　禾　泰壽（埼玉医科大学分子生物学）

それでは最初の講演に移りたいと思います．恒例に従い，村松正實先生の略歴をご紹介致します．

村松正實先生は，昭和30年に東京大学医学部を卒業後，内科に入局．その後，米国ベイラー医科大学に留学され，分子生物学への道を選ばれました．帰国後，癌研究所部長，徳島大学医学部教授を経て，昭和57年に東京大学医学部第1生化学教授に就任されました．平成4年，本学第2生化学（現　分子生物学）教授に招聘されました．今年になり，ゲノム医学研究センター所長に就任され，埼玉医科大学のゲノム医学研究をひっぱっていこうと意欲に燃えておられます．

それでは村松先生，お願い致します．



本日は我々の研究センターのために，わざわざ埼玉医科大学医学会総会のテーマとして選んでいただき，誠に有難うございました．また，このように多数の先生方もお見えで大変感激しております．

今日は，外からいろいろな先生をお招きして，それぞれの専門分野で，ゲノム及びゲノム医学の方向からお話をお願いするわけです．そのイントロダクションとして，今，ヒトゲノムはある程度わかったというけれども，どの程度しかわかっていないかという話を少しさせていただいて，それからゲノム医学研究センターの内容を少しばかり紹介させていただきたいと思います．

今日はお医者の方々もおられますが，学生の方もおられるようですし，非常に初歩的なところから出発して，その代わり，あっという間に少し難しいところへ入るかもしれませんけれどもお話をさせていただきたいと思います．

先ず，なぜDNAは大切なのかについでですが，それは，我々が患者さんを診て，体に異常がある際，多くの場合，どこか組織や器官に異常があります．その組織や器官は細胞からできており，細胞は大部分，その機能は少なくともタンパク質によって行われています．

ところが，このタンパク質の情報がDNAに書かれていてタンパク質を決めており，そういうことでDNAを知らなくては他もわからないということになるのです．

情報の流れは，ご承知のように，DNAからRNAという中間物質を経て，それから翻訳でタンパク質ができます．一方，複製と言って，DNA自体が複製することも出来るわけです．最初のDNAからRNAができるところを転写と言い，それはRNAポリメラーゼという大きな酵素がDNAに合わせてRNAを紡いでいくという形で情報を伝えます．そして，そのRNA（メッセンジャー RNA）の情報に従って，アミノ酸が繋げられタンパク質になります．そうして出来上がったタンパク質は，そのタンパク質のいろいろなアミノ酸に修飾が起こります．例えばシステインの間に結合が出来たり，いろいろなところに糖鎖がついたりして，一種の（これは2次元にしか書いてありませんが）立体構造でタンパク・タンパクの相互作用があって，それが生物の働きになることがわかっています．

ポスト・ゲノムという話は，あとから林崎先生あたりから出ると思います．

我々が昔習った頃も今も，大抵の人はDNAというのは遺伝を支配し，遺伝をするときだけ働いている

と思っています．しかしそうではなくて，我々の刻一刻の機能も，やはりDNAが支配しているということが非常に重要なことです．だから病気にもなるわけです．DNAは遺伝だけでなく，その生物の機能そのものも支配しているのです．

「ゲノムとは」という定義ですが，ある生物の遺伝子の総体を言います．全体と言っても良いのですが，総体と言った意味は，いろいろなコンポーネントがあるということを強調してそのように言いました．これは私の定義であります．

遺伝子はDNAという化合物からできており，通常，ある1つのタンパク質の構造を決めています．これは「通常」としか言えないので，このごろは大体1つの遺伝子が1つのタンパクを決めることは少なくて，数個のタンパク質を決めることが多いのです．遺伝子の中のいわゆるsplicingという機能が働いて，1つの遺伝子から数個のタンパク質を作ると計算されています．ですから，遺伝子が3万とか4万とか言われても，実際にできるタンパク質は10万を超えるであろうと考えられます．

生物種により，遺伝子の数，すなわちゲノムの大きさも異なり，大腸菌では約4,000というのはしっかりわかったようです．ヒトでは約10万，今はこの半分ぐらいと考えられており，少ない人は3万いくらというのですが，最近はどうも数え直したら6万ぐらいあったという説も出ているようです．この辺は実はまだゲノム計画は本当の意味で完成していないのでわかっていません．ということは，それだけの数のタンパク質を持っているということであり，生物種が複雑になればなるほどタンパク質の種も複雑になり，我々の行動や脳の機能が複雑になるわけです．

これはA・G・T・Cで書かれたSV40という小型ウイルスの環状DNAでそのゲノムです．これは5,300ぐらいのbp（ベースペア,塩基対）です．

要するに，ほとんど1番小さいウイルスのゲノムの大きさは5000いくらということです．

ヒトになりますと，そのゲノムは3×10の9乗塩基対，30億塩基対あると言われています．大体間違いないようで，いくつかのchromosome（染色体）に分かれています．大きな1番からだんだん小さくなってきて，22番が1番小さいのですが，そのほかにX・Yという2種類の染色体があって，全てのヒトはこういうものを1対持っています．お父さんから来たものとお母さんから来たもの，それにXとYを持っていると男に，YがなくなってXを2つ持っていると女になるというのが我々のパターンです．どの1つをとってもその中には何億という塩基対のDNAが入っているわけです．

生命の進化ということも頭に入れておいた方がいいのですが，哺乳動物は1億年以上前にでき，約6,500万年前に大彗星がぶつかって生物の90％が死にました．その前にもいろいろあるのですが，地球が45~46億年前にできて，生命が35億年ぐらい前に，そして20億年ぐらい前には真核生物ができて，哺乳動物はせいぜい1億年ぐらいではないかということです．

だから，我々ヒトなど，ホモサピエンスと言われるのは20万年ぐらいのことですから，本当に1ミリぐらいの歴史しか持っていないことになります．それだけ生命の歴史は古いのです．しかし遺伝コードを見ると，同じコードを使っているので，やはり生命は1つの起源から出ており，そこらを飛んでいるハエも我々も，同じ先祖からということは非常に重要です．それを人間だけは特別だと考えている人も何人かおりますし，そういう宗教もありますが，そうではないのです．

今まで我々科学者が得た証拠は，全ての生物は1つの原始生命体から出たものであることを裏付けています．

さて，ヒトゲノム計画とは何かというと，先程言ったように遺伝子の総体，全遺伝子DNAの配列，「A・G・T・C」で書かれていますが，それを全部読み取ろうという計画です．これは1989年から1990年にかけて，アメリカが全世界に先駆けて始めました．そして，昨年（2000年）に取敢えずdraft（概要）がわかりました．しかし未だギャップもあり，正確さも問題であり，まだまだはっきり全部わかったとはとても言えません．完全なものが出来るには早くとも，もう2~3年はかかるでしょう．このドラフトの方は去年，大体出来たと言って，今年，2つのグループからNatureとScienceに発表されたのですが，1つは，米国立衛生研究所（NIH）が主導して，世界の研究共同体を作り，日本も加わっております．もう1つは，途中から参入したけれど，ものすごいスピードでやった米のベンチャー企業で，セレラ・ジェノミクス社（もの凄い資本のバックアップがあったわけですが）と2ケ所で行いました．その結果はそう違わないので両方とも正しいのだろうと思います．

やり方は，例えば1つの染色体でもいいのですが，そのDNAはもの凄く長くて，そのままでは塩基配列は決められません．そこで，これをまず小さく切ります．小さいといっても，初めは100万とか50万塩基対，非常に大きな長いものに切って，それをプラスミドに入れBACライブラリという1つの集合を作ります．そしてこの集合の1つだけをたくさん増やして，その増やしたものをまたブツブツ切って1キロ塩基対くらいにします．このぐらい小さくなると，今度はようやくシーケンスができるぐらいの長さになります．

このようなことをやって，1番下でこのシーケンスがたくさんできるのを，端と端がつながるものを並べて次第に長くし，これをまた並べて，そして元がわかっていくという方法です．セレラ・ジェノミクスの方は，染色体に分けないで全DNAからのshotgun方式をとりました．

遺伝子塩基配列からどういう情報が得られるかというと，遺伝子の構造がA，G，T，Cのレベルでわかります．編成の全パターンの全部がわかることは非常に重要です．これ以下でもないし，以上でもなく，そういうことがわかることは，いろいろなことを考えるときに重要なのです．

それから，遺伝子ファミリーと言いまして，遺伝子が1つポツンとあるということはむしろ極めて少なくて，ある1つの遺伝子が見つかるとその兄弟がいっぱい見つかることが普通です．それを「遺伝子ファミリー」と呼んでいて，そういうものの構成・機能がどうやって進化してきたかということがわかります．

これは後でも言いますが，ヒトのゲノムだけやっていたのではだめです．先程も言った1つの生命の流れという動きがあります．今ここに生き残っているものしか見ることはできませんが，低い段階のものから，ヒトに向かって途中をいくつも取って，その間の変化を見ます．それで進化を見ることは非常に大切な情報を得ることに繋がります．

遺伝子上のいろいろな信号やモチーフ，シグナルなど，即ち転写のシグナルや複製のシグナル，いろいろなシグナルがわからなければいけません．これはタンパク質とは別の意味で，重要なインフォメーションです．そうすると，今度は遺伝子がいくつも見つかるわけです．おそらく数万以上見つかりますが，それの間のネットワークが非常に大切になってきます．DNA配列からタンパク質のアミノ酸配列がわかっても，その機能はすぐにはわからず，「さらに多くの分子生物学的研究が必要」となります．

そのようなことを手がかりとして，人間の最も重要な現象であります発生・分化，高次神経機能，脳の機能，医学的な方から見ると遺伝疾患の全貌が明らかになります．そういう各々の法則性を解明していくのが，今後のヒトゲノム，或いはポスト・ゲノムの方向であろうと思います．

先程少し言いましたが，ヒトのゲノムサイズは3×10の9乗，30億塩基対あります．大腸菌は約400万塩基対あります．大腸菌の場合，400万塩基対あって，その遺伝子の数は約4,000，ヒトは30億塩基対あって，その遺伝子の数は5~10万，数万だと思います．しかも，遺伝子からタンパク質を作る作り方が，大腸菌の場合は大体1対1ですが，ヒトは1対2とか3という値が出ていて，ゲノムは非常に多様化しています．しかし，この塩基対の中の遺伝子が占める部分は極めて少なく，非遺伝子部分が砂漠のようにべらぼうに多くて遺伝子の部分はオアシスの如く極めて少ないのです．ほんの10~20％以下という程度です．特にタンパク質をコードしている部分は確か1％程度です．

染色体は，ヒトの場合は22本＋X（又はY）です．一方，モルガンという単位（遺伝子間の距離）も組換えの頻度で決めたものです．ここにも遺伝子数は5~10万と計算しています．だから，今でもこんなものではないかと思います．

医学の方から見ますと，病気というのはこういうふうに考えられます．遺伝要因が強い病気と環境要因が強い病気があります．遺伝要因だけでほとんど100％起こるのは，例えばハンチントン舞踏病とかテイザックスで遺伝子を持っていれば100％です．もちろん，劣性の場合と優性の場合がありますし，優性の場合はペネトランス（透過率）とか，いろいろな条件が関与してきますが，とにかく遺伝でほぼ決まります．

1番，環境要因が強いのが外傷です．火傷や車にはねられれば誰でも外傷を起こすわけで，これは100％です．ところが感染症ぐらいになると，もはや環境要因だけではありません．昔はコレラやペストに罹るのはバクテリアによるので，これは全部環境要因と思っていたのですが，そうではなくて罹りにくい人もいるのです．例えばアフリカでエイズに罹らない女の人が見つかりました．レセプターの異常でエイズウイルスが入れないことがわかり，こういう感染症も環境要因だけでなく遺伝要因が関係していることがわかってきました．ヨーロッパでペストでたくさん死んだときも，おそらくペストに罹っても大丈夫な人が生き残ったのではないかと，むしろ考えられます．

この中間ぐらいに生活習慣病があって，高血圧・糖尿病・高脂血症等々，この辺は遺伝要因半分，環境要因半分です．しかも面白いことに，前者は遺伝子が大体1つと決まっているからすぐわかります．非常にはっきり遺伝で決まるのですが，後者は遺伝子が1つではなく，いわゆるmultigenicでたくさんの遺伝子が関係しているという複雑な病気です．しかも，関与する遺伝子によって，環境もどのぐらい影響するかということが動きます．ということは，遺伝子が病気を決めるというよりは，病気に対するsusceptibility（感受性）を決めていると言えます．

2月に出た確かセレラの論文に“Paradigm Shift in Biomedical Research”と書いてあります．これは英語だけれども学生の人も読めると思います．つまり今ま

ではStructural Genomicsで構造だったけれども，だんだんFunctional Genomicsで機能に入っていき，それから，ゲノムの全体を知るということは，今度はプロテオミクス，タンパクも全部知らなければならないということです．

それからmap-based discovery，これはgene（遺伝子）のマッピング，それから，いわゆるcandidate approachという方法で，ジーン・ディスカバリーを行っていくのです．このシーケンスができてしまうと，もうマップしなくても，ある遺伝子が取れたときにシーケンスを少しすると，それはゲノムの上のどこだとすぐわかるわけです．こういうのは例えばcloning in silico，シリコンの中で，計算機上でクローンするのですが，そういう時代になっています．sequence-based gene discoveryになっていくだろう．

それから，先程言った環境との関係ですが，問題の焦点がmonogenic disordersからmulti-factorial disordersになってきます．疾患感受性があって，それがいくつか溜まって，環境の影響で起こるものが非常に多くなります．DNA診断は，だんだんにもっと大きな意味でのsusceptibility（罹りやすさ）の診断になります．こうなってきますと最近流行りのスニップ（SNP：Single Nucleotide Polymorphism）が重要な役割を果たします．

それから，今でも必要があるのでやっていますが，昔は1つの遺伝子の構造を十分明らかにすればよかったが，今後は更にたくさんのジーン，或いはgene familyやpathwayやシステムを明らかにしていかなければならない時代になります．そういうgene actionやgene regulationの方向に物事は進んでいます．

進化の研究の重要性を考えると今まではヒトという1つのspeciesであったのが，今後はseveral species，他の多くのspeciesのゲノムの解明も必要になってくるということです．

医学の近未来像．今でもこの予想は変わらないと思います．最も期待されるテクノロジーの1つは画像診断です．今でも進歩しており，CTでいろいろなことがわかり，ヘリカルCTで，肺癌などが凄くよくわかるようになっています．

診断へのコンピュータの応用．いろいろな数値が，頭の中でどうしようもないのを，多変量解析で何かに収れんさせることが可能になって来るでしょう．

人工臓器．人工肝臓や腎臓は再生医学の進歩が必要でしょう．埋め込み型の心臓ではまだうまくいっていないようですがこれは材質の関係がもっと発達すれば，できるようになるでしょう．

遺伝子の応用．いよいよこれが出来るようになり，最初はインターフェロンなどサイトカインを大腸菌に作らせました．私もクローニングしましたし，そんなものを作るのに使われましたが，いまや診断・治療に遺伝子が使われ，それからタンパク質工学で，いろいろな新薬，新素材が出来るようになっているのが近未来であるということです．

ここから先は，新しくできた研究所をご紹介したいと思います．

「埼玉医科大学ゲノム医学研究センター」というのは，私が名前を付けました．‘Genomic Medicine’というのを思い付いたとき，私は数人のアメリカの友達に，この名前はどうだと聞いたのです．みんな賛成してくれまして，まだアメリカでさえこういう名前のリサーチセンターはないのだから，日本が早く作るのはいいだろうと言ってくれたので，自信を持って付けさせていただきました．

今，先に述べたゲノムというものを背景にして，その情報・技術を基に医学を研究する分野と私は定義づけております．ですから，これにひっかかるのであれば何でもいいので，必ずしもシーケンシングをじゃんじゃんやらなければいけないとか，そういうことではないのです．その点はお間違えのないように．

今5つの部門があります．実は6つ作ろうと思っているのですが，最後の1つは未定であります．

1つ目は「遺伝子構造機能部門」で名前はかなり一般的ですけれども，実際，伊藤敬助教授が担当しており，そこに人が何人か集まって始めております．主としてクロマチンの構造と機能を解明するのが目的です．クロマチンというかたちで遺伝子転写が支えられているわけですが，それがまだまだ謎に満ちており，その辺を解明していきます．これは病気の解明にも繋がります．

2つ目は「遺伝子情報制御部門」これもあまりにも一般的な名前ですが，東大の老年病科の井上聡講師に来ていただき，部門長になっていただいています．こちらはエストロゲンその他のnuclear receptor及びその標的遺伝子を主としてやっています．

3つ目は「発生・分化・再生部門」で奥田晶彦助教授が担当しており，数人で始めております．今，問題のES細胞（embryonic stem cell）を扱って，それがembryonic stem cellであるためにはどういう条件，どういう機構が働いているかを研究しています．

4つ目は「病態生理部門」です．ここは須田立雄教授が担当しておられます．須田先生は昭和大学を定年になられて，こちらへ来ていただきました．彼は朝日賞もお取りになったビタミンD，骨代謝の大家であります．ここは大きなチームで製薬会社との共同研究も

行われております．

5つ目は「神経科学部門」ニューロサイエンスです．実はこの名前もみんなで相談して付けたのです．広常真治助教授が担当して，脳ができる時の発生における細胞の動きを何が調節しているかを主に追究しておられます．1種の病気で，いわゆるジャイラスがなくなるような病気などの遺伝子を取ったり，いろいろな新しい仕事をしておられます．本学の精神医学との共同などもすでに始まっています．

最後（6つ目）に「遺伝子治療部門」を計画しておりまして，これは近々人を決めようと思っております．

以上で大体の私の紹介のお話はおしまいにします．どうもありがとうございました．

座長：村松先生，どうもありがとうございました．今日のシンポジウムの最初にすばらしいイントロダクションをいただき，またゲノム医学研究センターの紹介までいただきまして，ありがとうございました．

座長：時間が詰まって参りましたので，直ちに次のセッションに移りたいと思います．では座長を交代します．本日，最後に討論の時間を取ってありますので，最後にまた自由なかたちで討論したいと思います．

シンポジウム

マウスゲノムエンサイクロペディア

林﨑　良英
（理化学研究所ゲノム科学総合研究センター）

座長　松下　祥（埼玉医科大学免疫学）

それでは，林﨑良英先生をご紹介させていただきます．私は本学免疫学の松下でございます．

林﨑先生は，現在，理化学研究所ゲノム科学総合研究センター遺伝子構造機能研究グループの，プロジェクトチームリーダーでいらっしゃいます．先生は82年に阪大の医学部のご卒業，86年に大学院終了，92年から理科研のライフサイエンス筑波研究センター・ジーンバンク室研究員，ヒトゲノムプロジェクトの推進室，ゲノム機能解析研究グループ・プロジェクトリーダーを経て，98年から現職に就いていらっしゃいます．

先生は賞を各種，受賞なさっておりますが，98年には東京テクノフォーラム賞ゴールドメダル，遺伝子辞書の作成，今日お話しいただくタイトルと非常に近いものですが，2001年にはつくば賞を受賞しておられます．

本日は「マウスゲノムエンサイクロペディア」という演題でお話をいただきます．どんな百科事典ができつつあるのか，私も非常に楽しみに聞きたいと思っております．林﨑先生，よろしくお願いいたします．

ご紹介ありがとうございます．まず，埼玉医科大学のゲノム医学研究センターの開設，おめでとうございます．今日私がお話しするのは，このところずっと私はこのタイトルでお話ししていますが，実は内容が年度ごとに進行して変わっております．

「マウスの遺伝子の辞書」というタイトルで1995年から，日本の中でゲノム科学を再建せよということを，理研の当時の理事長であった小田稔さんに言われまして，何をやったらいいかを考えて，ずっとやってきたのがこの仕事です．

当時，とにかく再建するために，まずそこにおられます村松先生が理化学研究所の顧問になられ，それで今のようなプロジェクトをとにかく実施することが可能になったわけです．

今日はスライドを全部英語で書いてあり，訳しながら話をしますので少し遅くなります．後半はひょっとしたら飛ばすかもしれません．とにかく，どういうものができて，どういう使い方ができるかがわかればよろしいかと思います．

私たちが最初狙ったのは1995年と申し上げました．その95年には，世の中のゲノム科学はどういう方向になっていたかというと，アメリカ合衆国はヒトゲノムのシーケンスを実行することを決意した年であります．

そのほか，1990年代初頭から，製薬会社にサポートされたアメリカのベンチャーによって，EST（Expression Sequence Tag）という，RNAになったその一部分のかけらのcDNAを山のようにシーケンスするというプロジェクトが，すでにもう終わっていました．さて，私たちは何をしたらよいか，もうやるものはないのではないかと思ったのですが，こういうものをやることにしました．Full-length cDNAをやろう．「完全長cDNA」といいます．

それはRNAの端から端までの全長を含むものです．なぜこの完全長cDNAが重要なのかといいますと，1つは，RNAの完全なかたちがわかる．タンパクの全体のかたちがわかる．もう1つは，これは全長を含んでいますので，直接それでタンパク質を発現できます．かけらのDNAをコンピュータ上でいくらつないでも，それは物質としてつながっていませんので，実際にタンパクを作ることはできないわけです．ですから，タンパクを作ることにも重要である．

それから完全長cDNAは，染色体のDNAと比較することによってプロモーター，要するに転写を制御している領域がわかる．もう1つは，ゲノム・プロジェクトというのは，先程，村松先生が言われましたが，ショットガン・シーケンスというやり方があります．そのショットガン・シーケンスをつないでいくのに，このDNA，cDNAそのものが，それをつなぐのに有用な情報を与えます．そういうことで，これをやろうということになりました．

生体内に存在する全部の遺伝子を，まるごとフルレングスのかたちで取ってきたものを集めてバンクを作って，その構造を決定しようというのが私たちの目的であったわけです．

それをやるためには，まず当時何もなかったので，技術から作らなければいけないということで，完全長cDNAを作る．それからハイスピードのシーケンシングをするシステムを作る．それを用いてマウスゲノムエンサイクロペディアを作ることにしました．（図1）

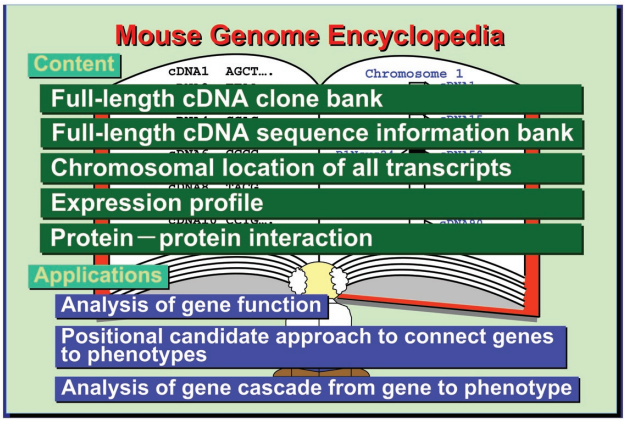
我々のこのマウスゲノムエンサイクロペディアは，全部の遺伝子をかき集めるというのですが，全部の遺伝子の内容というこの辞書には何が載っているか．それは完全長cDNAのクローンのバンクがあります．もう1つは，完全長cDNAのクローンのバンクを全遺伝子について集める．それから，それのシーケンスを全部決定する．

その完全長のcDNAとは，mRNA（転写単位）のことをいいますが，転写単位が染色体のどこにあるかを決定する．もう1つは発現タンパクになる．すなわち，その遺伝子がいつどこでタンパクになっているのか．mRNAになってタンパクになっているのかを見るための，発現プロファイルを明らかにする．

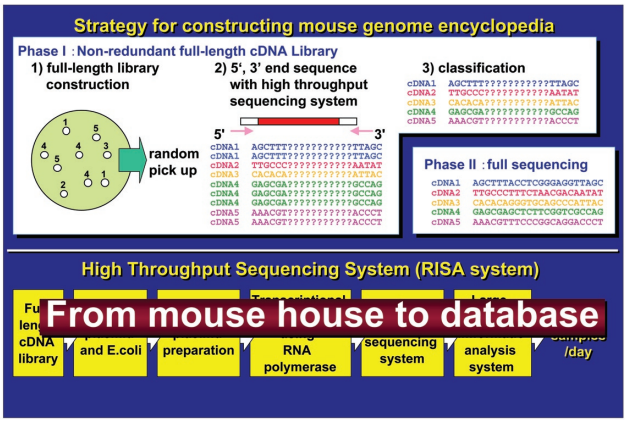
もう1つは，タンパクは何もそれ自身で機能するというのではなく，互いに相互作用しております．あるタンパクがあるタンパクに情報を伝えるためには，タンパクとタンパクの相互作用を明らかにしなければいけない．そういうことで，全遺伝子のどのタンパクとどのタンパクが相互作用するかというのを，ラフでもいいから，とにかく作ろうということにしました．これによって遺伝子の機能がわかります．

あと，疾患の遺伝子を見つけるためのアプローチの方法ですが，染色体の位置情報から見るやり方で，positional candidate approachというやり方があるのです．そのやり方は，フェノタイプ（phenotype：表現型），つまり病気と遺伝子を関連付けするために，非常にいい役割をするであろう．もう1つ，遺伝子がつぶれると，なぜ病気が出るかということのパスウェイが，これをやってわかるであろうということです．

どういう戦略を取ったか．まず，完全長cDNAのテクノロジーを作って，高品質の完全長cDNAを作る技術を作って，プレートの上にクローンをまきます．大腸菌のクローンをランダムにピックアップして，両端からシーケンスを組みます．そうすると同じクローンが2回出てきたり3回出てきたりします．それを除いて重複（redundancy）のないcDNAを作ります．（図2）



（図1）



（図2）

このcDNAの代表選手を選んで全長を決めるというのが，その次のフェーズです．端からこのシーケンスを全部出す．しかも，ほとんど全部の遺伝子をカバーしようと思って，しかも2～3年でそれをカバーするためには，1日4万個のサンプルを処理する操作が必要です．そのために，こんなシステムがなかったの

で，自分たちで作ろうと．今でこそゲノム科学のいろいろな装置がありますが，当時はありませんでしたので，それを自分たちで作ろうということです．

これはマウスを使ってやろうということで，マウスからデータベースを結ぶパイプラインとして，「RIKEN Integrated sequence analyser research sysytem（RISA system）」を作りました．

まず，完全長を得るための完全長cDNAをmRNAから逆転写するためのテクノロジーですが，これをelongation methodということで，伸長反応で伸ばすわけです．こういうテクノロジーを開発しました．（図3）

これはどんなテクノロジーかといいますと，普通，逆転写酵素はステムループ，RNAの二次構造のところで止まります．これは電話線がねじれているような格好です．そういうところで止まるわけです．これがあるから止まるので，これを止まらないようにするためにどうしたらいいか．

温度を上げればいい．分子運動が大きくなっていくのですが，温度を上げると逆転写酵素が失活します．そこで我々が見つけた発見ですが，トレハロース（trehalose）という二糖類を入れると，高温でも逆転写酵素が活性化して，端まで行くことを発見しました．

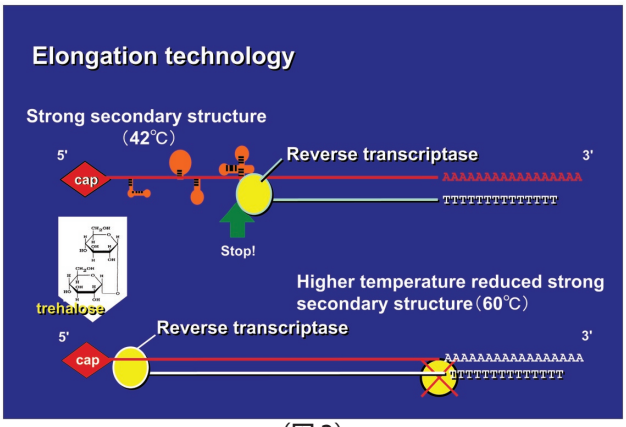
何でそんなことが見つかったかということですが，やけども卵焼きもそうですが，タンパクは熱をかけると変性します．高次構造が崩れるわけです．細胞の中でこういうものが起きますと，ものすごく毒性が強いので，それを元に戻す機能を持つシャペロニン（chaperonine）という物質があります．そのシャペロニンという物質が世の中で知られているのですが，作業仮説として，こういうふうに折りたたみを元に戻すような物質が，反応液の中にある．ひょっとしたら，酵素としますと，酵素を守ってくれるかもしれないというので，そういうものを探したわけです．（図4）

こんな文献がありました．トレハロースはこんな物質ですが，これが酵母菌にヒートショック，酵母菌をやけどさせます．そうするとトレハロースをいっぱい作るようになります．なぜだろうと思ったのです．

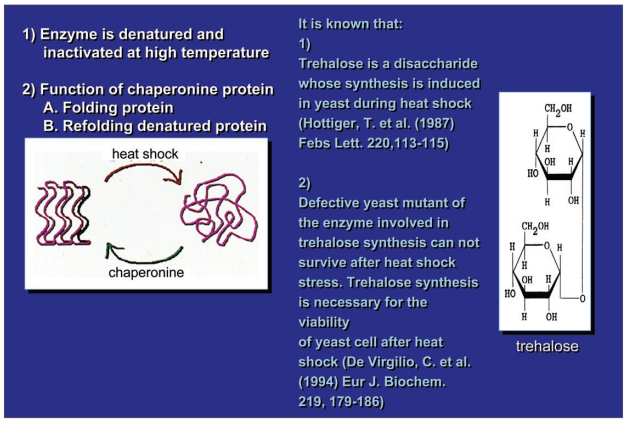
もう1つ，このトレハロースの合成型酵素が欠損したmutantにheat shockをかけると，全然生き返ってこないで，そのまま死んでしまう．heat shockに非常に弱いわけです．

そういうことから，シャペロニンを「物質ではない」と考えて実験しました．実際，5 KbのRNAを鋳型にして，DNAを合成することをしたのですが，正常の反応でやりますと5 Kbのバンドは見えますが，こういうパーシャルなcDNAがいっぱいになります．ところが60℃にしますと酵素が失活しますのでバンドがなくなります．ところがトレハロースを入れると，酵素は失活しないで，ちゃんと5 Kbのバンドが見えるけれども，パーシャルのcDNAがなくなるということで，非常に効果的である．（図5）

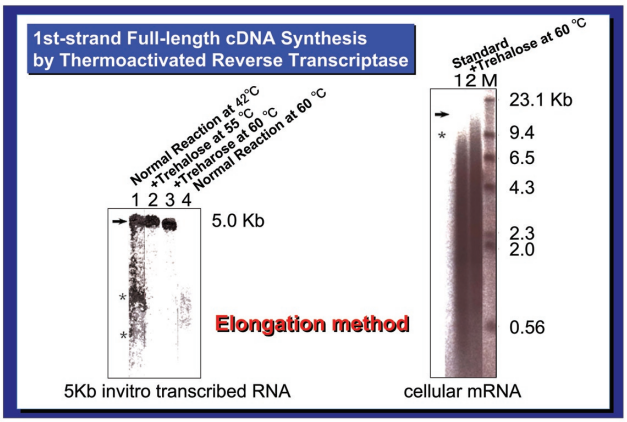
しかも，こういうものを入れますと，60℃でこういう反応をする．トレハロースを入れますと，これはcellular mRNAで，一番長いのは，普通の条件としては9 Kbぐらいですが，16 Kbぐらいまで伸びるようになるということで，非常に効果的であることがわかりました．



（図3）



（図4）



（図5）

これはセレクションする方法ですが，今日は時間の都合上，これを省きます．このような方法をいろいろ作った結果，我々のクオリティが非常に上がってまいりました．

特に最近，cDNAというのは，今までどれだけ頑張っても5 Kbか4 Kbが取れればいいところだったのですが，最近我々の新たなベクターを作りますと，平均差長が9とか8，一番長いのは十数Kbのクローンが取れます．ジストロフィーという大きな遺伝子が一発で取れます．

何でこんなことができるようになったかというと，新しいベクターです．何が新しいかというと，これは基本的にBACという非常に長大な遺伝子を保持するようなものを，cDNAに使いますと，長い遺伝子でもできるということで，こういうものが出来上がっています．この辺は飛ばします．（図 6）

最終的に，我々の完全長cDNAの長さですが，全体の長さでここに9 Kb以上というのがありますが，5’エンドの完全長率は非常によい成績を得ています．少なくともここでの完全長の定義はCap siteまでということではなくて，タンパクの開始codonを含んでいるパーセンテージを上げました．こういうふうに非常に高率に，タンパクの全長をコートするようなcDNAが得られてきたわけです．（図 7）

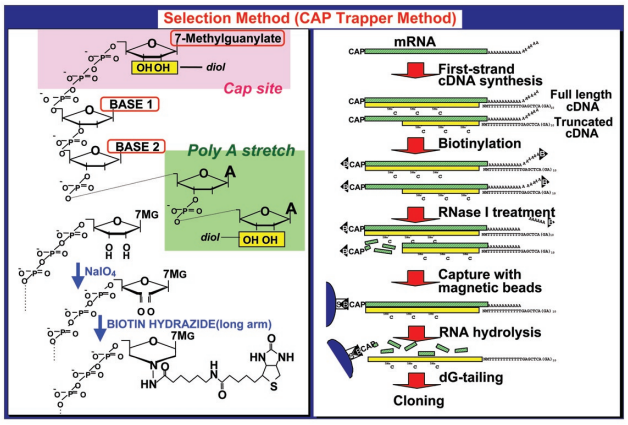
我々のcDNAのサマリーですが，普通に作ると大体2～2.5 Kbぐらいの平均差長があります．長いところを核として，非常に長いところだけ集めてきますと3.5 Kbぐらいです．一番長いベクターを使いますと5～16 Kbぐらいのものが取れてきます．完全長率は，普通，ライブラリを作りますと90～95％が完全長のcDNAで，少なくともタンパクの開始codonであるA・T・Gを含みます．（図 8）

それから，今から申し上げますが，どんどんデータベース，バンクを作っていきますと，もしパーシャルcDNA，不完全長でも，新しい遺伝子なら，バンクの中に入れますから少し妥協するということで，70％ぐらいの効果があります．

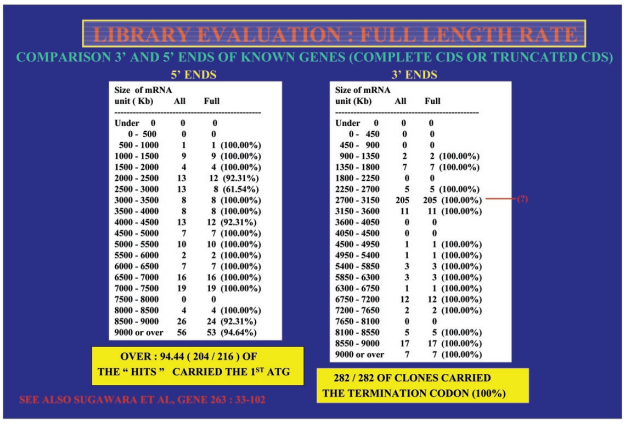
さて，実はそれを実際2～3年でやろうと思ったら，1日4万個のシーケンスを処理しなければいけない．これはクレージーだと言われたのですが，ともかく技術がなかったら作らなければならないということで，必死になって装置類を作ってきました．

これはプラスミド・プレパレーターです．普通，プラスミドを調整するのは手でやります．あれは1日100～200個やれば，翌日しんどくてできないのですが，この機械は全自動で4万個取ってくれます．非常に効率のよい機械です．（図 9, 10）

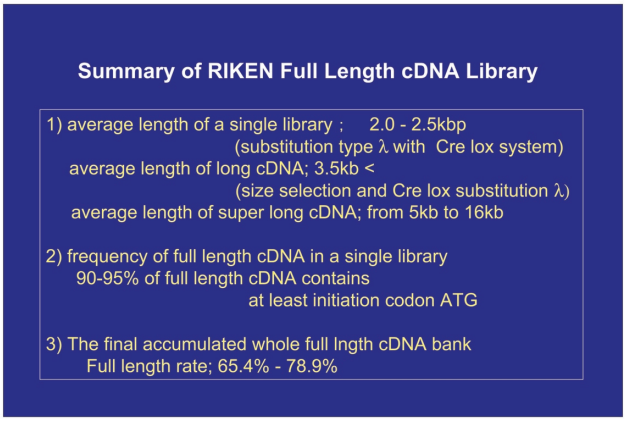
それから，PCRがたくさんできるようなものを作りましたし，最終的にシーケンサーも作りました．このシーケンサーは16×24の384のフォーマットのプレート，穴が384開いています．そこから，こういうガラス細管の中に直接インジェクションして電気泳動し，最終的にシーケンスを決定するような機械も作りました．シーケンサーそのものです．これは新技術事業団の生命活動のプログラム，村松先生が総括されていますが，その資金を得て，こういうものを開発してきたわけです．（図 11）



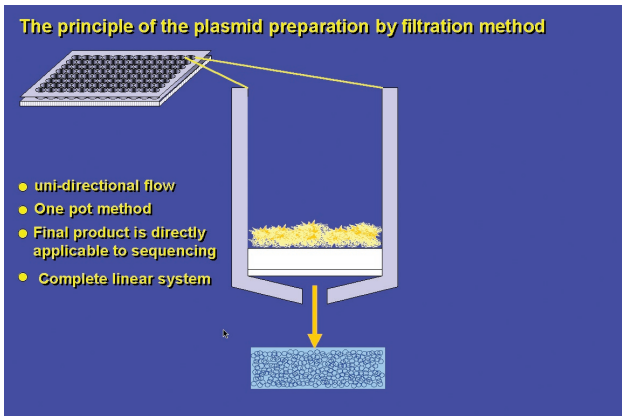
（図 6）



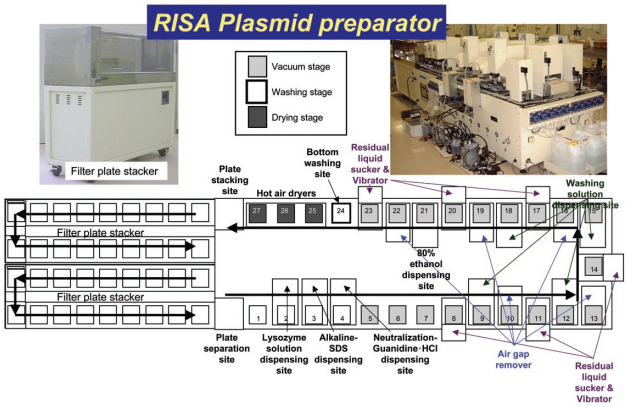
（図 7）



（図 8）



（図 9）



（図 10）

（図 11）

最終的にこういうパターンが出ます．これはおもしろい話ですので，念のために説明します．我々はプラスミドを4万個も取ることをなぜできるかというと，私は元は医者なのですが，こういう機械的なシステム化についてかなり勉強しました．

それは，量産しようと思ったら，液を入れるとき，試薬を入れるのは全部1方向（uni-directional flow）でなくてはいけない．もう1つは，絶対チューブへ移してはいけない．one-pot methodである．最後に得たものが，直接その次の反応に使える．それができると，自動車工場のようにラインの上に乗せていける．こういうものが必要です．

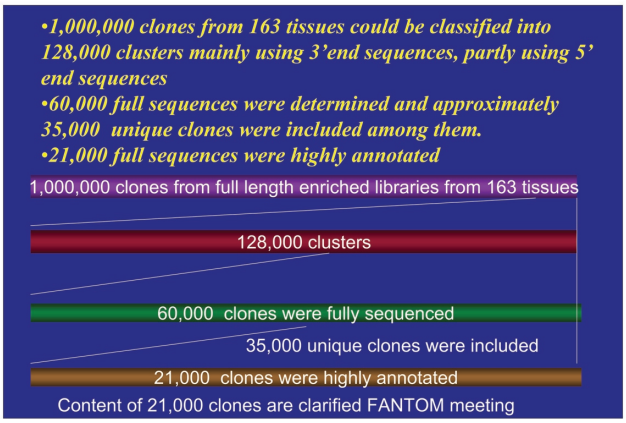
それで，まず大腸菌の培養液をやりますと，ここはガラスのフィルターがあります．membrane filterがありますが，リゾチウムを入れて，アルカリSDSを入れると大腸菌が壊れます．中和・吸着溶液を入れるとこのようになり，吸着するとプラスミドがガラス板にくっつきます．あとは洗浄液を入れて洗います．そのあと溶出溶液を入れて溶出しますと，最終的にプラスミド溶液が得られます．これはシーケンスそのものに使えます．

この図は，1方向から入れて1方向，uni-directional flowです．こういうチューブを使ってやるのですが，ライン化を進めるうえで非常に量産に向いているということです．参考までに付け加えました．

そういうことをしてラインができてきましたので，これで実際にライブラリを作り始めようと．ここにはマウスのいろいろな組織が書いてあります．163個の異なるステージ，異なる組織から取ったものの表です．こういうシーケンサーを使って，まず両端から読んで，どんどんシーケンスをしていきます．

これは今のシーケンスの結果です．163 tissueから100万クローンを取りました．これはノーマライゼーションという手法を使っていますので，これは普通に取ったライブラリでは，大体1億クローンぐらいから取った分に該当します．ものすごい量です．それを，両端のシーケンスを決めて，同じものは除きますので，12万8000ぐらいのものに落ち着きます．（図 12）

ところが，これぐらいのグループに分類しても，やはりまだredundancyがあります．そこで，しょうがないから全長を決めていきます．現在6万個のクローンの全長シーケンスが決定しておりますが，これはまだredundancyが入っています．同じものが2個以上含まれているものがあります．3万5千個，ユニークなシーケンスが取れてきました．



（図 12）

そのうち，この6万個に3万5千種類の遺伝子が入っております．しかも2万1000クローンに関しては，注釈づけまで行いました．今，世の中ではヒトゲノムプロジェクトが出ましたが，あれから見ますと，こういう高等生物の遺伝子の数は約3～4万といわれています．この中に，かなりの部分は含まれているといえます．

私どもはそういうものを注釈づけて，目でシーケンスを見ても何かわかりませんので，注釈づけしていきます．注釈づけをしていくために，例えば全く今まで知られていた遺伝子は知られていた遺伝子とする．その知られている遺伝子によく似た遺伝子，マウスの遺伝子によく似た遺伝子，これはsimilar-IIとか，定義づけをしてどんどんやっていきます．（図 13）

そしてヒトには知られているけれども，マウスに知られていないものとか，モチーフをまた見てみる．例えば転写調節因子が含まれるzinc finger domainがあるものはこれだけとか，そういうモチーフを見る．それから，タンパクはコードしているけれども，絶対に何かがわからないというようなものも入っています．

こういうものをどんどんコンピュータで分類して，あとは目で確認していきます．そのためには専門家を呼ばなければいけないというので，去年の夏に60人ぐらい，世界各国からこういうものに興味がある研究者を呼んで，ミーティングをしました．そしてこういうデータベースを作っていったわけです．これは何に似ていますとか，これはこういうものです，などと書いてあります．

これを分類します．あるものはcell division，分裂に関係あるものとか，energy metabolism，エネルギー代謝に関係あるとか，いろいろ書いてあります．

びっくりすることに，機能のわからない新しい遺伝子が山のようにあります．これはいったい何をやっているのだろうかというのが，今後の課題になってくるのです．

それをもう少し細かく分類したものがこれです．

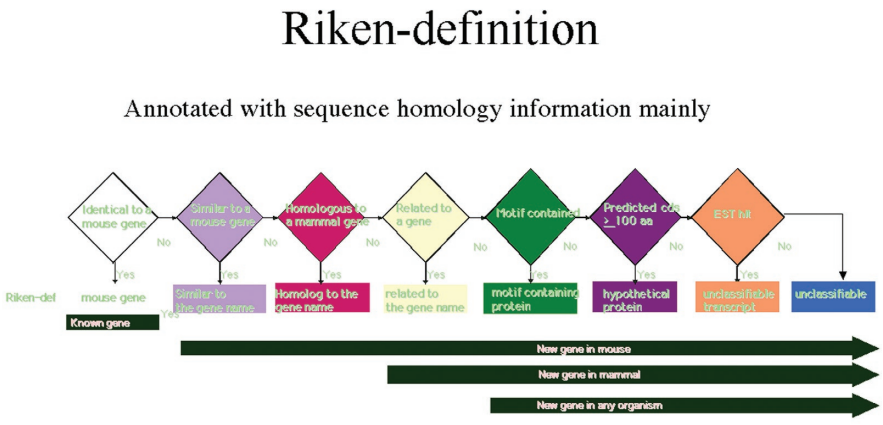
現在2万1千，これは去年終了しました．現在，‘typhoon set’とありますが，FANTOMというのはそのミーティング名の略です．Functional ANnotation of Mouse cDNA．annotationというのは注釈づけです．シーケンスに，一個一個目で見て（curation），これは何かという名前を付けていく．名前を付ける操作をするものを，我々のこういう国際的なコンソーシアム，FANTOM Consortiumと呼ぶようになったのですが，FANTOM1というのがその名前で，最初の2万1000個を終わらせました．

それが‘typhoon set’となり，これは今年の10月で4万5千個になります．たぶん‘cherry-blossom set’が来年4～7月ぐらいまでの間に12万8千，全部，シーケンスを決定していきたいと思っています．要するにこの全部の中に5万個のユニークな遺伝子がたぶん入っているだろう．そのうちの3万～3万5千はタンパクをコードしているユニークなジーンであろうと思われます．

いったんこういうシーケンスが決まりますと，ゲノムの配列がわかっていますので，コンピュータで比較しますと，先程も村松先生のお話に出てきましたが，hybridization in silicoで，コンピュータの中で考えて，ゲノムのどこにcDNAがあるかというのを見ていく操作をします．

現在，6万個のうちの3万8千個ぐらいがゲノムの上に張り付いているのですが，ときには1個のcDNAが，いろいろなところに張り付きます．これはよく似た相同的なファミリーがあちこちにあるということです．ときには1個の遺伝子から複数のcDNAが出てきます．これはalternative splicingといいます．exon（構造配列）のつながり方が違います（図 14）．

現在，我々のバンクはどの程度を含んでいるかというと，アメリカのNIHの中にあるNCBIというバイ



（図 13）

オインフォマティクス・センターがありますが，そこで1個ずつ遺伝子を目で見て，これは絶対遺伝子だというのを，セットを作っています．そのセットのことをレフセック（Ref-Seq:Reference Sequence）と呼んでいます．そのレフセックで今現在，絶対これは遺伝子だぞというのが，7千個報告されているのです．その7千個のうち大体6400個を，我々がランダムに取ったものがカバーしています．それ以外に1万8千個あるのですが，この個数から見ると，我々のバンクは全遺伝子の9割をカバーしているのではないかと考えます．

その中から，いろいろなタンパクのシーケンスがあります．いろいろなモチーフがあります．新しいモチーフが，実は見つかってきたりしています．例えば新しいタイプのロイシンジッパーとか，今日は全部挙げませんが，いろいろなものが見つかっています．

こういうことを実行しますと，世の中には2つのゲノム・プロジェクトがあります．DNAからRNAになってタンパクになります．このDNAのシーケンスを決める．これがゲノム・プロジェクトです．RNAの配列，クローンを集めて配列を決めるのがcDNAプロジェクトです．（図 15）

ゲノム・プロジェクトの方は21世紀になった今年の2月15日にヒューマンゲノムのドラフトが出ました．cDNA・プロジェクトの方は，今度我々がこれより1週間前，2月8日号に，我々の最初の2万1千個のセットを報告しました．

おもしろいことに，この2つは関係があります．なぜ関係があるかというと，実はこれは我々が強調するのですが，その中に，マウスのcDNAというのは，なぜマウスをやったか．ヒトの病気，医学に貢献がない場合は，あまりおもしろくない．ところがマウスは非常に重要です．ヒトでできない医学上の研究のほぼすべてを，網羅できます．

ヒトゲノムのゲノム配列だけでは，どこが遺伝子かわからない．そこで我々が取った遺伝子そのものの配列を見て，ヒト遺伝子のゲノムの中から，どこが遺伝子であったのかというのを想定する．ヒトゲノムの表のこの部分のくだりは私たちが書いた部分ですが，こういうところが非常に使えるということで，cDNAは非常に重要です．

さて，とにかく今，一生懸命集めているということですが，次は，集めてもその遺伝子が，いつどこでタンパクになっているか，mRNAになっているかを知るべきであると思います．そのためにマイクロアレイを使います．このマイクロアレイは，たぶん次の油谷先生が言ってくれるので，今日，僕は原理のスライドを書いてきていません．これは見てもらったらわかりますが，スライドガラスの上に，細かいDNAの点が打ってあります．ここの中に22,000個の点が打ってあります．（図 16）

この22,000個の点を打つために，こういう機械を作りました．この機械の先の直径が100 μ ですが，その中に異なるDNAを打っていくのです．1つのアームの先端がこういうふうに48ピンあり，1ストローク0.5秒で打つような機械ですので，1秒間に192点打ちます．

こういうものを20 K・2万種類の遺伝子がいつどこで発現しているかを，どんどん見ていきました．今，40 Kに伸ばしているところですが，brain , liver , kidney , lung , 脾臓，心臓，こういうもので，いつどこでどの遺伝子が発現しているかを見るデータベースを，どんどん作っていきます．現在のところ2万遺伝子が50個の種類のtissueで発現しているというデータベースを作りました．これを4万遺伝子に増やしつつあります．（図 17）

さて，タンパクが，次にタンパクとどう相互作用しますかというインタラクションをするデータベースも必要である．タンパク－タンパク・インタラクションはどういうことか．

これはあるcDNA , 10万個あるのか, 3万個以上は絶対ありますが，数万個あるのでしょう．そういうものはどれとどれが相互作用するかを，全部スクリーニングしていく．このようなタンパク・インタラクションのスクリーニングは，いろいろなところで役に立ちます．

基本的に，こういう完全長cDNAの一番重要なポイントは何かというと，タンパクを作ることができるわけです．それで，こういうスクリーニングのできるシステムを作りました．mammalian cell hybrid とありますが，これはスクリーニングシステムの原理です．

今日は詳細を省きますが, 100万組を1日にスクリーニングできるようなシステムです．（図 18）

こういうものを作って，どのタンパクとどのタンパクが相互作用するかをずっと調べて，それをこういうかたちのデータベースにしてあります．これはどういうことかというと，この黄色の四角は1つずつの遺伝子で，タンパクです．このタンパクは，例えば太い矢印で，このタンパクは強くこのタンパクに相互作用するということが，ずっと示されています．これがデータベースのかたちで入っています．

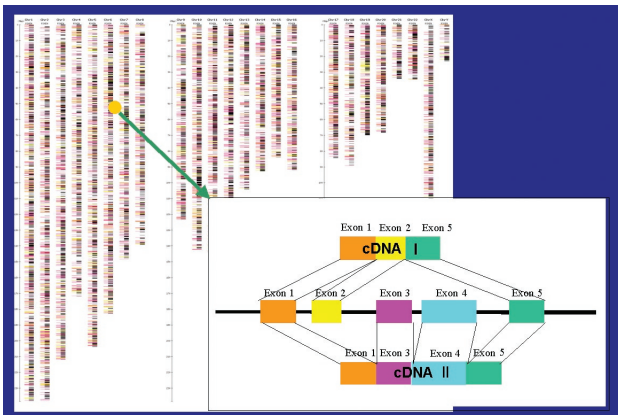
さて，こういうものを全部作っていくと，いったい何ができるかを少しだけお話ししたいと思います．

まず我々，医学領域の研究は，おそらくこれは堀川先生が話されるのではないかと思いますが，糖尿病の原因遺伝子は何かというのを調べます．それを調べるのは遺伝学です．GeneticsとGenomicsを使います．

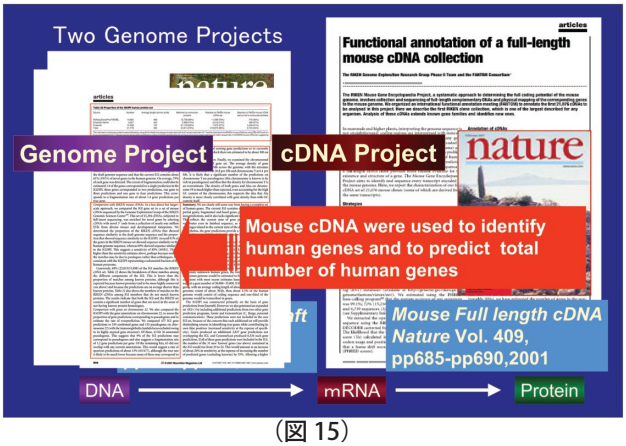
positional candidate cloningとは何かといいますと，親から子に病気が伝わると，その病気と一緒に伝わるゲノムの領域はどこかを探します．そうすると，その位置にその病気の原因遺伝子があるでしょうということで，それをずっと詰めていくやり方です．（図 19）

それで病気の原因遺伝子がつかまります．しかし，それで何がわかったか．例えばこのタンパクは，この病気の原因遺伝子だと一発でわかる場合もありますが，わからないことがいっぱいあります．それを解明するためには，パスウェイを調べなければいけない．この遺伝子のパスウェイはこういうパスウェイで，最終的にこういう病気が生じるというのを，ずっと転写のパスウェイを見ていく．この遺伝子がその次の遺伝子を活性化したり，抑えたりします．それがcascade（滝）のように流れていくわけです．こういうものがどういうルートをたどっているかを，見なければいけないことになります．

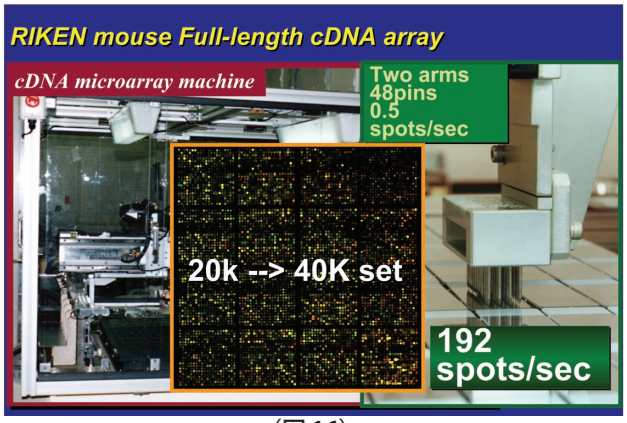
その一項に注目してみますと，これも非常に簡素化された図ですが，あるタンパクとタンパク，例えばこのタンパクと別のタンパクが相互作用して，何かのコ



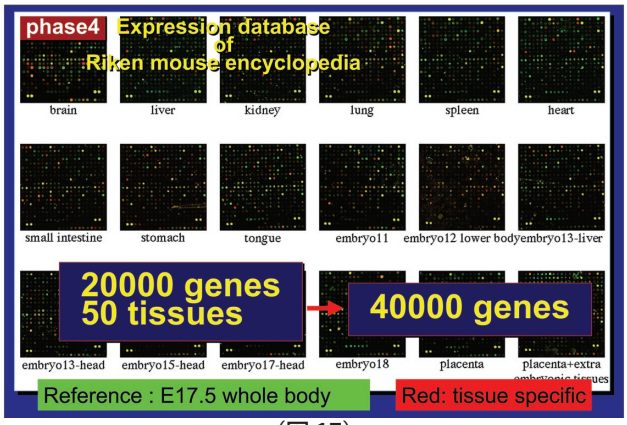
（図 14）



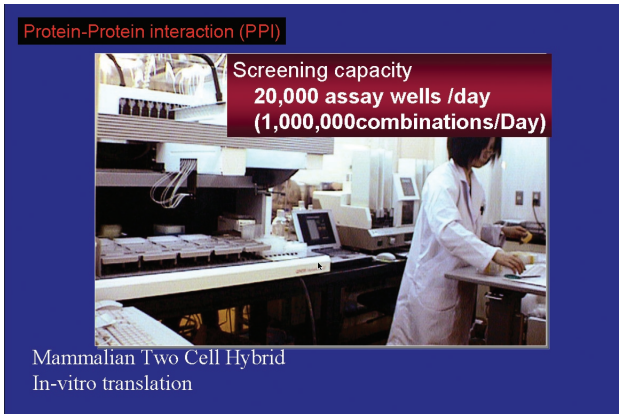
（図 15）



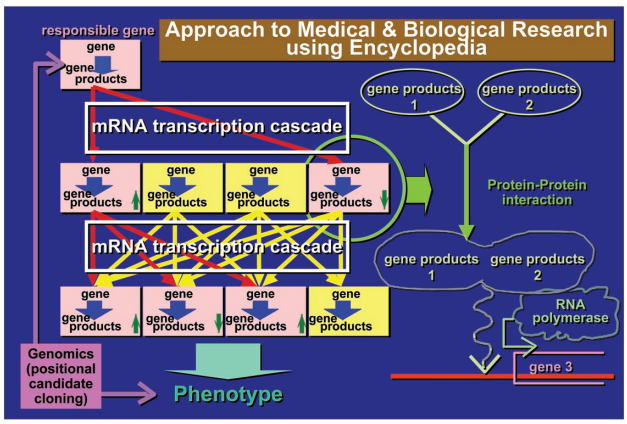
（図 16）



（図 17）



（図 18）



（図 19）

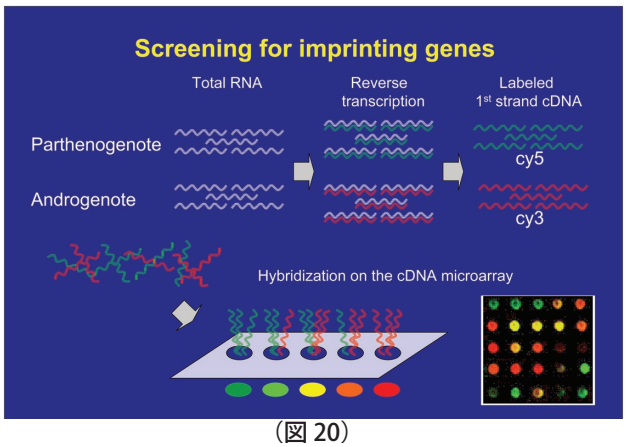
ンプレックスを形成して，その次の遺伝子を活性化する．RNAポリメラーゼを活性化するということで，こういう流れができるわけです．こういう流れを見ていくことが，非常に重要なことです．

遺伝子を同定していくために，例えばどういう同定方法がありますか．ヒトの遺伝子のマップ，病気のマップがあります．病気と一緒に伝わるゲノムの領域はどこかを調べて，病気の遺伝子が，染色体のどこにあるかが記述してあるマップがあります．逆に，完全長cDNAで（我々はcDNAを山のように取っていますから），ゲノムの配列が出てくると，どこにあるかが全部わかります．これだけでも結構，遺伝子がつかまってきますが，これプラスまた別の情報を入れると，もっとよくつかまるのです．

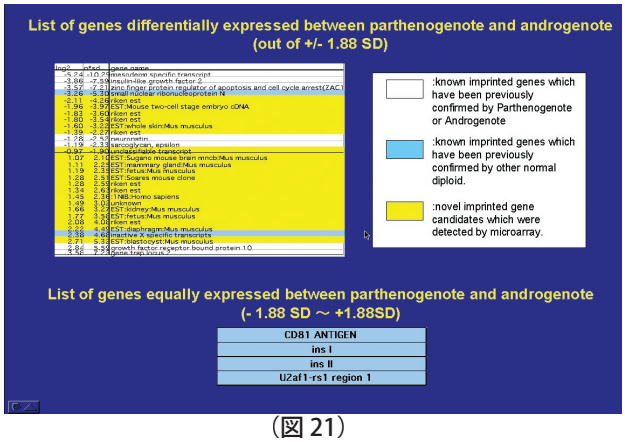
実例を言います．インプリント（imprint）という現象があります．このインプリントは刷り込み現象といいますが，例えば父親と母親の遺伝情報，卵子と精子から来る遺伝情報は均一ではないです．お父さんから来たゲノムの染色体からだけRNAができて，お母さんから来た場合はできないとか，その逆という現象があります．現に，お父さんから病気の原因遺伝子を受け継いだときだけ，発症するような病気がたくさんあります．逆に言いますと，転写がインプリントしている．お父さんからだけRNAがありますよというマップがあれば，これとこれを比べると，候補遺伝子（candidate gene）を挙げることができます．（図 20）

これでうまくいった実例を1つ言います．インプリントの遺伝子をまとめて取ってこようと思ったらどうしたらいいか．単為発生胚（parthenogenote）と雄核発生胚（androgenote）ありますが，パルセノゲノムというのは，単為発生胚です．この単為発生胚というのは，マウスなどで母親の雌性前核だけを媒体にして人工的に作る胚です．それは父親のゲノムが入っておらず，母親のゲノムしか入っていません．逆にアンドロゲノムというのは父親の遺伝子しか入っていません．この2つを用いて，こういう胚からRNAを取ってきます．このRNAを取ってきて，片方を緑の色素，片方を赤の色素でラベルして，標識をしておいて，先程，マイクロアレイ（各種の遺伝子が載っている）に配列しますと，この2つは競合して，母親からだけ発現している遺伝子が緑に，父親から発現している遺伝子は赤に染まるはずです．これで実際にやってみますと，いっぱいインプリント・ジーンが取れてきました．（図 21）

おもしろいことに，我々は実はもうシーケンスを全部やっていますので，ゲノムの中の染色体のどこに何があるか．横の1本のバーがcDNAの位置を示してい



（図 20）



（図 21）

ます．これで3万5千個，ほとんど全部がマップされていますので，この中のどれかという話です．

実はおもしろい1個の遺伝子が見つかりました．これはZAC1という遺伝子ですが，これはzinc finger domain（亜鉛をcurateするかたちの構造）を持った，あるタンパクです．このZACタンパクは，もともと肺癌や卵巣癌を抑える癌抑制遺伝子として取られてきた遺伝子です．現にこのZAC1が，今のやり方でスクリーニングしたらインプリントしていることがわかりました．父親からだけ発現します．（図 22）

おもしろいことに，このZAC1は染色体の69-24にマップされているのですが，ちょうどそこに新生児一過性糖尿病の遺伝子がマップされました．この新生児一過性糖尿病は，父親から遺伝子を引き継いだときだけ病気を発症します．（図 23）

そういうものを実際に見てみますと，例えばこれはtrisomyですが，こういうふうに父親からの領域が2倍になると，病気が発症する．すなわち父親からだけ発現している遺伝子ですから，発現量が増えるという現象のときに，この病気が出てくることになります．

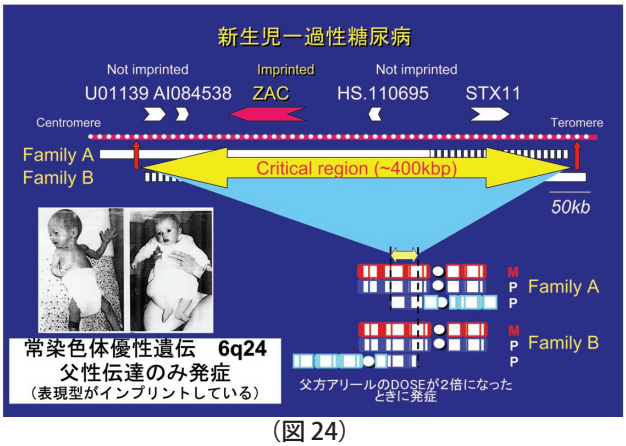
実際，こういうブレイクポイントを，染色体のトランスロケーションの領域がいろいろな患者さんがいま

すので，つぶさに調べると，ここからここまでの領域の中，400 Kbの中に入っていることがわかります．その中に実はZAC1という遺伝子があって，インプリントしていたわけです．（図 24）

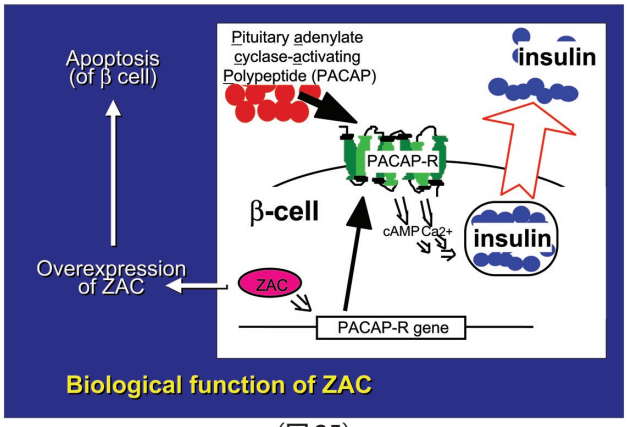
現に，これが病気の原因遺伝子です．なぜなら，Pituitary adenylate cyclase-activating Polyopeptide (PACAP) のレセプターがあります．このレセプターにシグナルが行くと，すい臓のベータ細胞からインシュリンを分泌することを刺激するシグナルになりますが，このレセプターの受容体の遺伝子を調節しているのがZAC1です．また，ZAC1がover expressionしますとapoptosisが生じます．こういうことで，糖尿病が発症しているのだろうということがわかったわけです．

このようなアプローチは，これは全部，こういう表現系として，病気がインプリントしているものです．こういうものを全部使いますと，同じようなやり口で取れていくのではないかと考えます．（図 25）

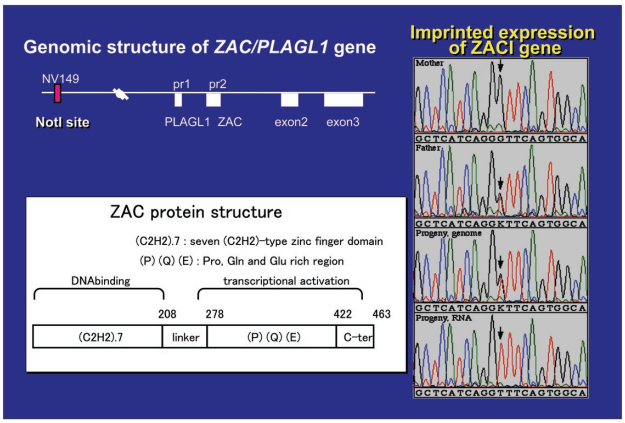
このようなデータを，人間が頭で考えてもしょうがないので，コンピュータに考えさせますので，データベースを作ります．このデータベースのマネジメントシステムを作ります．ジェノマッパーといいます．（図 26-29）



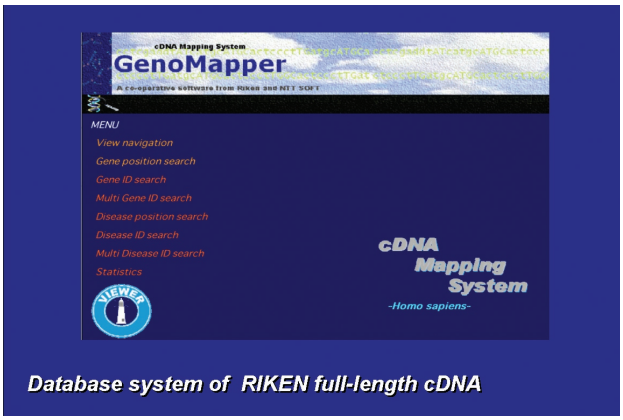
（図 24）



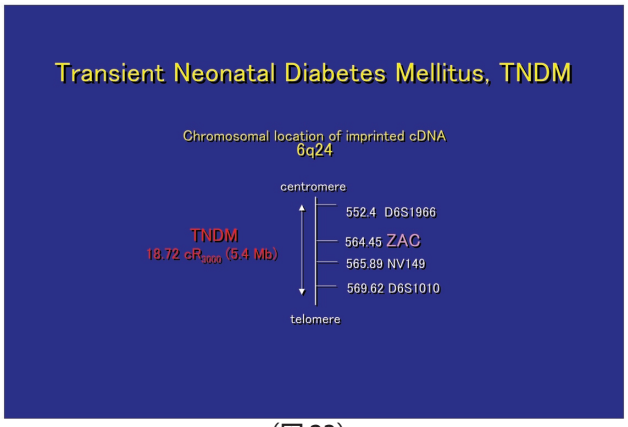
（図 25）



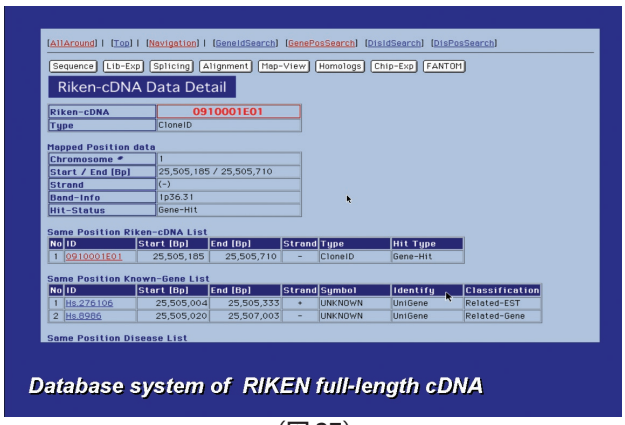
（図 22）



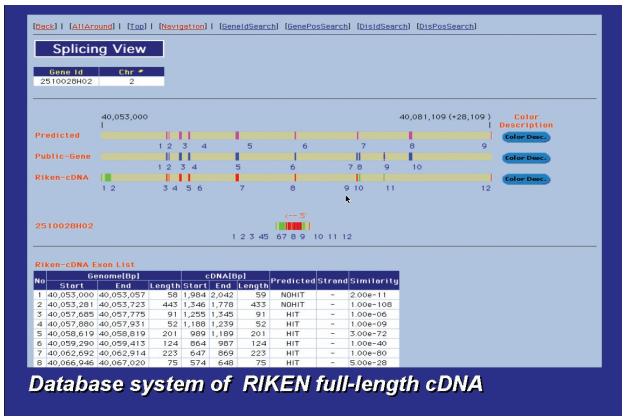
（図 26）



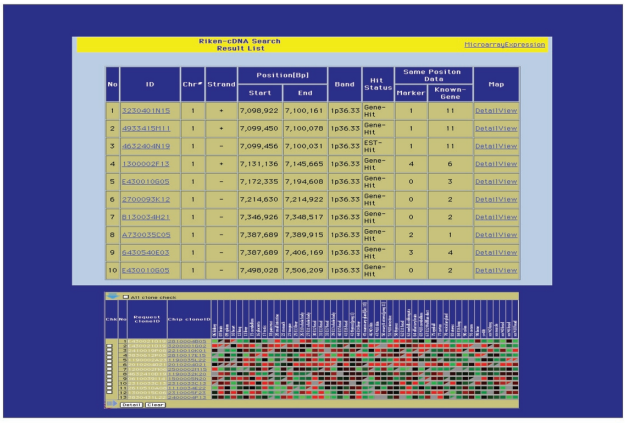
（図 23）



（図 27）



（図 28）



（図 29）

このジェノマッパーでは，クローンのIDをここに，染色体は何番で，染色体の何ベース目から何ベース目にマップされています．ストランドのプラスマイナスでは，マイナスですと．リージョンは染色体19のどこですと．こういうのをざっと入れていくわけです．

そうすると，ゲノムのA・G・C・T，遺伝子の予測プログラムで予測された配列（exon）が，こことここことここと教えてくれるのですが，実はコンピュータはしょせんコンピュータです．実際，RNAを取ったわけではないですから，実際にRNAを取ってできません．ですから，理研のcDNAを使いますと，こんなところに別のexonがあったり，こんなところに余分なexonがあったりすることがわかります．

実際，今のようなpositional candidate cloningをやるためには，染色体の番号とフランキングマーカーを入れますと，どんな遺伝子がそこにあるかを教えてくれますし，そのときの発現情報も教えてくれます．

病気の原因遺伝子をつかまえるためには，3つの重要な情報があります．1つは，染色体異常の位置情報，発現情報，タンパクのインタラクションの情報です．こういうものを使っていきますと，どんどん後方遺伝子を縮めていって，最後に病気の転・変異（SNP），1塩基の置換を見ることによって，その病気の原因遺伝子が同定されます．

2～3ここで実例を挙げますが，skin cancer（皮膚癌）の感受性の遺伝子が，遺伝学的にマップされていたのです．これは90年代にこんなマップが出てきたのですが，ここから全然進展していないのです．

というのは，これは広大な領域で，これはいったいどこの遺伝子かわからないということであったのですが，実際，我々のこういうやり方で，遺伝子がほぼ3万5000個マップされています．ですから，そこの領域にあるもののsourceでいきますと，7000個に絞り込むことができる．皮膚癌の感受性遺伝子は皮膚で発現しているということで，皮膚で発現している情報を入れますと，1400に縮められる．

また，これは少し難しいのですが，タンパクのインタラクションをしているものが，例えば先程言いました，病気の原因遺伝子に両方ともたまたまマップされた場合は，この両方ともcandidate geneである．こんな性質を利用しますと．この3つの場合もそうですが飛ばしますと，おもしろいことがわかります．

これは染色体を1番から順番に回したのですが，この赤のバーのところが感受性の遺伝子の領域です．たまたまここにある遺伝子で，皮膚で発現していて，なおかつタンパクの相互作用するものどうしはどれでしょうというのをコンピュータが教えてくれ，こういうものがcandidate geneとして挙がった．6個に絞り込めたということになります．

結局これはシーケンスすると，皮膚癌に罹りやすさを支配する遺伝子は，実はapoptosisに関係する遺伝子だったのですが，こういうこともどんどんわかってくることになります．

最後のまとめですが，我々はテクノロジーを開発してエンサイクロペディアを作りました．これはフェノタイプと遺伝子を結びつけるのに非常に役に立ちます．エクスプレッション・パターンもそれを絞り込むのに役に立ちますし，この中で分泌タンパクがあると，そのままプロテインのタンパク製剤になります．不完全長cDNAは，そのタンパクの三次元構造から，例えば創薬などに用いられるように，三次元構造はX-ray crystallographyとNMRを使いますが，それのタンパクを調達する材料になります．また，タンパクのインタラクションを，例えば阻害するものを見つけてみたりすると，これはdrug designになります．（図 30）

こういうことをやろうと思ったら，やはり非常に多くの人たちが関与しなければいけません．そこで，これは理研の我々のグループで，これは機械を作って

くれる技術部です．これは動物の細胞を取ってくれる化学合成屋さんです．これは病院関係です．企業群が先程の機械を作ってくれました．マイクロアレイをどんどんするコンソーシアム，それからFANTOM Consortium，これは一個一個目で見てやるannotationです．（図31）

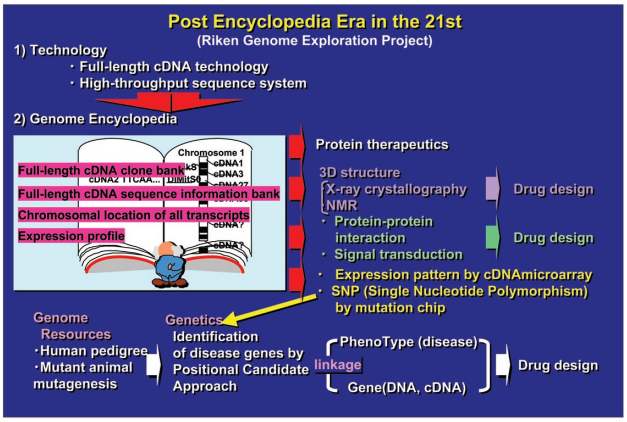
それから，うちの所長の和田先生と，顧問をしていただいている村松先生がアドバイスをしてくださいました．

もう1つ重要なのは，他とよく連携するためには，国立遺伝研のDDBJの五条堀孝先生が，非常に我々のデータをリリースするために，ものすごく手伝ってくれました．それから，世界のみんなが寄り集まって，FANTOMというコンソーシアムを作りました．（図32）

このように，ゲノム・プロジェクトを始めるときには，「ゲノムはアメリカに取られたか」と思いましたが，とにかく「完全長cDNAだけは日本のお家芸にしてやるぞ」と思ってここまで一応もってきたのですが，現在のところ，世界で一番大きなtranscriptomeです．

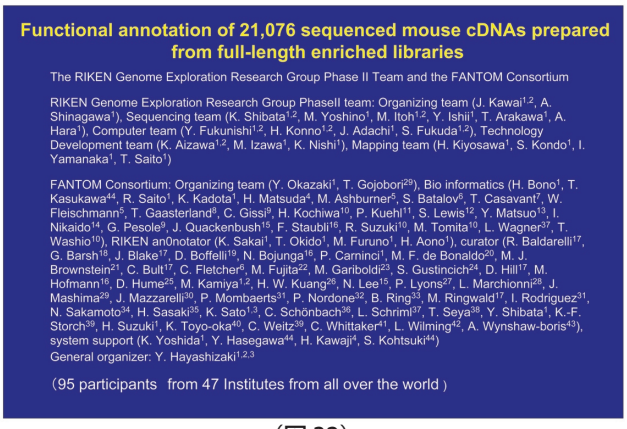
我々が作っただけではだめで，皆さんに使ってもらわないといけないので，みんなにディストリビューションするように今できています．注文してくだされればディストリビューションする会社もありますので，これをお使いになることはできます．

どうもご清聴ありがとうございました．



（図30）

（図31）



（図32）

座長：大変おもしろいお話をいただきましてありがとうございました．せっかくの機会ですので，スペシフィックな質問がありましたら1～2題受けたいと思いますが．よろしいですか．

先生，4万セットのうち，現時点では2万セットですね．仮にcommercially availableとして，一般研究者が手に入れるとして，どれくらいで手に入るのですか．

林﨑：会社に聞いてみないとわからないのです．私は会社のエージェンシーでないのであれなのですが，企業向けと学術向けで完全に分けてあると思います．2万個でいくらぐらいだったでしょうか．100万円と言ったと思いますが．

座長：1セット100万円ですね．

林﨑：1セット，全部のセットです．あれは2万数千あったと思います．

座長：ございませんでしょうか．どうぞ．

参加者：どれくらいの数のマウス特有の遺伝子がありますか？

林﨑：それは非常におもしろい質問ですが，結構あります．結構というのはどの程度かという質問なのですが，実はマウスの中で，ヒトにマップできないcDNAで，なおかつオープン・リーディング・フレームがはっきりとある，というのがあります．個数としては何十ではないですか．もう1桁か2桁上です．

座長：ほかにございませんでしょうか．

参加者：細胞内局在を研究している人たちはいるのですか．

林﨑：それは細胞内局在でやっているグループがあります．我々もこれを使って，*in situ* hybridizationをバーッとやってみたり，そういうのは1つはアメリカ

にMGCというプログラムがあって，特に脳でBMAPというプロジェクトをやっています．脳で*in situ* hybridizationをずっとやって，細胞内の局在を見たり，オーストラリアにも，やっているグループがあります．

参加者：そちらでは，局在の研究を進める可能性はないのですか．

林﨑：うちの研究室の中では，細胞内の局在というより，私たちはどういう順番でやるかというと，全部やるのはものすごく膨大なので，例えばある特定の病気，ある特定の何かというときに，少なくとも同じ細胞で発現しているものどうしを比べる，という戦略を取っています．特に病気にフォーカスしたものであれば，それどうしで先にやる．そうするとパネルのあるもののうちで，虫食い状に埋まっていく，そんなかたちです．

座長：それでは時間も押しておりますので，予定どおりここで10分，ブレイクをいただき，休憩させていただきます．

林﨑先生，どうもありがとうございました．

シンポジウム

多因子疾患における疾患感受性遺伝子のポジショナルクローニング -NIDDM1を例にして

堀川　幸男
（群馬大学生体調節研究所助教授）

座長　粟田卓也（埼玉医科大学第4内科学）

第4内科の粟田です．最初に，今回ご講演いただく堀川先生のご紹介をさせていただきます．堀川先生は1985年に東京大学の工学部応用物理学科を卒業されたというユニークな経歴があり，そのあと1985年（同年）に大阪大学医学部に学士入学されています．そして1989年卒業後，一般病院にて内科臨床研修後，1993年に大阪大学医学部第2内科に入局されました．第2内科は代々，代謝や糖尿病をされているところです．そのあと1996年にシカゴ大学生化学分子生物学部門に留学されました．そこは有名なDr.Graeme Bellがおられる糖尿関係の遺伝子では世界一であるというところで，そこで今回発表される2型糖尿病の感受性遺伝子同定をされました．2000年に帰国され，群馬大学生体調節研究所調節機構部門　遺伝情報分野に入られ，2001年に助教授に就任されました．

糖尿病や高血圧などの生活習慣病は世界でも非常に増えており，その遺伝子解析は非常に注目されていますが，先程の村松先生のお話にもありましたように，多因子疾患であるということで解析が困難であり，なかなか成功しません．堀川先生は今回，ありふれた2型糖尿病の遺伝子解析を行い，疾患感受性遺伝子のマッピングに初めて成功され，『ネイチャー・ジェネティックス』に発表されました．

そういうことで，今回「多因子疾患における疾患感受性遺伝子のポジショナルクローニング-NIDDM1を例にして」ということでご講演いただきます．

本日はゲノム医学研究センターの立ち上げという非常に栄えある場で，こういう光栄な機会を与えていただきまして，まことにありがとうございます．

今日，私がお話ししますのは，今，非常に増えてきている生活習慣病，その中でも横綱級の2型糖尿病の疾患感受性遺伝子の同定についてです．我々が世界で初めて，多因子疾患の何らかのゴールにたどり着いたのではないかというデータをご紹介したいと思います．

糖尿病は，日本人よりもアメリカ人の方が圧倒的に多いと思われていると思いますが，実際のパーセンテージでは，すでに10年前に日本がほぼ追いついており，人口比で4~5%程度です．そして同様の上昇を示して1998年には6%となり，日本とアメリカは同じぐらいのレベルです．あんなに太ったアメリカ人と日本人がほぼ同じパーセンテージを持っていることは，非常に驚くべきことで憂慮すべき問題になっています．

2型糖尿病の大まかな病態機構ですが，今日お話しします遺伝的背景，加齢や肥満，あとは運動不足等により，筋肉でのインスリン抵抗性，要するに血糖を下げるもの（インスリン）の作用不足が起こります．また脂肪の蓄積による，脂肪細胞からのさまざまなサイトカイン．これについては最近いろいろ同定されていますが，アディポネクチン，TNF-α等がインスリ

ン作用に影響します．さらにブドウ糖の貯蓄庫である肝臓からブドウ糖がこぼれ出す．これらが一緒になって，どんどん病態が進行していき，インスリンを出している膵β細胞の機能不全も重なり，境界型，そして最終的には顕性の糖尿病になります．これが2型糖尿病の中でも，一番メジャーな割合を占めるありふれた（common）糖尿病の病態の発症スキームです．

先程の病態の裏にある遺伝的背景には，2型糖尿病の中でも2つのパターンがあります．まず，モノジェニック（単一遺伝子），遺伝子1個で証明できる糖尿病系です．これが若年発症の糖尿病（Maturity Onset Diabetes in the Young，MODY）で，この糖尿病はインスリン分泌不全を症状のメインとするために，日本人の糖尿病のモデルとして適当であるため研究されています．このMODYタイプは，ある1つの遺伝子で病態発症が証明できるものです．

もう1つは今日の話の主人公ですが，多遺伝子型の糖尿病です．これは環境因子の下に，さらにさまざまな疾患感受性遺伝子が複雑に絡まり合って起こります．これがありふれた2型糖尿病と考えられ，感受性遺伝子の同定をするのは非常に困難だと考えられます．

この遺伝子効果をバラの花にたとえますと，MODY型は，MODY遺伝子という1つの遺伝子のフェノタイプ（表現型）に対する効果が非常に強く，見ておわかりのとおり，赤と白とはっきりとした表現型の差を出せます．こういう遺伝子は少数の，大規模家系を使って同定されてきました．

ところが今からのゲノム医科学は，環境因子と非常にマイルドな効果の遺伝子（gene）が重なって，ある完成した表現型を作る疾患を対象にしており，個々の遺伝子の効果は小さく下のバラのように表現型の差が出にくいのです．ただし，これからはこのような遺伝子を同定しなければならない時代になっています．

これが最初に示した単一遺伝子型糖尿病の原因遺伝子でMODYタイプの1~6番です（図1）．一番同定の早いものが8~9年ぐらい前で，HNF-4 α，HNF-1 α，HNF-1 βなどとありますが，この中でグルコキナーゼという解糖系の酵素以外は，すべて核内の転写因子が糖尿病の原因になっています．

これからお話しする多因子型の2型糖尿病は従来より，「遺伝学者の悪夢」といわれていました．なぜならば単純な遺伝形式をとらず，複数の原因遺伝子，疾患感受性遺伝子が複雑に絡まり合って起こるからです．さらに，そこに環境因子（environmental factor）の関与も考えられます．さらにありふれた2型糖尿病疾患・対照者では，リサーチをするときに，健常者と糖尿病の素因を持つ者とをはっきり区別することが非常に困難です．その1つ目の理由は血糖による糖尿病診断です．すなわちある基準値を設定していますが，ある値までが糖尿病である値からが健常者かを数値で決めていることです．それがクリアな2つのグループを形成できない1つの理由です．

もう1つの理由は，糖尿病素因を持っていても，発症が中年以降であることです。このことは又，発症時既に両親が亡くなられており，連鎖解析をするうえで家系のサンプルをすべて取ることが難しいということでもあります．実際問題としてそれを解決するには，同じ家系で兄弟2人が糖尿病であったり，3人が糖尿病であったりという兄弟に注目して解析をしていく手法が必要になります．これについては後で詳しく述べます．

原因遺伝子を同定するアプローチですが，まずはあてずっぽうの方法，candidate gene approachです．糖尿病であればインスリンやインスリンの受容体，ブドウ糖を運ぶものなど，生体の中にさまざまな関係するmoleculeがあるわけですが，そういうものをランダムに解析するというものです．そこで何らかのポジティブなデータが出た遺伝子がいくつか発表されています．最近になって，いろいろなグループがコンビネーションして，ラージスケールのスタディが発表されました結果，PRAR γとアミリン，IAPPというタンパク，この2つのみがラージスケールのスタディでも有意であるということが，現在までに報告されています．

あてずっぽうだけではどうしようもないので，その次のこととして何とか連鎖解析を用いて，まず領域だけでもアバウトに決めたい．しかし，先程言いましたが，2型糖尿病のように発症の遅い糖尿病では，大家系を組むことができないため，従来の連鎖解析は行いにくいのです．そこで糖尿病を発症している兄弟だけを使って，その兄弟が共通にシェアしている染色体の領域を決めていこうではないかということで，それが

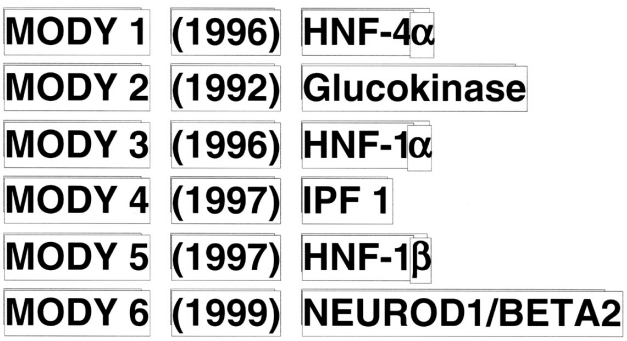


図 1

さまざまな民族で行われているわけです．

簡単な原理ですが，両親2人がdouble heterozygoteの場合，子どものジェノタイプを考えます．専門用語ではIBDと呼ぶのですが，この2人がシェアしているアレル（allele）の数を0個，1個，2個と表して解析していくわけです．0，1，2のパーセンテージは理論上，何の偏りもなければIBDが0になるのが25％，1になるのが50％，2になるのが25％，すなわち1対2対1になります．ところがある染色体の領域のマーカーが2型糖尿病に関係しているとすると，IBD＝0が1，2の方に偏っていきます．そういう偏りをある統計学の手法で解析していくのが，この罹患同胞対解析です．

実際に，さまざまな民族でこのような罹患同胞対解析（Affected Sib-Pair analysis：ASPアナリシス）が現在まで行われています．国家的なプロジェクトとして，アメリカのFUSION，GENNIDという非常にラージスケールのスタディがあり，他にそれぞれの国で各グループが200~300の同胞対を集めて行っています（図2）．

本日お話ししますのは，この中でメキシコ系アメリカ人2型糖尿病候補遺伝子座として知られている，NIDDM1についてです．これは世界最初なので「1」ということですが，NIDDM1は染色体（chromosome）2番の終末部に同定されていました．こういうゲノムワイドスタディーでは非常に多くのfalse positive（偽陽性）が出ますから，かなりきつめにロッドスコアの基準を採っています．従来の解析方法では3ですが，実際にこういうゲノムワイドになると，3.6や4といったものが有意であるとしております．実際にこの2番の終末部，このNIDDM1の領域でのロッドスコアは4で，そういう厳しいクライテリアでも有意でした．

ここで用いられたサンプルはメキシコ系アメリカ人のもので，合衆国のテキサス州の半島の先の方のエリアから採取しました．

Starr countyのメキシコ系アメリカ人の簡単なプロファイルですが，すべてのcounty（テキサス州）の中で，最も高い糖尿病の関連死亡率を持っており，ここのエリアでは45歳以上の20~25％が糖尿病なのです．

統計学は再現（replication）できることが重要ですから，独立した2種類のサンプルを使っています．第1サンプルが170家系で330同胞対，第2サンプルは82家系で118同胞対．あとはランダムコントロールとして112名．これを用いて解析を行っています．

全ゲノムでの解析ですが，非常に弱い有意差を入れると大体11~12個の候補ローカスがありましたが，その中で今回お話ししますNIDDM1は一番高いロッドスコアが出たものです．これがNIDDM1の領域です．1 LOD confidence intervalというのは，ロッドスコアが1下がるときの幅です．この12センチモルガンの間に，80％以上の確率でNIDDM1原因遺伝子があるだろうということです．この民族で2番目に高いピークが染色体15番にありますが，それを考慮する（タイピング上で分類するときに両方を同時に考える）と，4のロッドスコアが約7（正確には7.28）に上がり，この「1 LOD confidence interval」という領域を12から7センチモルガンに縮小できます．

実際に，縮めたあとをどのように詰めていくかです．非常にラフではありますがヒューマンのゲノムシーケンスが終わっている現在では必要にはならないのですが，その当時はまだこの領域のシークエンスはありませんでした．ですから，我々はまず人工染色体を使って，物理的地図（physical map）を先の7センチモルガンに作成しました．そして従来の考え方より，発現している遺伝子（expressed sequence tag：EST）または既存の遺伝子が原因と考えやすいため，これを最初のターゲットとして解析をスタートしました．

そしてコントロール群112名，患者群は110名全群と，NIDDM1領域に非常に強い連鎖を認める37名，15番と2番の両方に連鎖を強く示す20名という3群に分けました．この3群を比較して，もし原因遺伝子アレルがあるならば，これらのグループにつれて頻度が高く，さらに連鎖へのcontributionが大きくなるだろうという仮説の下に研究を進めていきました．

これはその7センチモルガンのフィジカル・マップです．PAC BACとか，いわゆる人工染色体をつないでみました．そうすると，この染色体2番の終末部は非常に相同組換えが多いせいか，7センチモルガンでも実際の見積もりより距離は小さく，実際にはこのあたりでは7センチモルガンが1.7メガベースでした．普通は1センチモルガンは1メガベースですが，この

2型糖尿病と連鎖が認められている遺伝子座

	location	*marker*	results
Pima Indian	11q23-25	*D11S4464*	LOD=1.7
	1q21-23	*D1S1677*	LOD=2.48
Mexican American	2q37	*NIDDM1*	LOD=4.02
	15q	*CYP19*	LOD=1.27
Utah Caucasian	1q21-23	*APOA2*	LOD=4.3
Finnish	12q	*NIDDM2*	LOD=3.65
French	1q21-24	*APOA2-D1S484*	MLS=3.04
	3q27-qter	*D3S1580*	MLS=4.67
FUSION			
Finnish	20p	*D20S849*	MLS=2.15
	20q	*D20S107*	MLS=2.04
		D20S886	MLS=1.99
	11q	*D11S901*	MLS=1.75
	2p	*D2S319*	MLS=0.87
GENNID			
Caucasian	5q	*D5S1404*	LOD=2.80
	12q	*D12S853*	LOD=2.81
	Xq	*GATA172D05*	LOD=2.99
Mexican American	3p	*D3S2432*	LOD=3.91
African American	10p	*D10S1412*	LOD=2.39

図2

あたりでは240キロベースでした．

我々がESTや既存の遺伝子に見つけていった遺伝子多型（スニップス：SNPs）です．非常に単純な，ときにはGとA，CとTといった1塩基の差，ときにはinsertion-deletionという非常に小さい塩基のずれを多型といいますが，このSNPsを同定して領域を縮めていこうという方針で進めました．

先程の1.7センチモルガンですが，このようにSNPsをとって，さまざまな組み合わせでハプロタイプを構築し，その頻度差をコンピュータで解析していきます．そうしますと1番，2番，19番それぞれ単発では，患者全群，NIDDM1領域に連鎖を示す群，2番と15番の両方に示す群というサブグループ間で，ほとんど頻度は変わりません．ところが，それらを組んだところ1番，2番，19番で初めて，有意差を得ることができました．したがって，我々のフォーカスを1，2，19番で囲まれる領域に絞りました．この段階で約220 キロベースあります．

次に，コアな部分の66 Kでの，メキシコ系アメリカ人で見つかった全SNPsです．実際にはいろいろな民族を読んでいますから，これ以上SNPsがあります．青がメジャーアレル，高い頻度の方のアレルのホモ，そして赤がヘテロ，黄色がホモですが，カルパイン10の領域では若干多型が少なく，less polymorphicで連鎖不平衡が強いことが見て取れます．

先程のものから，明らかに100％連鎖不平衡，ですからインフォメーションとしてはSNPsが2つあっても1つにしかならないものは削って，それ以外の約100個を，先程のサンプル数110名の患者と，112名のコントロールで比較して解析しました．このコアな66 Kの中には3つの遺伝子がありました（図3）．

主役のカルパイン10という遺伝子とGプロテインの35番，RNPEP like（アミノペプチダーゼBに似た遺伝子）の3つです．そして一個一個のSNPsに関して，オリジナルの連鎖への寄与と，あとは関連解析で単一SNPでの頻度差，ハプロタイプでの頻度差を考えながら進めました．最初は，単発のSNPsで，43番というカルパイン10のイントロン3にあるSNPsが非常に有意な候補として浮かび上がってきました．

全部のSNPsのデータですが，例えば43番について示しますと，このように頻度差はグループごとに上がっていきます．そして，43番でメジャーアレルをホモで持っている人を取り出すだけで，ロッドスコアが9になりました．統計学的なpermutation studyの結果，この43番のSNPデータをもとに分類したときのロッドスコアは，有意に高いロッドスコアであることがわかりました．したがって，我々はこの43番を最初の非常に強いcandidate SNPとして，このあとの解析を進めていきました．

これは43番を0点にとって，66 Kの領域で連鎖不平衡を計算した絵です．連鎖不平衡というのは，指標としてDプライムとRのスクエアがよく使われますが，その2種類の43番を起点にしたデータです．そうすると，この領域では43番から約15 Kぐらいの両側にわたって，連鎖不平衡が保たれていることがわかります．

次に2点比較方法ですが，SNPsすべてについて，

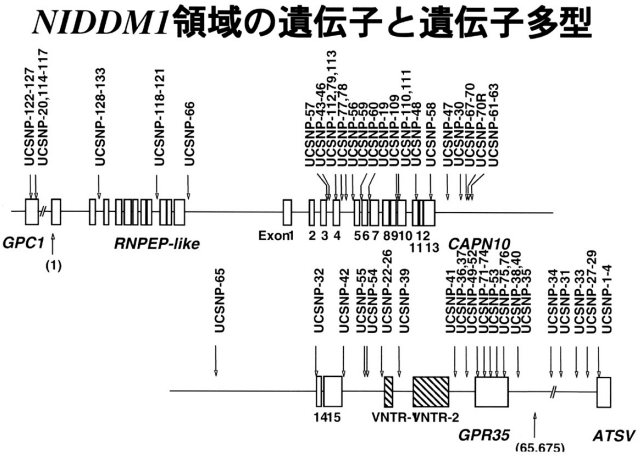


図3

2個ずつ連鎖不平衡の状態を，先程のDプライムやRスクエアを計算して表した図です．先程のコアな領域を大体網羅しているのですが，このあたりがカルパイン10の領域になります．このあたりは非常に強い連鎖不平衡があることがわかります．

43番，19番，63番というSNPsをピックアップして，実際にハプロタイプを構築してみます．1番がメジャーアレル，2番がマイナーアレルですが，1–1–2というコンビネーションを持つハプロタイプと，1–2–1というコンビネーションを持つハプロタイプがそれぞれ，この疾患群につれて頻度が上がります．そして我々はレプリケーションを確認するために，メキシコ系アメリカ人だけではなく，フィンランドのサンプルやドイツ人のサンプルを同時に検討しました．

先程はハプロタイプの頻度のみを解析したのですが，どうしても最終的にはジェノタイプ，ダイプロタイプ（diplotype）を検討しなければいけません．そこで実際に患者でダイプロタイプを組んでみたところ，1–1–2と1–2–1という組み合わせをしたときに，危険度（odds ratio）が有意になることがわかりました（図4）．そして，これは他民族でも高い傾向を示しています．これが有意にならないのは，サンプルのNが少

ないためですが，このコンビネーションで一番高い危険度を表すことがわかりました．

ここからは機能的な裏付けが必要になりますので，我々は43番SNPに対して機能的な裏付けを行いました．先程の1–1–2と1–2–1の両方に共通して入っており，非常に高い連鎖への寄与を現した43番SNPに，何らかの特殊な意義づけはできないものだろうかということでやったのが，EMSA（electrophoresis mobility shift assay），ゲルシフトのデータです．ヒューマンのHepG2，肝細胞株と実際のヒトのpancreatic isletから採取したタンパク抽出物を用いて，43番のSNPsのあたりにプローブを作り，実際に行いました．

そうしたところ，Aアレル側に強く付き，Gアレルには弱く付く何らかのプロテインがある事が判明しました．即ちこの領域には何らかのプロテインがバインドして，カルパインのエクスプレッション等に何か影響を与えるのではないかと考えました．実際にそれを*in vitro*の系でtransfectionして行ったところ，この43番につくタンパクによって，レポータージーンの発現の差を我々は確認することができました．

即ち細胞レベルでは何らかのタンパクがこの43番にインタラクトして，カルパイン10の発現を変えるのではないかということが示唆されたわけです．次にヒトの個体レベルでの検討ですが，ピマ・インディアンという非常に糖尿病頻度の高いグループのコントロールのサンプルを使って解析をおこないました．今度は先程の43番のジェノタイプによって3つに分け，ヒトの個体のレベルでカルパイン10の発現が違うかどうかを調べました．実際にGG,GA,AAとこのように分けたものでカルパイン10の発現レベルが違うことがわかりました．

さらに，炭水化物の消費量を測定すると，カルパイン10の発現レベルが低い人はカーボハイドレートの消費が低い方にいくという結果が出て，カルパイン10の発現が低い人には，何らかのインスリン抵抗性，インスリン作用の不足があるのではないかというデータを得ることができました．

今度は，実際にマウスやラットの脂肪組織や，筋肉，骨格筋といったブドウ糖を消費するtarget tissueを使って行った実験です．カルパイン10の発現を抑える方法として，今はまだカルパイン10は特異的なインヒビターは持ち合わせていないので，ここで実際に使ったのは非特異的カルパイン・インヒビターです．カルパインはスーパーファミリーを形成していますから，必ずしもカルパイン10だけに対するインヒビターではありませんが，ブドウ糖の取り込みが落ちました．インスリンを作用させてブドウ糖を取りこまそうとした時に作用不全が認められました．すなわち，カルパインの発現を落としたときに，ブドウ糖の取り込みが落ちるのではないかということがこの実験で示唆されました．

次は75グラムOGTT（75 g経口ブドウ糖負荷試験）という，大量のブドウ糖負荷に対して，生体がどのように反応するかという，基本的な糖尿病検査を行ったときのデータです．43番SNPのみではなく，疾患感受性のハプロタイプの組み合わせによって，どのように影響されるかを見たテーブルです．ここでも1–1–2，1–2–1というat riskなハプロタイプのコンビネーション，このダイプロタイプを持つものが，HOMA-Rという，臨床の場でいわれるインスリンの抵抗性，作用不足のマーカーにおいて有意に高いデータをを示しています．

さらにここでわかったことは，EIR（Early Insulin Response），すなわち初期のインスリン分泌相が低い．インスリンの分泌の立ち上がりは悪いが，最終的にはHOMA-Rに示されるように，インスリン作用不足・抵抗性を示すために，2時間のインスリン値と血糖値は上昇するという，相対的なインスリン不足があるというデータを得ることができました．したがってカルパイン10は，インスリンの作用と分泌の，両方に影響があるのではないかということが示唆されました．

2型糖尿病の最初のスライドですが，この病態機構においてカルパイン10はインスリン分泌と作用不全両方に関係があることが示唆されています．実際には今後ノックアウトマウスやトランスジェニックマウスを使って，具体的なカルパイン10のスペシフィックな効果を考えていかなければいけないのですが，現在は以上のようなことが示唆されているということです．

このカルパインですが，カルシウム依存性の中性領域のプロテアーゼ，つまり中性領域で働くカルシ

カルパイン10遺伝子変異と糖尿病発症危険度
(UCSNP-43,-19 and -63)

	Mexican American		Finnish	German
	Set 1 (170 families)	Set 2 (69 families)		
112/112	0.29 (0.08-0.99)	0.92 (0.29-2.95)	1.99 (0.18-22.13)	-
112/121	2.80 (1.23-6.34)	3.58 (1.43-8.92)	2.55 (0.79-8.29)	4.97 (0.64-38.41)
121/121	0.67 (0.31-1.43)	1.07 (0.45-2.59)	1.92 (0.92-4.00)	0.65 (0.33-1.28)

図 4

ウムに依存するプロテアーゼの略です．昔からプロテアーゼはただ細胞の中のゴミ掃除のために必要と考えられていましたが，カルパインはタンパク質の基質のmodifierとして重要な働きをすることが分かっています．実際の臨床の場でも，例えば最近では白内障（cataract）や筋ジストロフィーなどいくつかの病気で，カルパインが原因となる病態がいわれています．基質の種類にも細胞骨格のタンパクや膜タンパク，転写因子がありますし，いろいろな基質を使って生体に影響を与えるものだということがわかってきています．

このカルパインには，典型的（typical）なカルパインと，非定型（atypical）なカルパインがあります．4番目にCalmodulin-like Domain（カルモジュリン様領域）を持っているクラシックなカルパインと，そうでないものを持っている非定型なカルパインと2種類あるわけですが，このカルパイン10は非定型のカルパインです．実際にはこのドメインTというのは，カルパイン10においてはドメイン3のデュプリケーションという構造を持っています．

もともとこのCalmodulin-like Domainが，カルシウムによる活性化に大事だといわれていましたが，最近になってこのドメイン3も，カルシウムの依存性に重要な役割を持つことがわかってきています．したがって，それをタンデムに持っているというカルパイン10は，EFハンド（すなわちカルシウムにrelateするハンド）は存在しないのですが，強くカルシウムと関係するのではないかということが想像されます．

実際にもうゲノムシーケンスが全部出ており，既存のカルパインのプロテアーゼドメインでBlastをかけると，ヒューマンにおいては全ゲノム上に1~14番までしかないのではないかという結論が出ています．遺伝子の数としては14ですが，それぞれにスプライシングなどを持つものもあり，実際にはさまざまなフォームがあります．例えばこのカルパイン10は，私がクローニング中のものだけでも，アイソフォームが8つありました．いろいろなスプライシングで，多機能な表現型を示すのではないかといわれています．他のカルパインのノックアウトマウスや細胞株の解析が進んできて，カルパインの正体がだんだん明らかになってきています．カルパイン10は現在進行中です．

染色体2番の解析に使った15番のところにはカルパイン3がありますが，このカルパイン3は筋ジストロフィーの原因遺伝子でもありますし，骨格筋にスペシフィックに発現しているということで，糖尿病での末梢での抵抗性との関係も想像できます．

現在までのところメキシコ系アメリカ人では遺伝学的には10個ぐらいの候補ローカスがあり，それが複雑に絡み合って，相互作用（interaction）をしながら，ジェネティックな背景を作る．そこにノン・ジェネティックなファクターがあり，2型糖尿病という，ありふれた糖尿病が発症するのではないかというスキームを考えています．

これからになりますが，このNIDDM1は要するに最初に連鎖解析，罹患同胞対解析で，あるエリアに絞り込んで，そのあと多くのSNPsを同定することによってゴールしました．しかし，時代が変わってきており，もはやヒューマンのゲノムシーケンスを我々は手にしています．既にセレラ社ではSNPを280万~300万個同定しているという時代が来ています．ですから，2つの群（患者群とコントロール群）を集めてしまったら，SNPsを全部両群でタイピングしてしまおうという流れが，今の主流になりつつあります．

もともと糖尿病の原因のSNPsは倹約遺伝子といわれ，生存に有利なものです．それが飽食文化になったために，糖尿病になっているということが考えやすいわけですから，皆さんの中にもこの感受性遺伝子は，高い頻度で残っています．したがって，非常に長い歴史の中で，何十~何百代にもわたって組み換えをしていますから，緻密なマーカーがないと短い連鎖不平衡はとてもdetectできません．したがって，このSNPsを優先的に全部タイピングしてしまおうと．現在，実際にはデータとしてはありませんが，私も日本の中でそういうグループでやっており，こういう同定法が主流になってきています．

これは連鎖解析とランダムなアソシエーション・スタディの相関図です．従来の連鎖解析（罹患同胞対解析）では，相対危険度（GRR）が2で必要なサンプル数は3000を超え，非現実的です．感受性遺伝子の頻度が横軸ですが，4ぐらいあれば何とか1000サンプル以内で，同定できるでしょうが，2になると非現実的になります．

それに対してアソシエーションスタディー（関連解析）というのは，1.5と非常に低い相対危険度を持っているものまで，多様の頻度にわたって何とかカバーできます．約1000サンプルぐらいあれば，何とか同定できるのではないか．ですから，このアソシエーション・スタディが感受性遺伝子の同定には非常に有用であるという理論の裏付けのもと，全ゲノムでのランダムなタイピングを今，施行しています．

実際にするといっても，やみくもにする前にいろいろと考えておかなければいけません．例えばSNPsを5万，800サンプルずつを2グループ与えられたときに，10のマイナス6乗の有意水準を決めて，検出率がやっと94％をとれるというところに1ステップでもって

いく．

即ち，この有意水準で5万個のSNPsをタイピングしても，偽陽性（type on error）が0.25あります．逆に言えばこれぐらいに有意水準を設定しておかないと，これぐらいでは収まりません．これが0.05であれば2500個のにせものが出てきますから，このぐらいにしているわけです．これでは大体8000万タイピング，1個が100円として80億円いるわけです．

それを2ステップにすると，ファーストステップで250人，セカンドで550人と分けて，これを両群に分けて，最初はありきたりな0.05という弱いクライテリアでいくと，90％の感度で2500個のにせものが出てきます．それだけを550サンプルで今度は2回目をやって，10の4乗程度の有意基準で80％をとって，先のやり方と同じ検出率をとるやり方ですと，2700万のタイピング，すなわち27億になり，約53億円の倹約ができます．

このように，今後はタイピングをするときにある程度理論づけを持って進めていきます．しかし，最終的にはこれは運任せの部分もあり，実際にはこの中に本物が入っているかどうかなど，さまざまな問題があります．また連鎖不平衡が非常に強いところにSNPsがいくらか出たときに，どれが本物の感受性SNPであるか．そういうことが非常に難しいわけで，具体的には難しい問題をいろいろと含んでいます．

多因子の同定は非常に難しいのですが，アメリカのCBSの開祖者であるエドワード・マローが言った「歴史は，難しいからできないということは決して認めない」という言葉を肝に銘じてやっていくしかありません．多因子疾患の感受性遺伝子を同定することは非常に困難であるということを身にしみて経験していますが，ゲノム研究センターの立ち上げにあまり悲観的なことは言いたくありません．こういう苦労がありますが，何とか疾患感受性ハプロタイプだけでも同定できれば，それで患者の疾患罹患率を予測し，疾患発症を予防することができるわけです．ただ，最終的にはある1つのSNPにたどり着ければいいですし，それがタンパク質の変異であれば話は非常に簡単ですが，現実にはそうではなさそうです．それが現在の状態です．

このNIDDM1のプロジェクトは私がシカゴに留学しているときのものですが，シカゴのグループの面々と，ピマ・グループやイギリス，デンマーク，ドイツなどいろいろな方々のサンプルや協力のおかげで，何とかゴールできたのではないかと思います．

今日の話はこれで終わります．

座長：ありがとうございました．多因子疾患の遺伝子解析について，その現状と展望を含めてお話をいただきました．非常に大変で，今は力ずくでやるしかないのではないかということや，労力とお金が非常にかかるというお話もされました．何か質問はありませんか．

村松：NIDDM1は他の民族ではどうなんですか．

堀川：この2番のエリアの話になりますと，2型糖尿病の原因遺伝子座についてはガーデンバラエティといって，世界中でいろいろなロケーションが出ていますが，クロモソーム20番やクロモソーム1番のある領域では他民族を超えてオーバーラップしています．しかしこの2番に関してオーバーラップしているのはメキシコ系アメリカ人とフランス人のサンプルです．全くgeneticには違うと考えられるわけですが，合併しています．

日本人でも解析していますが，単純にただタイピングして危険度（odds ratio）を組むだけでは有意には出ません．なぜかというと，ランダムコントロールにその組み合わせが多くあれば危険率は有意には出ません．ただ，日本人は今後，脂肪食のintake上昇などで，コントロールの中の何人がこちらに動くかということが，まだ全然わからない状態なので，それだけで単純な評価はできません．実際には，クランプテストなどの細かい検査をしますと，1–1–2，1–2–1は，日本人の中にいても，そういうインスリン作用の不足を表します．

参加者：カルパイン10のノックアウトはまだできていないのですか．先生が前にいらしたシカゴでやっているのですか．

堀川：当然，シカゴで作りはじめています．けれども，chimeraにはなりますが，germ-lineに乗らないのです．それは意味があってそうなのか，今，解析をしています．

参加者：あるいは例えばカルパイン10に結合するタンパクとか，サブストレートなどに関しての情報は．

堀川：それはとろうと思ってやっているわけで，次に油谷先生から話があると思いますが，私が考えていますのはポリジェニックですから，基質もいろいろなジーンが絡むでしょう．それを1ユニットで同時にとらえていくためには，マイクロアレイが最適であるといろいろなものが同時にどう動くかがわかりますから，そういうことを考えてやっています．

参加者：イントロン領域の3にあるSNP43につくタンパクがあって，ミューテーションによってつくかつかないかで，発現を見られて，そこはエンハンサーのところのSNPだったのでしょうか．

堀川：そういうことです．要するに発現を見て，それは今日はスライドには持ってきていませんが．

参加者：発現を確認されたのですか.
堀川：差異としては非常に弱いのです．ただ，それは当然，*vitro*上はレポータージーンの発現です．要するにルシフェラーゼの前にあの領域をつないで見たのは，若干ですが有意差は得られました．実際にそれを*vivo*に持ち込んでやったのが，先程のフィギュアです.
参加者：もう一度まとめると，イントロンの中にある領域は，エンハンサーであったと．その領域も確定されているのですか．そのミューテーションの何ベースぐらいのところまでつくか.
堀川：正確には，そこの塩基をふってとかはやっていませんが，使ったのはコアな24ベースです.
参加者：その引っ付くタンパクは同定されたのですか.
堀川：実際にはそのタンパクを決めたいという段階です．既存のコンセンサス・シーケンスはないです.
参加者：カルパイン10の組織はどこで出ているのですか.
堀川：カルパイン10はユビキタスです.
参加者：ということは，ユビキタスでしたら，患者群と正常群とで採れますね．例えば白血球でもいいわけでしょう．それで*in vivo*の発現量は実際に差があるのですか.
堀川：末梢血でということですか．末梢血は調べていません．ただ，調べてお示ししましたのはmuscle biopsyのデータです.
参加者：確かに非常にイントロンについては議論のあるところで，『ネイチャー・ジェネティックス』でもなかなかアクセプトされなかったと伺っています．臨床的な関連で，ブドウ糖負荷試験で差があったというのは，日本人でされたのですか.
堀川：今日，お示ししていますのはイギリス人のデータです．まだプレスはしていませんが，日本人でも差は出ています．同じようなconsistentなデータです.
参加者：あとはエイズの治療に使うプロテアーゼ・インヒビターが，糖尿病みたいなものを起こすということが出て.
堀川：それは自分たちに興味深いのですが，エイズに効果のあるプロテアーゼ阻害剤はシステインプロテアーゼ阻害剤ではありません.
座長：堀川先生，興味深いお話をありがとうございました.

シンポジウム

癌の機能ゲノミクス解析

油谷　浩幸
（東京大学先端科学技術研究センター）

座長　久武幸司（埼玉医科大学分子生物学）

分子生物学の久武です．次の演題は油谷浩幸先生で，「癌の機能ゲノミクス解析」という演題でお話をしてもらいます．油谷先生の経歴を簡単にご紹介しますと，先生は1980年に東京大学医学部を卒業されています．そのあと研修をされ，第3内科に入られました．第3内科から1988年にアメリカのMITの癌研究センターに留学されています．1994年に日本に帰られ，そのあと東京大学にある先端科学技術研究センターに最初は助教授として，今年からは教授として仕事をされています．今の研究テーマとして，癌のシステム生物学を研究されています．

DNAチップを用いて遺伝子発現を調べる方法が，最近はよくやられています．我々のように昔から分子生物学をやっている人は，遺伝子の発現というとノーザンブロット法をやっていましたが，最近はDNAチップを使って数百，数千，さらには万の単位で遺伝子の発現を調べることが行われています．今日の先生のお話は，そういう技術を使って癌の性質を調べるということです．最新の非常におもしろいお話が聞けるのではないかと思います．

今日はゲノム医学研究センターの開設にあたり，こういう会でお話をさせていただく機会をいただきましてありがとうございます．村松先生には研修のころからいろいろと研究のご相談などに乗っていただき，なつかしい思いがします．

すでに林崎先生がマイクロアレイのお話をされていますし，私どもはシステムは違いますが，マイクロアレイとはどういうものかも含めて若干ご紹介申し上げたいと思います．

先程の村松先生のお話にもありましたように，生命のセントラルドグマというものがあり，DNAのインフォーメーションがRNAに伝わって，RNAからタンパク質になるという中で，最近の学問はとにかく網羅的に見ようと．つまり，患者を診るときも，漢方ではありませんが，全体を見ることが非常に大事になってきます．これまでのように遺伝子一つ一つの形を見る，機能を見るということでなくて，その総体を見ることが大事になってきます．その学問はオーミックス，いわゆるゲノム，トランスクリプトーム，プロテオームという言葉がだんだんこの科学の場で氾濫してきます．

オーミックスという場合に，ゲノムがDNAの総体であれば，そのRNAの転写されたプロファイルの全体像がトランスクリプトームであり，プロテインの総体がプロテオームです．プロテオーム解析については，今後，技術的なブレイクスルーがないと，まだまだ何十万というタンパク質の全貌を得ることは難しいのです．そういう意味で，現在ある数万の遺伝子の全体の少なくとも発現の量を見ることは，十分に可能になってきたということです．

そういうことをすると何ができるかですが，その遺伝子がどこの組織に出ているかで，病気の起きている細胞で，遺伝子の発現のプロファイルを調べることで，病態の理解につながることが期待されます．また，癌細胞などでの変化を見れば，例えばAPC（Adenomatous Polyposis Coli）という大腸癌の遺伝子

は，東大医科研の中村先生などがVogelsteinらと一緒にとられたわけですが，こういうものは癌抑制遺伝子ですから，発現が低下していることが認められます．

あるいはMycというoncogeneであれば，これは8q24にある遺伝子ですが，発現が非常に亢進しているということが見られます．この場合にはDNAのレベルでも普通，増幅していることが多いわけですが，発現量が非常に亢進しています．rasのようにポイントミューテーションでactivation（活性化）が起きるような遺伝子の場合には，発現レベルにはそれほど異常はないけれども，活性化が認められます．癌という病態はこういうさまざまな遺伝子変化の結果として出来上がってくるわけです．

その全体像を見るDNAチップとは何かということですが，定義としては遺伝子のクローン，cDNAのクローンを非常に高精度に高密度に配列化したものです．最近はDNA合成機の精度も上がり，100ベースぐらいのものであればin vitroで作れることになっていますから，そういう合成のDNAを配列することも可能です．塩基配列を調べる，遺伝子の発現量を調べる，または多型を調べる，染色体の数を調べるといったさまざまな応用が可能です．今日は癌の分類という話を中心に申し上げますが，将来的にこれが治療法の選択にもつながります．多型の解析に使われた場合には，薬剤の感受性，副作用の予測に使われるだろうということが期待されますし，臨床検査の分野ではO-157など病原性の大腸菌なのかどうかという検査にも応用が十分に期待されます．

現在，私たちが使っているプラットフォームとしては，ジーンチップのシステムと，マイクロアレイという，スライド上にクローンをスポットしていくタイプのテクノロジーの，両方があります．それぞれ長所と短所があり，ヒトの遺伝子であれば，現在は数万の遺伝子が解析できます．これは既知の遺伝子の数プラスEST（Expressed Sequence Tag）ということで，実際の遺伝子のユニットとしてのローカスの数よりは多くなっています．マウスにおいても，現在のシステムを使えば3万6000ということになります．

これはマイクロアレイの実験の行程についてどこかのウェブサイトから取ってきたものですが，ガラス上にこういうピン，あるいは現在はインクジェットの機械を用いてスポッティングを行っていきます．そして物理的に，普通に実験に使うスライドグラスに，1枚2万というスポットで遺伝子を貼り付けていくわけです．そのときに対照サンプルと，自分が調べたいサンプルを別々の蛍光色素で，例えば赤と緑に標識すれば，それによって，あとはどちらのサンプルで遺伝子が増えているかがわかります．

技術的に機材としてはこういうスポッターを用います．拡大するとこのような小さいピンが並んでおり，このピンで遺伝子を打っていくわけです．この中にガラスのスライドをたくさん並べておけば，一度に何十枚ものスライドが作れ，そのあとはこのようなhybridizationの装置を用います．この機械はうちの研究室にあるもので，12枚が使え，この中でハイブリダイゼーションと洗浄を行います．そのあとは共焦点のレーザースキャナーでスキャンを行います．これはハイブリダイゼーション装置の拡大です．これはスキャンしたイメージで，通常はCy3 , Cy5という励起波長の異なる蛍光色素が使われます．

一方，アフィメトリックス社のシステムであるジーンチップはほとんど既製品ですが，半導体加工技術である光リソグラフィという技術で，光マスク（フォトマスク）の穴のあいているところだけに光が当たります．その光が当たったところだけ化学反応が起きることを用いて，シリコンの基板上に約20~25ミクロン四方の小さい単位で，別々の異なったDNAを合成します．そうすると，現在は40万種類ぐらいのDNAがここで合成でき，大体1枚のアレイで，1万~1万2000個の遺伝子の発現プロファイルが測定できます．

これがスキャンしたイメージで，先程のスポットとは違い，一つ一つの光マスクの形が正方形になっています．このように全く発現していない場合にはほとんどシグナルがないけれども，発現が認められる，ハイブリダイゼーションが行われた場合にはシグナルが得られます．2段になっているのは，似たような配列を下に配置しておくことによって，ミスマッチを用いた配列に対して，それは一応バックグラウンドで差し引けるようなシステムになっています．

遺伝子のなるべく3'側の，ほかの遺伝子とホモロジーのない部分に，このように25ベースのプローブ配列をデザインして，1つの遺伝子を16ぐらいのプローブペアで代表させて測定を行います．

今日はこのようなシステムを用いての腫瘍の診断，あるいは治療法の選択ということで，まずは肝細胞癌の例を説明します．用いた腫瘍は肝細胞癌，小児腫瘍，この対照になる非癌部および正常肝です．これは大腸癌患者の肝転移巣の非癌部です．つまり，こちらは肝炎ウイルスの感染を伴った肝腫瘍の非癌部組織です．それをアフィメトリックス社から現在売られているU95アレイで解析するわけです．これはUnigeneのビルド95というバージョンに合わせて作ったもので，1万2000個のヒトの遺伝子が解析できます．

癌と癌でない部分の区別はそれほど難しいもので

はありません．こちら側の上半分は非癌部で，下側が癌部ですが，このように明らかに発現のプロファイル，一つ一つの列のバーコードのパターンが遺伝子が上がっている下がっているということを示します．そしてこの一つ一つが遺伝子で，ここには1万2000個のうち変化のあった3000個ぐらいの遺伝子について，示してあります．このパターンによって仕分けると，癌の方では発現が増えている遺伝子もありますし，癌になることによって発現が低下する遺伝子もありますから，このように明らかにパターンが違うことがわかります．

その内訳を見ていきますと，こちらの正常，大腸癌肝転移，あるいは肝炎患者の非癌部組織もほぼ分かれていますし，小児腫瘍はここに分かれていますし，この上が肝細胞癌です．こういう単純な二次元の2方向のクラスタリングにおいてもこの程度の分類は可能ですが，さらにこの中に未分化型の肝癌，および中分化型，高分化型という肝癌の細かい分類を試みようとしますと，未分化型の肝癌は小児の肝芽腫の未分化型（embryonal type）と非常に近いということもあり，このクラスタリングのパターンが非常に複雑になってきます．

ここでは，別の多変量解析の1つの手法である主成分分析法を用いて，三次元上に第3成分までを表示して，パターンとして分類できるかを表示してあります．外側のこのあたりに見えますのが小児の腫瘍であり，この赤いところが高分化型，中分化型の肝癌，そしてこちら側に正常群があります．そういう目で見れば，大体グループを形成しているのがわかります．

ここでは正常肝，慢性肝炎と，こちら側が非癌部組織，こちら側が癌部です．約1~2例のはぐれものだけが逆になっていますが，分類は十分に可能です．ただ，この中で慢性肝炎患者と正常肝は，先程のクラスター解析と同様に区別が可能です．

肝細胞癌の場合には日本の人口の約2％がHCV感染陽性ですし，1％強がHBV感染陽性ですが，当然，その両方から肝癌が発症してくるわけです．その違いによる差がプロファイルにあるかを見てみますと，癌部に関してはこの水色のHCVの患者と黄色のHBVの方ではあまり差がありません．一方，こちら側では，ある程度水色の群と黄色の群がまとまっている傾向が認められます．このことについてはこのあとにまた戻ってきまして，この差はいったい何だろうかということについてご説明を申し上げます．

先程の林﨑先生のお話にもありましたが，ゲノム情報，ゲノムの配列はすでにかなり決まってきています．ですから，こういう発現情報とゲノム情報の間に何か関連を見いだせないかということで，私どももゲノム情報上にこういう発現情報を関係づけることを試みています．

癌がどんどん進行していくうえで，染色体が一部欠けていくことが認められます．あるいは，先程のMycという遺伝子のように増幅が起こることがあります．これはDNAの上で確実に変化が起きているわけですが，一方，epigenetics，imprintingという状態で代表される，DNA配列自体には何も変化が起きていないけれども，遺伝子の発現レベルが増えたり減ったりするということが実際に起こっています．癌ではこのように状態が狂っている，暴走してしまっていることがしばしば認められます．ということで，これは従来の例えば染色体が欠けているかという解析では見いだせないわけです．そこで1つの手段として，現在，トランスクリプショナルなプロファイルで何かそういうものが認められないかということを試みています．

従来の古典的なKnudsonのtwo-hit theoryでは，ここに癌抑制遺伝子（tumor suppressor gene）というものがあったとした場合に，そこに1つのミューテーションがあって，その反対側の母由来・父由来という2つの染色体があった場合に，片方が欠失を起こしている．そういうことで両方のコピーが不活化されることによって腫瘍が引き起こされるということですが，必ずしもこのように物理的に染色体が失われるわけではありません．ジェネティックにこの部分が欠けているということですが，こういうものは従来，マイクロサテライトの解析，CGH（Comparative Genomic Hybridization）という染色体コピー数を数えるような方法で，解析が可能でした．一方，エピジェネティック（epigenetic）な変化で，遺伝子のある領域がサイレンシングされている，ある特定の遺伝子が不活化されているという場合には，このLOHはマイナスで，コピー数としては2nが維持されているままですから，LOHとしては見つからないということが起こるわけです．

ここで遺伝子発現のインバランスをマップで見ようということで，1万，2万という遺伝子を解析したあとで，NCBIのGenbankのマップ情報にそれを貼り付けたわけです．

例えば肝癌ですが，染色体が1~22番，および性染色体が縦に並んでいます．先程の林﨑先生の講演の中で，マウスのマップが1番から丸く並んでいた絵を思い出していただければ，似たようなことをここでやっているわけです．そして，ここの部分は肝細胞癌で落ちているような領域，こちらでは肝細胞癌で遺伝子が逆に上がっているという領域がマップしてあります．

そして，例えばここに白い線がありますのでここを

拡大してみますと，短腕（p）で上がっているような部分，あるいは長腕（q）の方で上がっている部分，逆に下がっている部分が認められます．

先程，肝癌の分化度の話を少ししましたが，高分化（well differentiated），中分化（moderately differentiated）の比較的正常に近い肝癌と，あるいはそこからさらに未分化（poorly differentiated）になった状態で比べます．つまり，高分化型から中分化型，中分化型から未分化型へ段階的に癌が悪性化していくわけですが，その間でこういう発現レベルの変化が染色体の領域として変わる部分はないかということです．

最初のこの絵は，この2つのこことここをダイレクトに比べて，全体を見ているわけです．そうしますと，ここの領域は最初の段階ですでに発現亢進を行っていますが，この段階では変わりません．逆にここの部分は徐々に発現低下がaccumulationしていくことが理解できます．

この赤く示しているところに何があるかを見てみますと，このような遺伝子が並んでいまして，この水色で示している部分，この星印がついているようなものが地図の上で確認できたものです．それ以外についても大体，肝細胞が通常発現しているような遺伝子が並んでいるだけで，現在のところは腫瘍にとりわけ関係があるというものは見当たりませんが，いわゆる癌化に伴って分化度が下がるに従って，発現が減っていきます．hepatocyte（肝細胞）の機能を担う遺伝子群がこの領域にクラスターしていた，同じファミリーに所属する遺伝子がこの領域にあったことを反映しているのだと思います．

こういう発現レベルをゲノムワイドに領域的に見ていくということで，もちろん染色体としての異常を検出することも期待できますし，あるいは今お示ししたような遺伝子のクラスターを認めることもできると思います．そして，そのステージによってどの段階で遺伝子発現の変化が起きていくかについても，見い出すことができるのではないかと期待しています．

次に，こういう解析を診断に使えないかということで，肝腫瘍を例にとって，その手法について説明します．いろいろな分化度によって区別ができるであろう，あるいはC型肝炎，B型肝炎によって違うであろうということでしたが，それを識別するための遺伝子，実際にどういう遺伝子が違っているのかをどのようにして拾い出すかということで，ネイバーフッド（neighborhood）解析，あるいはnon-parametricなマン・ホイットニー（Mann-Whitney）解析を説明します．

ネイバーフッド解析というのは，ホワイトヘッド研究所のエリック・ランダーらのグループによって，2年前に出た論文で提唱されたものです．このように2つの群，例えば正常と癌を比べた場合に，変化があるものを拾い出せばいいわけですから，なるべく平均値の差が大きくてばらつきが少ないものが，いいマーカーになるでしょう．その値をP値として定義して，そのP値の大きいものを選んできます．

使った症例はこのようなグループで，年齢以外は特に両群間で差はありません．肝炎の臨床をされている先生方にはご理解いただけると思いますが，B型肝炎とC型肝炎の患者さんというのは，年齢が高くなってから発症しますので，いたしかたがないところです．

またクラスター解析ですが，この非癌部組織のみをここに示しています．それもウイルス感染を伴った非癌部組織だけです．緑色がB型肝炎で，赤がC型肝炎です．先程の多変量解析の主因子分析の結果とほぼ同じで，B型肝炎とC型肝炎の発現パターンはそれぞれクラスターを作っているのがわかります．

この2つの間で，先程のP値の高い遺伝子を選びだそうということになります．実際に計算したP値が，全くランダムにそういうものが起こった場合に，どの程度のP値をとるだろうかということを計算した，計算上の全くランダマイズした場合と，さらに信頼区間を考慮した場合に，自分たちが計測した分布がどの程度有意かどうかということを検定します．それはpermutation testといいますが，このテストでこのように実際の測定した値が計算上のランダムな分布とP値の分布と変わっていなければ，それは有意ではないということになります．

この下のパネルを見ていただきますと，肝癌と非癌部組織のデータです．実際上，B型肝炎とC型肝炎ではランダムな状態よりも外れていますから，C型肝炎特有な遺伝子発現のパターンがあるだろうということかわかります．一方，こちら側の2つは癌組織HCCの中での分布ですが，両群間で差がないことがわかります．

こちら側がC型肝炎の非癌部組織で発現が亢進している遺伝子，こちら側がB型肝炎で発現が亢進している遺伝子です．この赤で示しているのはすべてインターフェロンで発現が亢進する遺伝子群です．つまり，C型肝炎ウイルスはRNAウイルスであり，それに対してB型肝炎ウイルスはDNAウイルスですから，C型肝炎の患者の肝組織ではRNAがウイルスに感染することによって起こる生体のレスポンスが，発現のプロファイルに反映していることがわかります．

一方，癌になった組織においては，そういうウイルスに対する応答は関係ないということで，あまり顕著な差がありませんでした．癌というのは継続的に変化

が蓄積してきますから，高分化型，中分化型，さらに低分化型の肝癌の中でどういう遺伝子の発現の変化が起きているかを見てみました．

先程のpermutation testを使って，高分化と中分化が区別できるかです．これは病理学的にも難しいことがありますが，かろうじて有意差がありそうです．そして，中分化と低分化については非常に差が大きそうでした．

遺伝子を拾い出してみますとこのようになり，この遺伝子のセットは高分化と中分化を比べて差のあるもの，つまり高分化で高くて中分化で低いというものと，高分化で低くて中分化で高いというものです．当然，中分化で高くなっているものは低分化でもそのまま高いですし，中分化で差があるものは，そのまま低分化でも差があることがここでわかります．

こちらで抽出された約25ずつの約50個の遺伝子については，中分化と低分化の間でその差を見ていますが，中分化から低分化へとより分化度が落ち，脱分化することによって，発現が下がるものがきれいに拾われてきています．こちらの25個の遺伝子についても，中分化から低分化になることによって上がっていきます．大体，こちらで下がっているものは，より分化度の高いものでも下がっていることがわかります．

このように見ていきますと，この約100個の遺伝子のパターンを見れば，分子レベルでも肝癌の分化度が大体判定できることがわかります．実際にこういう100個の遺伝子でスコアリングシステムを使えば，分化度についてはほぼ決定することができます．

最後にこういう情報をどのように実用化していくか．最近はtranslational medicineなど盛んに言われますが，この中から新規の腫瘍マーカーが探索できないかということが行われます．

肝癌の場合にはAFP（α-フェトプロテイン）という，すでに臨床で長年使われているマーカーがあり，こういうものがDNAチップでどのように見えてくるかということになります．非癌部と肝細胞癌の組織を比べると，このように非常に突出しており，この線が30倍以上発現レベルが違う遺伝子のラインですから，既存の腫瘍マーカーであるAFPは，このようなRNAの発現レベルでも十分高いので拾われてきます．このように点は小さいですが，いくつか同様な動きを示しているものもあるので，そういうものについて，今，拾い出しを行ってみました．

AFPの問題点は，肝癌の患者だけではなく，その他の腫瘍でも数百程度までは上がることがありますし，肝硬変の状態や劇症肝炎からの回復期でもしばしば上昇が見られるということです．また，肝硬変の患者の場合には，特に小さい癌があるのものではないかということで臨床の先生方の頭を悩ませています．そこでAFP以外のマーカー，現在はPIVKA Ⅱなどが使われることがありますし，AFPの亜分画も使われています．その際に，よりたくさんのマーカーがあれば，より確実な早期診断が可能になります．

これは胃癌ですが，肝癌においてもいくつかの遺伝子の発現レベルが高分化，中分化，低分化で高くなっており，非癌部でも肝硬変で若干上がってくる傾向がありますが，このような遺伝子がいくつも見つかってきました．

先程のAFPをこういうマップで見ますと，高分化でほとんど上がっていません．臨床でご覧になられた場合にも，高分化の小さい肝癌では血清中のAFPレベルは上がってきませんが，この組織のRNAレベルを見てもAFPはあまり上がっていません．実際にこのチップの上でも，10例調べてせいぜい1例か2例しか上がっていませんでした．ですから，こういうものはより鋭敏なマーカーになる可能性があることが期待されます．

今の中の1つのモノクローナル抗体を作成して患者の腫瘍を見たところ，腫瘍でタンパク質がきちんと増えているのが確認されました．これはノーザンブロット，これはジーンチップのスコアですが，約50例の肝組織を検討して，6割ぐらいの患者でこのモノクローナル抗体についてはタンパク量の増加が認められています．

こちらの上の2例が陽性例で，こちらはネガティブですが，免疫組織染色をすると網の目状に見えているかがわかると思います．この網の目状は膜にタンパクがアソシエートしているということですが，実際上はこれは細胞外にあるタンパクですから，これが免疫療法ないしモノクローナル抗体によるミサイル療法に使えないかということも，現在，検討を進めています．

アフィメトリックスのシステムはこういうものです．ある程度信頼性がありますが，まだ臨床診断に使うには高価なシステムですし，自分の好きな遺伝子セットを自由に解析することはまだできませんので，フレキシビティはありません．将来これが臨床検査で使われていくためには，もっと遺伝子の発現が簡単に測れるように，より高感度にしたものが望まれます．また，研究開発をするうえでも，さらに簡易なシステム，自分の好きな遺伝子をすぐにチップにできるシステムが望まれていくわけです．

これからこちらのセンターでも解析を行われるということですが，我々が感じている技術的な課題としては実験精度，再現性，感度で，まだ細胞に1遺伝子1コピーしかないというものが確実に測れるようなシステムではありません．また，臨床検体を測るという場合には，患者の貴重なサンプルでもありますから

微量検体です．しかも，腫瘍であってもその中で腫瘍成分，血球成分，血管の成分などがありますので，検体の均一性が問題になります．つまり，実験の細胞株でないかぎりは，こういう組織の中のある部分の腫瘍とこちらの腫瘍では，また組織系が違うということもありますので，顕微鏡下にマイクロダイセクションすることも必要になってきます．

また，冒頭にマイクロアレイとジーンチップのシステムの両方をご紹介しましたが，その2つのプラットフォームの間でのデータの互換性もなかなか困難です．現在もその間での共通のデータベースが提唱されてはいますが，なかなか実現しないでいるのが世界的な現状です．加えて，技術的な問題としてもどのように標識をしていくか，RNAの増幅を行うか行わないかということや，その標識についても間接標識を行うかというさまざまな問題があります．これはどうにか使えるシステムにはなっていますが，今後，改良の余地がまだまだあるというのが実感です．

最後に機能情報，これはシステム情報学へということですが，今後，癌というものだけをとっても，さまざまなコンピュテーションなシミュレーションが期待されていきます．今日はDNAチップということを中心にお話ししましたが，プロテオームの解析情報，タンパク質の相互作用の情報などがデータベース化されれば，そういうものを理論的に肉付けをして，癌をシステムとして扱うことがいよいよ可能になってくるのではないかと期待しています．

例えばこれはβカテニンのパスウェイですが，先程出てきたAPC遺伝子やβカテニンといったものの発現レベルをパスウェイごとに一望することができれば，この癌で何が起きているかを理解しやすいインターフェースを作ることができます．

正常肝，肝癌をこのように並べてみますと，ここでは発現量を赤から黒へguradualに表示していても，人間の目にはよくわかりません．しかし，差分をとることによって，このように差が出てきて，発現量の低いものが青で，赤いところでは，例えばTCF-4というβカテニンからのシグナルが伝わる，直接，遺伝子のDNAに付く転写因子の発現が増えていることがわかります．

しかし，このパスウェイだけを見ていても癌が理解できるわけではなく，例えばTGFβという抑制シグナルがsmadに入ってきて，このsmadがTCF-4のダウンストリームにある遺伝子の転写の調節を行います．こういうほかからのシグナル，あるいはここのシグナル系から外へのシグナルという，いわゆるシグナルのクロストークを考慮しないと生物は理解できません．こういうインタラクションを，コンピュータの助けを借りながら理解していくことが非常に大事になると思いますし，特にこういう網羅的な情報を獲得して，その情報を整頓していくことがますます重要になってくると思います．

今日はゲノム情報とのカップリングをご説明申し上げましたし，今後は文献情報とのカップリングということで，生物情報のデータベースを作っていく．林﨑先生からご紹介がありましたマウスのエンサイクロペディアが，まさにこういうものの代表になるわけですが，こういうデータは単なるデータです．今はデータの中からどうにか情報を抽出することまではできていますが，これにおぼれることなく，将来的には何らかの生物学的な発見をしたいと考えています．

これがうちのスタッフです．特に情報解析についてはこの広瀬先生の研究室，あるいは同じキャンパスにある岩田先生の研究室との共同で行っています．

座長：ありがとうございました．DNAチップを用いて癌を分類したり，診断するということですが，先生が指摘されたようにまだ問題点がかなりあるようです．しかし，おそらく近い将来には臨床で使われるのではないかと思われます．質問等がありましたらお願いします．

参加者：微量サンプルの問題に少し興味がありますが，例えば臨床的にニードルバイオプシーでとれるようなものを扱って，メッセージを例えばランダムプライマーで増幅した解析結果と，リッチなメッセージをダイレクトに使った解析結果を比較した場合に，フォールスポジティブはどの程度出てきますか．また，どれくらいのサンプルを使えば信頼しうるデータになるのでしょうか．

油谷：RNAのレベルとして私たちがルーチンに使っているのは，トータルRNAで5マイクログラムです．細胞数にするとLCMでマイクロダイセクションで1万個が大体100ナノグラムですから，細胞として1マイクログラムが10の5乗ぐらいで，針の太さにもよると思いますが，ニードルバイオプシーでも十分可能な量にはなると思います．その場合に，臨床診断に使う部分などをいろいろ考えていきますと，なかなか倫理的に難しいところもありますが，もう少し量を減らしていけば現在のテクノロジーでも十分可能です．

ただし，減らしていった場合，特に現在，ランダムプライマーの場合には，アフィメトリックスのシステムは遺伝子の3'側に特にプローブの配列を用いています．cDNAのアレイと違い，ランダムプライマーで増えてきた配列の場合には，それがもともと使ってい

る配列よりももっと上流の方，5’側にバイアスされてしまう可能性があり，きちんとしたシグナルがとれない遺伝子が出てくる可能性はあると思います．それがこの場合には欠点になります．

ですから，アフィメトリックスのシステムの場合には，現在はT7RNAポリメラーゼを使って3’側から増幅をしていますが，それをさらに微量検体から増幅する場合には1回ランダムプライマーで戻します．そういう場合に若干バイアスが出ることがありますし，ランダムプライマーの反応でうまく反対側から戻ってこない場合に，約2割ぐらいの遺伝子は，発現が見られなくなる危険性は伴います．

参加者：癌の組織を採ってくるとき，当然，正常組織の混入もあるから，レーザーキャプチャーのデータはどうされていますか．

油谷：肝癌の場合には，腫瘍の6~7割は腫瘍細胞なので．

参加者：やはり，「いいや」という感じですか．

油谷：ええ．言い訳になりますが，癌組織のプロファイリングということで許していただいて，癌組織を顕微鏡で見る場合には，癌細胞を取り囲む組織のリアクションも診断に入ってきます．実際に例えば内皮細胞で発現が亢進するというものもありますし，腫瘍血管特異的なトランスクリプトもありますので，今回のデータについてはLCMのデータではありません．

参加者：オリゴヌクレオチドのアフィ（アフィメトリックス）のチップのときに，ああいうアフィチップとオリゴヌクレオチドをプリンターで貼り付けたものとを比べたことはありますか．

油谷：あります．

参加者：どうですか．

油谷：オリゴヌクレオチドのどこを選ぶかによって，すごくデータが変わります．傾向として増えているものは，アフィの結果と相関します．アフィも大体1年に1回ぐらいはプローブを見直して，見直すごとにだめなプローブをはじくので，徐々によくはなっていきます．

参加者：少し違って，アフィと同じシーケンスをプリントで．テクニカルなことを聞いているのです．

油谷：アフィは発表すると言っていて，まだ25ベースのシーケンスを発表していないのです．（注．現在は公表している）

参加者：そうですか．遺伝子はわかっているのですね．

油谷：もちろん，遺伝子はわかっています．その25ベースの配列がどの領域か．

参加者：どこかがわからないのですか．

油谷：その数百ベースの中のどこかです，ということまでしかわかりません．ですから，cDNAのここの500ベースなら500ベースから，彼らは25塩基を十何か所選びましたと．その十何か所がここととここということは，まだブラックボックスです．

参加者：それは言わないと，やはり・・・．

油谷：そういうプレッシャーが今は国際的に非常強くなって，去年から出すと言っていますが，いよいよ来年の早いうちに出すようです．そうすれば，またデータの検定が楽になりますし，自分たちでもう一度プローブのデータを計算し直すことが可能になります．

参加者：シーケンスがわからないらしいのですが，そういうデータは世の中にあるのですか．そこを知りたいのです．データというのはアフィのようなやり方でつないでいったチップと，別のところで合成機で作ったものを貼り付けていった同じシーケンスを見た場合に，データの質の差はあるでしょうか．

油谷：in situで合成するというのは，アジレントはインクジェットですが，そこで塩基を配って25塩基を合成していますから，あそこのデータはそれに近いのだと思います．それなりにロゼッタはデータを出していますし，それほど問題はないのではないかと思います．フォトリソグラフィという方法では，1サイクルごとの合成の効率が普通の合成機よりも悪いですから，実際にどれだけDNAがそこに合成されているかというのは，わからないのです．

そこが非常にミステリーで，林﨑先生の質問にはきちんと答えられませんが，ほかのSAGE法やそのあとのRT-PCR，ノーザンで，私たちももともとは比べましたが，発現量の比較的多い遺伝子については，かなり絶対量を反映した相関関係が出ます．実際上のオリゴを貼り付けた場合についてのきちんとしたデータは配列がわかっていないので，アフィメトリックスは知っているかもしれませんが，ほかのところでは特にオープンになっているものはないと思います．

座長：ほかに質問はありませんか．なければ終わりたいと思います．油谷先生，ありがとうございました．

シンポジウム

ゲノム研究の倫理的側面
―埼玉医科大学倫理委員会における審議の実情と今後の問題点

山内　俊雄
（埼玉医科大学精神医学教授・倫理委員会委員長）

座長　齋藤一之（埼玉医科大学法医学）
　山内先生についてはよくご存じのことかと思いますので，詳しいご略歴は省略させていただきます．先生は昭和61年に本学におみえになり，精神医学の主任教授，神経精神科センターの所長というご本務を持ちながら，まさに八面六臂のご活躍です．私も学務委員会でずっとご指導を受けておりますが，本学の倫理委員会についても指導的な役割を担ってこられました．
　一昨年のこの会でテーマになりました性同一性障害の治療に関して，日本で性転換手術を本学で開始する非常に指導的なお立場です．倫理というと，とかく机上の空論という面が多いのですが，実務的な視点を兼ね備えられた切り口で，非常に明晰な論理性を持っておられる先生だと，いつも畏敬の念を持って見上げているところです．
　それでは山内先生，よろしくお願いいたします．

1．はじめに
　本日のテーマであるゲノムの問題と関連して，「ゲノム研究の倫理的側面」についてお話しさせていただきたいと思います．ただいま，大変先端的なマウスの研究，あるいは糖尿病，腫瘍などの領域における分子遺伝学のお話を伺いましたが，こういう研究をするときにどうしても配慮しなければいけない重要な点は，倫理的側面です．その点について話をするにあたって，倫理委員会についての紹介をさせて頂きたいと思います．と申しますのは，倫理委員会の設立の経緯や審議の状況について本学でお話しする機会がこれまでありませんでしたので，この機会に埼玉医科大学の倫理委員会がどんな位置づけであるのか，あるいは審議の状況はどうであるのかといったことをお話ししたいと思います，そのうえで遺伝子解析の研究の審査の状況，審議を通じて明らかになった一般的な問題点，遺伝子解析研究における問題点，あるいは後程申し上げます三省の指針なども含めて，お話を申し上げたいと思います．
2．埼玉医科大学倫理委員会の位置づけと問題点
　埼玉医科大学倫理委員会設立の経緯についてですが，最初は昭和62年に治験審査委員会という薬に関係した委員会として発足しました．これが平成元年に倫理委員会となったのですが，それは全国の大学に倫理委員会が設立される気運と期を一にして設立されたもので，平成2年2月に第1回の委員会が開かれています．本学には，治験審査委員会や組み換えDNA実験安全委員会，腎移植連絡会議といったものがありますが，その高次概念ということで，倫理委員会はその上に属するという位置づけで考えられました．
　ところで，いろいろな審議が行われる中で，倫理委員会のあり方や立場を明確にすることが求められるようになり，平成6年には本学における倫理委員会の役割とほかの委員会との関連があらためて論じられました．本学には脳死判定委員会や心臓移植検討委員会，治験審査委員会，新医療用具審査委員会と

いったいろいろな委員会がありますが，これらの既存の委員会で倫理性の判断に困難が生じた場合には本委員会に諮っていただく，あるいは，こういう既存の委員会で対処できない申請については，倫理委員会が対応する．それでも対応困難な状況が起きたときには，どう対処するかをあらためて議論をするということにしました．

平成10年には動物を対象とした研究の倫理的問題についても論議され，基本的には倫理委員会では扱わないことが再確認されました．つまり，ヒトを対象とした研究の倫理性を検討する委員会であることを確認したのですが，それでは動物の問題についてはどうするかについては，今後の課題とされました．その他に病名の告知や遺伝子解析の結果の告知の仕方についても，検討課題としてとりあげられてきました．このような審議の一方で，学内での倫理的意識の高揚をどのようにするのか，学生の教育も含めた啓発活動の必要性についても検討がおこなわれています．

ところで，ここで問題になりますのは，倫理委員会は，申請されたものを対象として審議するわけですから，申請をしなければ審議のしようがありません．逆の言い方をすれば，審議の申請をするかどうかについては研究者の倫理性が問われているということです．もし申請しないで倫理的責任が生じたとすれば，研究者はその責任を自ら負うべきであることはいうまでもありません．また，逆に倫理委員会が承認した研究であっても，研究の責任は研究を行ったその人にあるのだということも当然である，という姿勢であります．

3．審議内容の概要

こういう姿勢のもとにこれまで，「宗教上の理由により，輸血治療を忌避する患者への対応のガイドライン」「性同一性障害に関する総合的治療，ならびに臨床的研究」の審議をはじめ，2001年11月までに116件の審議をおこなってきました．

そのような折りに，ヒトゲノム・遺伝子解析研究に関する倫理指針が平成13年3月に出ました．本学の倫理委員会では今年の10月までに116件の申請研究の審議をしましたが，中には再審査もありますので，実際の延べ審議数はもっと多くなります．この3年ほどの申請件数は急速な伸びを示していますが，その背景には，皆さんの倫理的な意識の変化が反映されているものと思われます（図1）．

審議した申請研究の内容をみますと，図2に示すように，薬物療法に関するものが多いのですが，これは治験審査委員会で扱わないようなもの，例えばある患者に認可されない特定の薬を使って現在の症状の救済をしようといったものが含まれています．治療法については，薬以外のいろいろな治療法，例えば磁気刺激などが含まれています．あるいは，ある状況において生体の機能がどのように変わるかのという実験的研究（生体機能解明）もかなりあります．今日お話しする遺伝子解析は，数のうえからは4番目です．先程ご紹介いただきましたように，性転換手術の問題も我々が長い間にわたって審議してきたものです．あとは生殖医療，移植，それからエホバの問題も本委員会では長期にわたって検討した問題です．その中で遺伝子解析関連は13件あるのですが，これは本年になって急激に増えており（図1），今後この領域の審議が，倫理委員会における重要な課題になるだろうと考えています．

実際にこの13件がどのようなものかみてみますと，表1に示すように，エストロゲン応答遺伝子の解析，概日リズム障害を伴う疾患の遺伝子解析といった，ある疾患についての原因遺伝子の解析が多く，内科，皮

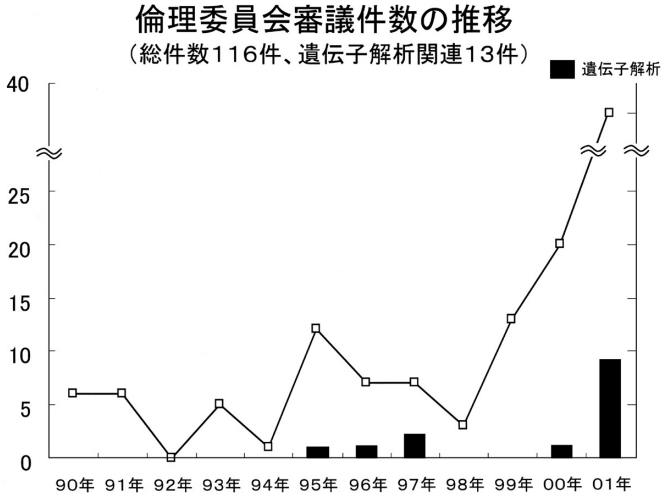


図1

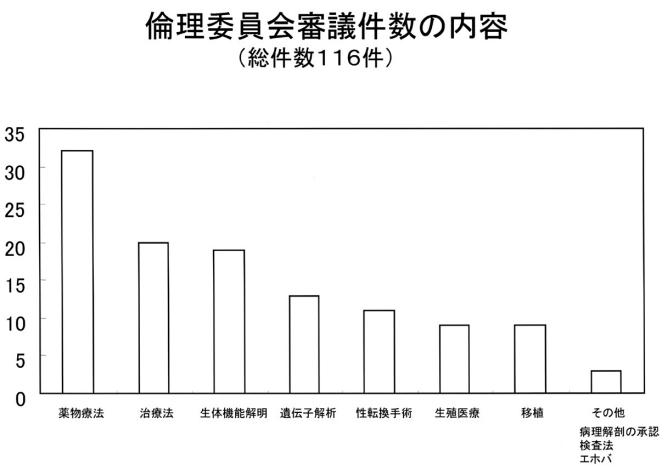


図2

膚科，輸血関係などいろいろな領域にわたる，臨床的な立場からの研究であります．

4．倫理委員会審議における一般的な問題点

倫理委員会がこれまでの審議を通して問題になった一般的な点を表2に示します．まずはじめに，審議は書類に基づく審査ですから明確な研究目的や計画が示されている書類でなければいけないということです．対象選択というのは，例えば健常人をコントロール・スタディとしてとるときに，どのようにして，どういう手段で健常人の選択を行うのかを含めて対象選択の方法が示されているべきでしょう．また，いつ，どのくらいの期間に，どのような方法で研究を行おうとするのかということも明記する必要があります．さらにまた，マテリアル（試料）をどのように採取するかも重要です．たまたま，ある検査をするときに採った試料を別の目的に使うことや，たとえば黙って血液を少し余分にもらうということが許されないのは当然です．予想される危険性，利益，不利益，個人の秘密の保持に対する配慮と説明，得られた結果をどのようなかたちで公表するか，本人に情報をどのようにフィードバックするかということも問題になります．

しかも説明と同意は，文書をもっておこなわれることが原則ですから，これまでに述べたようなことが，説明文書にもきちんと述べられていなければいけません（表3）．

5．遺伝子解析研究で特に問題となる点

ですから，遺伝子解析研究においても，このような一般的な問題点が明確に記載されていなければいけないわけですが，中でも遺伝子解析研究で特に問題となるものを3つほど表2に挙げておきました．

1つは病名告知の問題です．例えば遺伝性の要因が非常に強い疾患を持つ患者に対して，どのようにして病名告知を行い，研究の同意を得るかということがあります．また，そのことに関連して，そこで得られた個人情報がどのように本人や家族にフィードバックされたらいいのかという問題があります．

例えばある研究で，遺伝性疾患に罹患している子どもとその親から試料をいただきたいということだったのですが，その子どもは意識障害があって判断能力がなかったので，親から説明と同意を得るということで倫理委員会は研究を許可しました．しかし親の方では，検査の結果，どちらの親に責任があるかということになると困る，家庭争議に発展すると困るということで同意が得られませんでした．このような説明と同意の問題は遺伝子解析研究で非常に重要な点です．その他には一般的な問題として，先程から申し上げましたように，試料をどのようにして採取するか，あるいはほかの目的で採った試料を研究に使ってもいいかという問題が遺伝子解析の研究でも起こります．

6．三省指針について

我々はこのような問題について検討を重ねて，13の研究について答えを出してきたのですが（表1），平成13年3月29日付で文部科学省，厚生労働省，経済産業省がいわゆる「三省指針」を出しました．そこには，今言ったような疑問に対する回答が含まれています．研究を行う皆さんには，ぜひ熟読していただきたいと思いますが，今日は簡単に先程の3つの論点と絡めて説明をしたいと思います．

この三省倫理指針の項目としては，基本的な考え方や研究者の責務などがあげられており，研究責任者，あるいは個人情報管理者，倫理委員会の責務という問題や，提供者に対するインフォームド・コンセント，遺伝情報の開示，遺伝カウンセリング，試料等の取り扱いについてそれぞれ克明に述べられています．

1）病名告知・情報公開

最初に，先程の病名告知の問題と情報のフィードバックについて，三省指針はどのように言っているかについて説明したいと思います．この中では，「知る権利」と「知らないでいる権利」ということが述べられていますが，「基本的に提供者は，研究の結果明らかになった自己の遺伝子情報を知る権利を有する」というのが大前提です．ただ，そこで問題になるのは，試料は匿名化されなければいけない，だれであるかがわかってはいけないということも，もう一つの原則になっています．その場合，連結不可能，つまりあとで調べようと思っても，調べることができない，したがってだれであるかがわからない方式をとったときには，どんな結果が出たとしても，提供者は知ることができないわけですから，最初にそういう状況にあるということを伝えて，同意を得ておかなければいけないわけです．

また，得られた遺伝情報が必ずしも診断に結びつかないことも少なくないわけですから，役に立つ情報が提供されるとはかぎりません．したがって，その情報の意味や有用性，あるいは診断についての意味合いについても十分に説明して，情報の持つ意味や，それがわかったからといって，すぐに診断や治療に結びつくわけではない場合のあることについても説明をしなければいけないわけです．特に多くの研究は多施設で，大規模にやりますので，匿名化が行われていると一層，フィードバックが困難になります．ですから，そのような状況であることをきちんと説明したうえで，同意を得なければいけないわけです．

一方，先程の子どものケースでお話ししたように，研究の結果明らかになった自分の遺伝子情報を，知

らないでいる権利もありますので，それについても十分に説明されるべきです．ただ，予防が可能であったり，治療や薬剤の副作用の予測が可能であるような結果が出たときには，提供者の人命を救うという意味で，遺伝情報と判断が提供者に伝えられる方がいいという場合もあるわけです．ですから，そのことについてはインフォームド・コンセントのときに説明する必要があります．時には検査した本人だけではなく，血縁者にまで影響が及ぶ場合もあります．その場合でも，原則として提供者本人の承諾がなければ情報を伝えてはならないのですが，それが疾病に関する重要な意味を持っていたり，予防治療が可能とみられる場合には，倫理委員会の審議を経て血縁者にその判断を伝えることができます．遺伝解析をした結果出てきたものをどのようにフィードバックするか，あるいはしないかということは，遺伝子研究の中では非常に重要なテーマであるわけです．要するに，いろいろな場合を考えてインフォームド・コンセントを取りなさいということです．

もし，インフォームド・コンセントを取るときには予想されなかったような，重要な結果が出たときには，倫理委員会にもう一度，審議を依頼したうえで，どうしたらいいかを決めることになります．この情報の提供とフィードバックは，遺伝子の解析研究では非常に重要な意味を持っていますので，この点については十分に配慮する必要があるのではないかと思います．しかも，単に結果を知らせればそれでいいというわけではありません．得られた結果の意味や，時には遺伝カウンセリング等の社会的・心理的な支援を提供することも求められています．

2) 差別の禁止

もう1つ，差別の禁止についてですが，アメリカなどでも遺伝子解析で得られた結果が生命保険会社などに伝わって，いろいろな問題を起こしていますし，今後，日本でもそういう事態が起こり得ると思います．そのために「提供者の遺伝情報は，ヒトとしての多様性を示す基盤であり，提供者は研究の結果明らかになった遺伝情報が示す遺伝的特徴を理由に，差別されてはならない」とされてます．我々は心して，そういう問題を考えなければいけないと思います．

3) 個人情報匿名化と管理

個人情報匿名化と管理の問題も，遺伝子解析の研究では非常に重要な課題です．外部に情報が漏洩しないようにすることや，個人情報から個人を識別する情報の全部または一部を取り除いて，符号化・匿名化することが必要です．匿名化の方法として，「連結可能匿名化」といわれるものがありますが，これはある順番で番号をつけ，あとで対応表をみれば個人が特定できる場合をいいます．一方，「連結不可能匿

表 1

遺伝子解析研究

申請番号	申請課題名	申請者	所属	申請年月日(平成)
28	エストロゲン応答遺伝子の解析	村松正実	第二生化学	7年10月17日
32	概日リズム障害を伴う疾患の遺伝子解析	山内俊雄	精神医学	8年1月18日
41	遺伝性進行性ミオクローヌスてんかん症におけるメチル転移酵素の遺伝子発現に関する分子遺伝学的解析	横山富士男	精神医学	9年5月26日
42	概日リズム障害を伴う疾患の遺伝子解析(2)	山内俊雄	精神医学	9年8月20日
77	脊椎後縦靭帯骨化症の原因遺伝子の解析に関する研究	都築暢之	医療センター整形外科	12年9月29日
93	骨髄異形成症候群の遺伝子研究	松田　晃	第一内科	13年7月7日
95	造血幹細胞移植時の移植片対宿主病の重症度を支配する遺伝子群の研究	川井信孝	第一内科	13年5月25日
96	Ph染色体陽性慢性骨髄性白血病の多様性に関する原因遺伝子及びその遺伝子多型に関する研究	矢ヶ崎史治	第一内科	13年5月25日
104	尋常性乾癬感受性領域の全ゲノム高解像度マッピング	土田哲也	皮膚科	13年9月20日
106	シクロスポリンの作用機序および薬剤感受性に関する遺伝子研究	別所正美	第一内科	13年7月2日
109	消化管疾患(潰瘍、炎症、腫瘍)の発生機序に関する研究	太田慎一	第三内科	13年10月12日
112	混合性結合組織病患者の自己抗原リボヌクレオプロテイン遺伝子の変異の解析	大久保光夫	医療センター輸血・細胞治療部	13年10月9日
115	血管内皮一酸化窒素合成酵素の遺伝子多型と川崎病に関する研究	小林　順	小児科	13年9月27日

表 2

審議を通じて明らかになった問題点

1．一般的な問題点

1）研究目的・計画の明確さ
2）対象選択の方法
3）研究期間，研究の方法
4）適切な試料の採取と利用
5）予想される危険性，利益，不利益
6）個人の秘密の保持
7）結果の公表，本人への情報提供
8）適切な説明と同意の文書

2．遺伝子解析研究で特に問題となる問題点

1）病名告知の問題
2）個人情報の匿名化と情報の feed back の問題
3）試料の採取と目的外使用

表 3

説明と同意の文書に記載すべきこと

① 研究の目的
② 研究の方法
③ 予想される効果
④ 予想される危険
⑤ 試験により患者が受ける苦痛
⑥ 他の治療方法の有無（選択の自由の説明を含む）
⑦ 研究からの離脱について
（医師の判断で中止する場合・患者の希望で中止する場合）
⑧ 人権保護（公正におこなうこと，プライバシーを守ること）
⑨ 事故の対応（どのような時，どのような緊急措置をとるか）

名化」はそういう対応表もなく，対応のしようもないというものです．いずれかの匿名化をしますが，その管理のために情報管理者を設けなければいけないとされています．

これについては，本学の倫理委員会で許可した申請が多施設研究だったために，情報管理者がいないという理由で，本学では研究が進まない状況になっているものがあります．早急に対応しなければいけないと思っています．情報管理者の役割，責任についても指針では細かに規定しています．

試料の採取と適用外試料については省略しますが，本来の目的ではないことで採ってあった試料はどう用いることができるかということも含めて指針に書かれています．

4) 全体の流れ

これらの流れをまとめますと図 3のようになります．研究しようとする責任者，あるいは担当者は研究計画書を作って，それを研究機関の長，本学では学長に申請します．そして学長から倫理委員会に諮問があり，それを受けて倫理委員会は審議し，その結果を答申し，それに基づき研究機関の長は判断をします．許可が得られると，研究者は試料提供者に説明し，同意を得る，あるいは血縁者や代諾者にお願いするということになります．そういった手続きを経て研究が行われるわけですが，そのときに個人情報管理者を設けて，匿名化をすることが特に求められています．また，遺伝カウンセリングも重要な1つの柱になっています．もう1つは，本当にそれが適切に行われているかどうかを実地調査し，その結果を倫理委員会あるいは個人情報管理者に伝えるよう求めています．

7．本学倫理委員会の抱える問題

三省指針をもとに，埼玉医大がヒトゲノム，あるいは遺伝子解析研究をするときに抱える倫理的問題点について考えてみたいと思います．

1) 委員の構成

倫理委員会の委員構成の問題があります．三省指針に書いてあるところに従うと，倫理・法律を含む人文・社会科学領域の有識者，自然科学領域の有識者，一般の立場の者から構成されなければいけないと書かれています．また，外部委員を半数以上置くことが望ましいが，困難であれば複数名置かなければいけない．しかも，外部委員の半数以上は人文・社会科学領域の有識者，一般の立場の者でなければいけない．男女両性で構成されなければいけない．運営に関する細則でも，一般の立場の委員が1名以上出席しなければならないなど，いろいろと事細かに述べられています．

現在の埼玉医科大学倫理委員会のメンバーで女性委員は五十嵐先生お一人です．両性で構成しなければいけないという意味では良いのですが，一般あるいは人文・社会学の方が入るといった点で問題があります．

もうひとつの問題は，迅速審査という制度があり，時間をかけずに審査をする方式を認めていますので，これも今後，導入してかなくてはなりません．また，個人情報管理者の問題があります．これは試料や遺伝情報の匿名化，補助者の監督という役割をもち，違反した場合の罰則も厳しく記載されています．

そのほかに遺伝カウンセリングという問題がありますが，本当の意味で家族，血縁者あるいは提供者に正確な情報を提供し，疑問に適切に答え，遺伝性疾患等に関する理解を深めて不安や悩みに答えるという援助・支援をするカウンセリング制度を整備し，習熟した医師，医療従事者が協力して行わなければならないので，その制度を作る必要があります．また，今後は実地調査の制度を作り，適切にこのシステムが運営されているかをチェックする外部査察のような機構も，本学として備えなければいけないのではないかと思います．

そのようなことで，本学の倫理委員会もまじめにこの問題に取り組もうとしていますがまだいくつもの課題をかかえていますので，今後も解決に努力していきたいと思いますので，是非，皆さんにその趣旨を理解していただいて，適切な運用にご協力していただきたいと思います．

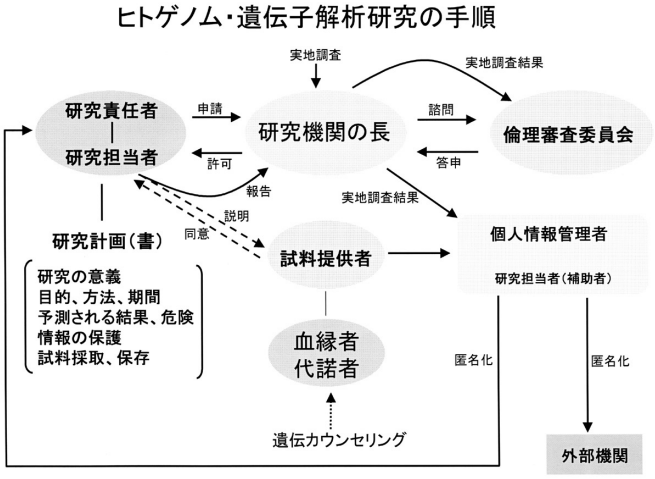


図 3

座長：ありがとうございました．現状と問題点についてわかりやすくお話しいただました．この問題は話を始めますと長いのですが，フロアからのコメントやご意見はおありでしょうか．五十嵐先生は唯一の女性の倫理委員として，委員会の活動の中でこれに関連して

ご意見・ご発言はありますか.

五十嵐：女性といっても年を取っておりますので，あまりそういう意識をしてくださらないと思っておりますが，今，そこに出ました条件だけは満たしていると思いますので，その言葉で許していただきたいと思います．条件として，人文科学など3つ，4つ出ましたが，女性であることで．本当は辞退しようと思いましたが，今度は外部としてもう少しということで，よろしいでしょうか.

村松：一言だけ追加をさせていただきます．私は分子生物学的にDNAをどのように見るかということから相談され，委員として出席させていただいていますが，審査を受ける用紙に書いてくる文章にはなってないものが多く，悩まされております.

第1には，本気で書いていないということがあるのではないかと思うのです.

第2には，そういうものの書き方をトレーニングされていないことがあります．ここにおられる方はいいと思いますが，欠席されておられる方は，目的や方法などをしっかり書いて，委員が読みやすいものを書いていただきたいのです．何をやろうとしているのか，何を目的としているのか，本当にわかりません．委員としてそういうむだな時間を費やすことが多いということを，私の方から追加しておきます．よろしくお願いします.

座長：ほかにありませんか．では時間ですので，どうもありがとうございました.

ご発表はこれで終了しましたが，予定の時間を少し過ぎてしまいました．本当は15分ほどディスカッションの時間を取ってあったのですが，特に外からお忙しい中をおいでくださった先生方に，ここはお聞きしたいという方がいらっしゃいましたら，どうぞ.

なければ，シンポジウムはこれで終了させていただきます．終了にあたり，本医学会の会長であります東学長から一言，お願いしたいと思います.

東学長：本学のゲノム医学研究センター設立記念にあたり，このようなゲノムに関するきわめて先端的なお話を伺いまして，ありがとうございました．お忙しい中，しかも夕方から数時間にわたりお話をしていただき，本当に感謝の言葉もございません．また，プログラムを見たときには，最後に山内先生がどのようにお話しになるのかと考えておりましたが，このゲノムに関する諸々の倫理的側面からの見方，考え方，やり方をきわめてわかりやすくお話しいただき，ありがとうございました.

座長：ありがとうございました．これにて公開シンポジウムを終了いたします.