

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 栗田卓也 (埼玉医科大学第4内科学)

第4内科の栗田です。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 などがありますが、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

MODY 1	(1996)	HNF-4α
MODY 2	(1992)	Glucokinase
MODY 3	(1996)	HNF-1α
MODY 4	(1997)	IPF 1
MODY 5	(1997)	HNF-1β
MODY 6	(1999)	NEUROD1/BETA2

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、IBD=0が1, 2の方に偏っていきます。そういう偏りが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ィールですが、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。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が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。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、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いのですが、その当時はまだこの領域のシーケ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	D1S4464	LOD=1.7
	1q21-23	D1S1677	LOD=2.48
	2q37	NIDDM1	LOD=4.02
Mexican American	15q	CYP19	LOD=1.27
	1q21-23	APOA2	LOD=4.3
Utah Caucasian	12q	NIDDM2	LOD=3.65
Finnish	1q21-24	APOA2-D1S484	MLS=3.04
	3q27-qter	D3S1580	MLS=4.67
FUSION	20p	D20S849	MLS=2.15
		D20S107	MLS=2.04
	20q	D20S886	MLS=1.99
		D11S901	MLS=1.75
		D2S319	MLS=0.87
GENNID	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。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、

NIDDM1領域の遺伝子と遺伝子多型

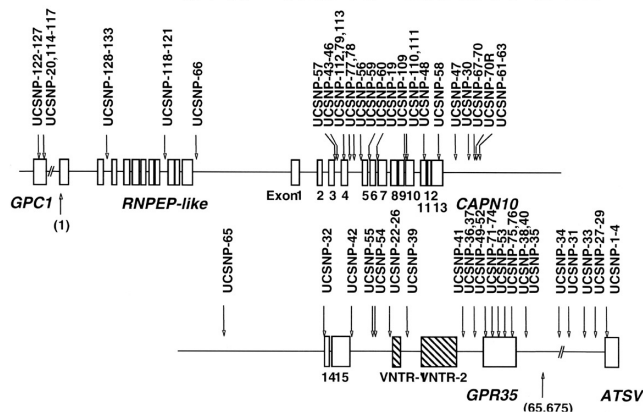


図3

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をやったのが、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сп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、レポータージーンが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行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なカルパイン・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だけに對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。実際の臨床の場合でも、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られています。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られています。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。既にセセラ社では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は節約遺伝子といわれ、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と。ですから、このアソシエーション・スタディー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、検出率がやると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という弱いクリテリアでいくと、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53億円の儉約ができます。

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。こういう苦勞がありますが、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、今は力づ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。なぜかということ、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する情報は。

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か。

堀川：正確には、そ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。