

<p><u>Nagao Y, Hiromatsu Y, Nakashima T, Sata M.</u></p>	<p>Graves' ophthalmopathy and tongue cancer complicated by peg-interferon <math>\alpha</math>-2b and ribavirin therapy for chronic hepatitis C: A case report and review of the literature.</p>	<p>Molecular Medicine Reports</p>	<p>1</p>	<p>625-631</p>	<p>2008</p>
<p><u>Nagao Y, Kawakami Y, Yoshiyama T, Sata M.</u></p>	<p>Analysis of factors interfering with the acceptance of interferon therapy by HCV-infected patients.</p>	<p>Med Sci Monit.</p>	<p>14</p>	<p>45-52</p>	<p>2008</p>
<p><u>長尾由実子, 今福信一, 佐田通夫.</u></p>	<p>肝臓病の方の皮膚や粘膜には、さまざまな症状が現れます。とくにインターフェロン治療中には注意が必要です。-肝臓病と皮膚・粘膜の病気-</p>	<p>肝外病変シリーズ</p>	<p>2</p>	<p>1-22</p>	<p>2008</p>

---

#### IV. 研究成果の刊行物・別冊

# 第1章 肝炎の基礎

## (1) ゲノムワイド関連分析による多因子疾患遺伝子の探索

西田奈央, 徳永勝士

### 重要なポイント

- 大規模なSNP解析技術の進展に伴い、ゲノムワイド関連解析は数多くの疾患感受性遺伝子探索研究に用いられるようになっていく。
- Affymetrix 500K Human Mapping Array Set, および Affymetrix Genome-Wide Human SNP Array 6.0によるSNPタイピングプラットフォームを構築し、ゲノムワイド関連分析を行う環境・装置を整備した。
- 大規模なSNP解析技術は、ヒトゲノム多様性に関連するさまざまな研究分野に画期的なインパクトを及ぼすことは疑いないが、同時にその膨大な多型情報を十分に活用するノウハウを確立することが必要とされている。

### はじめに

ゲノムワイド関連分析法は、ヒトのさまざまな多因子疾患に関わる遺伝子を探索する戦略として近年大きな注目を浴びている。多因子疾患感受性遺伝子を探索する統計遺伝学的手法には、大別して連鎖分析 (Linkage analysis) と関連分析 (Association analysis) がある。連鎖分析は患者家族を対象として文字どおり疾患遺伝子と多型マーカーの連鎖を検出する手法であり、ゲノム全域に分布する1万から数万種のSNPs (Single Nucleotide Polymorphisms) を用いればよい。一方、関連分析の代表であるケース・コントロール関連分析法は、非血縁の患者集団と健常者集団を対象として疾患遺伝子と多型マーカーの連鎖不平衡 (Linkage disequilibrium) を検出する手法であり、ゲノム全域に分布する数十万種以上のSNPsを解析することが必要と

なる<sup>1)</sup> (統計遺伝学的手法の原理ならびに集団遺伝学的基础については別に解説した<sup>2)</sup>)。

従来のゲノムワイド探索では、ゲノム中に約10cMの間隔で存在するマイクロサテライトを遺伝マーカーとして用いた連鎖分析法が広く用いられてきた。近年では、ゲノム中での存在密度が高いSNPを遺伝マーカーとして用いたゲノムワイド連鎖分析法が行われるようになり、その手法を用いて家族性乳癌の原因遺伝子としてよく知られている *BRCA 1/2* のほかに、新たな5カ所の候補遺伝子領域を特定したという報告もなされた<sup>3)</sup>。

一方で、ゲノムワイド関連分析は、本稿で紹介する大規模なSNP解析技術の進展も相まって、数多くの疾患感受性遺伝子探索研究で用いられるようになっていく。例をあげると、WTCCC (The Wellcome Trust Case Control Consortium) は7種類のCommon diseases [双極

性感情障害 (BD)、冠動脈疾患 (CAD)、クローン病 (CD)、高血圧 (HT)、関節リウマチ (RA)、1型糖尿病 (T1D)、2型糖尿病 (T2D) について、それぞれ2,000人の患者とコントロールとして健康者3,000人の計17,000人を対象とした大規模なGWAS (Genome-Wide Association Study) を行った<sup>4)</sup>。また、米国NIHはGWAS計画を提案し、いくつかのCommon diseasesについて大規模な研究チームを公募した。さらに、大規模な疫学研究として知られるFramingham Heart Studyで収集された試料のうち9,000検体について、本稿で紹介する技術を用いて解析することにより、心、肺、血液、睡眠疾患に関与する遺伝子変異を探索する計画が発表され、ゲノムワイド関連分析およびゲノムワイド連鎖解析などを行った結果が2007年にまとめて報告された<sup>5)</sup>。

SNPタイピング法としてよく知られている方法は、個々の多型部位を含むゲノム断片を特異的にPCRで増幅した後にアリルを識別する方法である<sup>6-8)</sup>。一方、近年になって多型部位特異的なPCR (Polymerase Chain Reaction) を行わずに大規模なSNPタイピングを行う方法が実用化された<sup>9-11)</sup>。

その一つであるAffymetrix社によって確立された方法では、まず制限酵素反応でゲノムDNAを断片化し、続いてそれら断片の両端にアダプター配列を付加して増幅した後に、マイクロアレイを用いたアリル特異的なハイブリダイゼーションを行う<sup>10)</sup>。2007年5月には、90万種を超えるSNP解析用プローブおよびCNV

(Copy Number Variation) 解析用の94万種類を超えるプローブを搭載したキットが市販された (Affymetrix® Genome-Wide Human SNP Array 6.0)。われわれはGenome-Wide Human SNP Array 6.0および、約50万種類のSNP解析用プローブを搭載した500K Human Mapping Array Setを用いたSNP解析プラットフォームを構築し、文部科学省科学研究費特定領域研究「基盤ゲノム」におけるヒトSNPタイピングセンターとして、いくつかの多因子疾患についてゲノムワイド関連分析を実施している。

本稿では主として技術面を取り上げ、ヒトSNPタイピングセンターに構築した500K Human Mapping Array SetによるSNP解析プラットフォームについて紹介し、最後の章でGenome-Wide Human SNP Array 6.0の概要を紹介する。なお、統計遺伝学的側面については別稿を参照いただきたい<sup>12)</sup>。

## 1 技術原理

500K Human Mapping Array Setは、制限酵素によるゲノムDNAの断片化とマイクロアレイによるタイピングの手法に改良を加えることにより、大規模なタイピングを行える手法として確立された<sup>10)</sup>。解析対象となるSNPは、公共のSNPデータベースおよびPerlegen社に登録されている約220万種のSNPから遺伝学的情報量が最大化されるように、また連鎖不平衡やHapMapプロジェクトからの情報も考慮して選

### 用語の解説

#### ケース・コントロール関連解析

ある疾患に罹っている患者群 (ケース) と健康者群 (コントロール) とで遺伝子・ゲノム多型の頻度に差があるかどうかを検定することにより、疾患関連遺伝子を探索するための統計遺伝学的方法。

#### ハーディー・ワインベルグ平衡

自然淘汰が働かず、突然変異によって新たな対立遺伝子が生じず、また移住や混血などが起こらない十分に大きな規模の集団においては、対立遺伝子の頻度は世代を経ても変化しないという集団遺伝学の基本的法則。



扱われた約50万種のSNPである。

500K Human Mapping Array SetによるSNPタイピングは、ゲノムの複雑さを低減しマイクロアレイへのハイブリダイゼーション効率を上げるための酵素反応ステップと、洗浄・染色装置およびマイクロアレイ用スキャナーを用いた検出ステップに分けることができる(図2-1-1)。

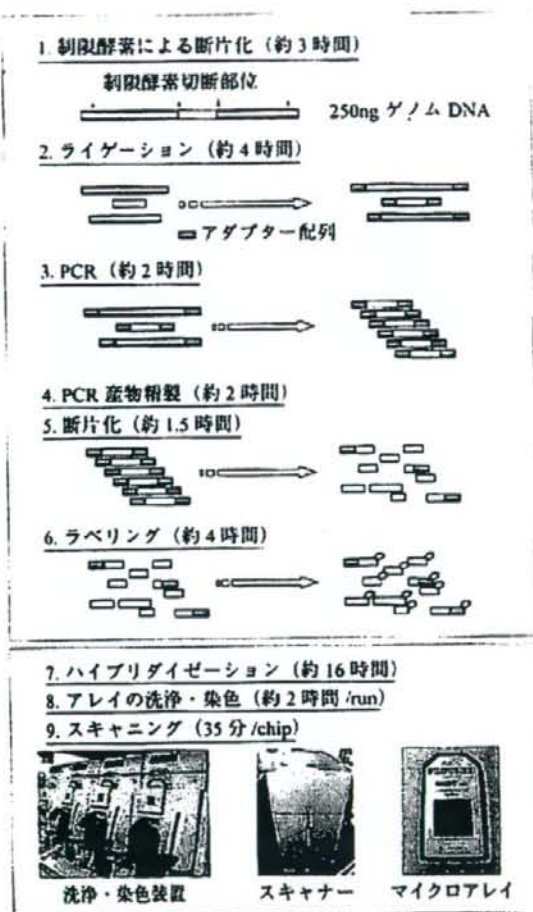


図 2-1-1 500K Human Mapping Array Set による SNP タイピングの原理

制限酵素による断片化からラベリングまでの酵素反応で得られた産物を洗浄・染色装置およびマイクロアレイ用スキャナーで検出することにより、約50万種類のSNPのタイピングを行う。

## 1. ゲノムDNAの断片化とアダプター配列の付加

50万種のSNPタイピングは、2種類の制限酵素 (Sty I, Nsp I) を用いてそれぞれ約25万種のSNPを独立にタイピングすることで実現される。制限酵素によるゲノムDNAの断片化を行った後、断片化されたゲノムDNAの両末端にアダプター配列をライゲーション反応により付加する。アダプター配列は、続くPCRで使用されるプライマーと相同な配列をもち、また制限酵素認識配列を突出端としても2本鎖DNAである。2種類の制限酵素 (Sty I, Nsp I) のそれぞれに対して用意されるアダプター配列は、制限酵素認識配列を除いて共通の配列をもっているため、共通のプライマーを使用してPCRを行うことができる。

## 2. ゲノムDNA断片の増幅

PCRでは、目的の長さをもったゲノムDNA断片 (200 ~ 1,100 bp) だけが選択的に増幅される。ここまでの酵素反応により、もともと30億塩基対のゲノムDNAが5億塩基対程度のPCR混合産物となる。マイクロアレイへの効率的なハイブリダイゼーションには、ゲノムの複雑さを低減することが大きな役割を果たすと考えられている<sup>[13], [14]</sup>。

## 3. 増幅産物の断片化とビオチン導入

続いて、PCR産物の精製を行った後、DNase I 制限酵素によりPCR産物の断片化を行う。断片化されたPCR産物は平均長で180bp以下となる。マイクロアレイへの効率的なハイブリダイゼーションには、ゲノムの複雑さを低減することに加えてPCR産物の断片化が重要になる。最後にTerminal deoxynucleotidyl transferase 酵素反応により、断片化されたPCR産物の末端にビオチンを導入する。

## 4. アレイへのハイブリダイゼーション

続いて、専用のマイクロアレイを用いてハイブリダイゼーションを行う。マイクロアレイには解析対象となるそれぞれのSNPに対して合計24本のプローブが用意されている。プローブは25塩基長のオリゴDNAで、SNP部位を含む塩基配列をもっている。また、2種類のアイルに対して完全に相補的な塩基配列をもつプローブ（PMプローブ）と一塩基のミスマッチを含むプローブ（MMプローブ）が用意され、4種類のプローブが1組のプローブセットとなる。

SNP部位を25塩基長のプローブの中心に置いたプローブセットを基本として、SNP部位を中心から4塩基上流（+4）にずらしたプローブセットから4塩基下流（-4）にずらしたプローブセットまで、7組のプローブセット（-4, -2, -1, 0, +1, +3, +4）のなかから3組のプローブセットを選択する。3組のプローブセットはゲノムDNAの両鎖に対して用意されているので、合計6組のプローブセットが各SNPに対して用意されている。また、遺伝学的に重要だとされる約52,000SNPsに関してはプローブセットを合計10組使用し、計40本のプローブを用意している。すべてのプローブセットのPMプローブとMMプローブのシグナル強度差から非特異的なクロスハイブリダイゼーションを定量化することで、正確な遺伝子型の判定を行うことができる。

## 5. アレイの洗浄・染色

マイクロアレイへのハイブリダイゼーションが終了した後、洗浄・染色装置を用いてマイクロアレイの洗浄および蛍光染色を行う。蛍光染色は、蛍光分子で標識されたストレプトアビジンが、上述のビオチン導入されたPCR断片に結合することにより行われる。また、洗浄・染色装置内ではビオチン修飾された抗ストレプトアビジン抗体を用いてシグナルの増強が行われる。

## 6. アレイのスキャンニング

最後に蛍光染色されたマイクロアレイを専用のスキャナーで画像データとして読み取り、続いて専用の画像解析ソフトウェアを用いて各SNPの遺伝子型を決定する。

## II. システム構築

## 1. ハードウェアの整備

## 1) 設備

## 500K Human Mapping Array SetによるSNPタ

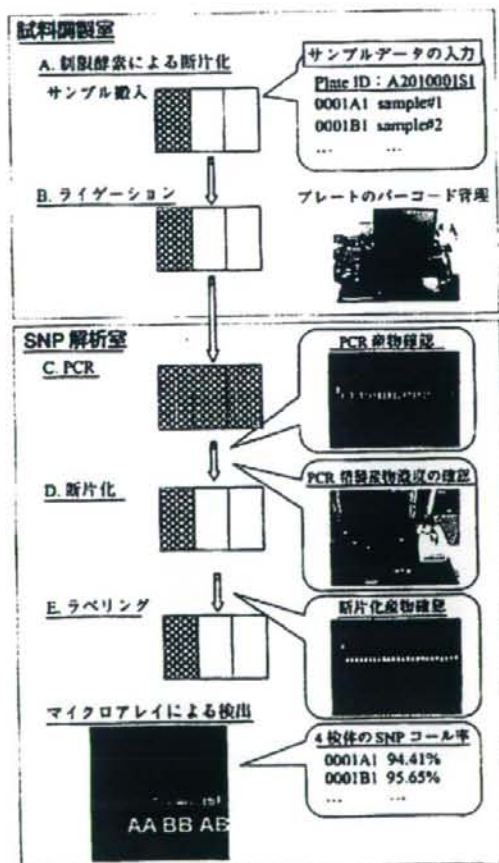


図2-1-2 SNPタイピングシステムの構築

500K Human Mapping Array Setを用いたSNPタイピングを効率的に行うためのシステム環境を構築し、作業マニュアルを作成した。96ウェルマイクロタイタープレートを使用して32検体ずつ解析を行う。



イピングを効率的に行うために、環境、装置を整備し、作業マニュアルを作成した。まず、ゲノムDNAへのPCR産物のコンタミネーションを防ぐために、試料調製室とSNP解析室を設けた。試料調製室ではゲノムDNAを保管し、PCRまでの酵素反応を行うのに必要な装置（サーマルサイクラーなど）を用意した。制限酵素による断片化からPCRの反応溶液の調製までを試料調製室で行い、PCR以降の酵素反応はSNP解析室で行った。また、4台の洗浄・染色装置を用意し、1回のランで計16枚のマイクロアレイを洗浄・染色することができるようにした。すべてのマイクロアレイはバーコードで管理され、洗浄・染色が終了したマイクロアレイはオートローダー付きのマイクロアレイ用スキャナーに装填され画像データが読み込まれる。オートローダー付きのマイクロアレイ用スキャナーは計64枚のマイクロアレイを装填することができる。バーコードを参照しながらすべてのマイクロアレイの画像データを自動的に読み込むことができる。

## 2) マイクロタイタープレートのバーコード管理とサンプルのID化

SNPタイピング作業のルーチン化にあたって、制限酵素による断片化からラベリングまでの五つの酵素反応ステップで使用すべてのマイクロタイタープレートをバーコードで管理する(図2-1-2)。また、1枚のマイクロタイタープレートで32検体分の酵素反応を行うこととし、制限酵素による断片化で使用するマイクロタイタープレートのウェル位置をサンプルと対応させることでサンプルのID化を行った。制限酵素による断片化以降の酵素反応で使用するマイクロタイタープレートでもレイアウトを変えずに酵素反応を行うことで、ウェル位置をサンプルIDとして解析結果を得ることができる。

## 3) チェックシートによる進行管理と

### サンプルの品質管理

酵素反応の各工程を管理するためにチェック

シートを作成し、反応工程の進行を随時確認しながら進める。PCRおよび断片化の酵素反応の後にはアガロースゲル電気泳動を行い、PCR産物および断片化産物の平均長がそれぞれ200～1,100bp、180bp以下となっていることを確認する。また、PCR産物精製後の濃度を吸光度計で測定して2 $\mu\text{g}/\mu\text{l}$ に調整し、濃度が不足したサンプルは破棄する。加えて、同一のマイクロタイタープレート上で酵素反応を行った32検体のうちから4検体だけを先行してハイブリダイゼーションを行い、コール率に問題がないことを確認した後で残り28検体のハイブリダイゼーションを行った。28検体のハイブリダイゼーションを行う際に、次のマイクロタイタープレートから4検体を加えて合計32検体のハイブリダイゼーションを順次行っていくこととした。

## 2. ソフトウェアの開発

500K Human Mapping Array SetによるSNPタイピングでは、GeneChip<sup>®</sup> Operating Software (GCOS)とGeneChip<sup>®</sup> Genotyping Analysis Software (GTYPE)という2種類のソフトウェアを使用する。GCOSソフトウェアは洗浄・染色装置およびマイクロアレイ用スキャナーを操作する際に使用し、またGTYPEソフトウェアはマイクロアレイのシグナル強度データから遺伝子型を判定する際に使用する。GTYPEソフトウェアで決定された約25万種のSNPの遺伝子型は、Sty I、Nsp Iごとにテキストファイルとして転送することができる。

われわれは500K Human Mapping Array SetによるSNPタイピングから得られる約50万SNPの遺伝子型情報を用いて、ケース・コントロール関連分析を行うためのソフトウェアを開発した。ケース・コントロール関連分析をするに当たって、Sty IおよびNsp Iごとにまとめられた約25万SNPの遺伝子型データを検体ごとに統合し、さらに検体をケース群およびコントロール群に分けて、新たなテキストファイルと

して作成する機能をソフトウェアに加えた。

続いて、作成したケース群およびコントロール群の解析結果のテキストファイルを使ってケース・コントロール関連分析を行った。この際、各コントロール群における各遺伝子型の観察数からハーディー・ワインベルグ平衡の検定を行うこととした。ケース・コントロール関連分析の結果はレポートファイルとしてまとめられ、専用のビューアーを用いて表示することができる。



### Genome-Wide Human SNP Array 6.0によるSNP解析

SNP解析の技術原理は第1章で紹介した500K Human Mapping Array Setと基本的に同じで、2種類の制限酵素 (Sty I, Nsp I) を用いてゲノムDNAを断片化し、続いてそれら断片の両端にアダプター配列を付加してPCRで増幅した後、マイクロアレイを用いたアリアル特異的なハイブリダイゼーションを行う。

両アッセイで大きく変わった点として挙げられるのが、500K Human Mapping Array Setでは1検体につき2枚のマイクロアレイを必要としたが、Genome-Wide Human SNP Array 6.0ではSty I, Nsp IそれぞれのPCR後の産物を混合してからマイクロアレイにハイブリダイゼーションをすることにより、1検体につき1枚のマイクロアレイですむようになった点である。ただし、90万種類以上のSNP用プローブおよびCNV用の94万種類以上のプローブを1枚のマイクロアレイに集積するために、1種類のSNPに対するプローブの選択方法が大きく変わった。

500K Human Mapping Array Setでは前述したとおり1種類のSNPにつき24本のプローブを用意していたが、Genome-Wide Human SNP Array 6.0では最適な1本のPMプローブを4スポット用意するという方法に変わった。また、SNP解析用プローブの選択方法の変更に伴い、新

たな遺伝子型決定用ソフトウェア (Genotyping Console 1.0) が用意された。

500K Human Mapping Array Setに搭載された約50万種類のSNPのうちの約44万種類のSNPに加えて、タグSNP、X染色体やY染色体上のSNP、ミトコンドリアSNP、データベースに新たに登録されたSNP、組み換えが多発する領域に存在するSNPなどを含めた合計90万種類以上のSNP解析用のプローブがGenome-Wide Human SNP Array 6.0には搭載されている。複数の施設で行われたGenome-Wide Human SNP Array 6.0によるSNP解析の結果から、平均コール率が99%以上となり、また、HapMapデータベースに登録されたSNPとの遺伝子型一致率は99.7%を超えることがAffymetrix社から報告されている。実際に東京大学ヒトSNPタイピングセンターで行った日本人健康者サンプル200名のGenome-Wide Human SNP Array 6.0によるSNP解析の結果でも平均コール率が99.73%となり、Affymetrix社が公表する水準を満たす結果となることが明らかとなっている。

### おわりに

90万種類以上のSNPを一挙にタイピングできる技術の実用化によって、従来は存在しなかった広範かつ詳細なヒトゲノム多型情報が得られる時代となった。このような情報は、疾患遺伝子探索研究のみならず、人類の進化や人類集団の歴史を解明する糸口を提供し、ヒトゲノム多型性に関連するさまざまな研究分野に画期的なインパクトを及ぼすことは疑いない。

しかしながら同時に、われわれは、まだ得られる膨大な多型情報を十分に活用できるノウハウをもっていないことも指摘しておきたい。われわれの用意した統計解析ソフトウェアは個々のSNPについては関連分析を行えるものの、ハプロタイプについての関連分析はできない。市販のソフトウェアにも、限定した領域でハプロタイプ関連分析ができるものはあるものの、



ゲノム全域にわたって一挙に分析できるものはない。

また、ケース・コントロール関連分析により疾患との関連が検出されるSNPのなかには多くの擬陽性が含まれていることがわかっており、真の疾患関連遺伝子を検出するためには得られた膨大な多型情報のなかからタイピング精度の高いSNPを選択することが重要になる<sup>15)</sup>。さらに、タイピングしたサンプルのなかには、血縁関係にあるものやサンプル重複などが含まれている可能性がある。血縁者がサンプルに含まれていることが擬陽性の増幅につながるものが知られており、それらの関連分析に不適切なサンプルを取り除くことも真の疾患関連遺伝子を検出するためには重要となる<sup>16)</sup>。

さらに、多数の検体について得られたゲノムワイド多型解析情報から、これまでまったく知られていなかった新しい遺伝子/遺伝子相互作用が見出される可能性がある。しかし残念ながら、従来の統計学的手法や計算アルゴリズムでは、このように膨大なデータを実用的に処理できない。

このように、ゲノムワイド多型解析情報は、バイオインフォマティクスに関わるさまざまな研究者にとって挑戦に値する多くの課題を提供してくれるとともに、その達成によって従来にない実り豊かな成果をもたらしてくれるに違いない。

## 文 献

- Ohashi, J., Tokunaga, K. : The power of genome-wide association studies of complex disease genes: statistical limitations of indirect approaches using SNP markers. *J. Hum. Genet.* 46 : 478-482, 2001
- 徳永勝士 編: 人類遺伝学ノート. 南山堂, 東京, 2007
- Gonzalez-Neira, A., Rosa-Rosa, J.M., Osorio, A., et al. : Genomewide high-density SNP linkage analysis of non-BRCA1/2 breast cancer families identifies various candidate regions and has greater power than microsatellite studies. *BMC Genomics* 8 : 299-311, 2007
- The Wellcome Trust Case Control Consortium : Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447 : 661-683, 2007
- The framingham heart study 100,000 single nucleotide polymorphisms resource. *BMC Med. Genet* 8 (Suppl 1), 2007
- Syvänen, A.C. : Accessing genetic variation : genotyping single nucleotide polymorphisms. *Nat. Rev. Genet.* 2 : 930-942, 2001
- Kwok, P.-Y., Chen, X. : Detection of single nucleotide polymorphisms. *Curr. Issues Mol. Biol.* 5 : 43-60, 2003
- Nishida, N., Tanabe, T., Takasu, M., et al. : Further development of multiplex single nucleotide polymorphism typing method, the DigiTag2 assay. *Anal. Biochem.* 364 : 78-85, 2007
- Rachagani, S., Gupta, I.D., Gupta, N., et al. : Genotyping of beta-lactoglobulin gene by PCR-RFLP in Sahiwal and Tharparkar cattle breeds. *BMC Genetics* 7 : 31, 2006
- Matsuzaki, H., Loi, H., Dong, S., et al. : Parallel genotyping of over 10,000 SNPs using a one-primer assay on a high-density oligonucleotide array. *Genome Research* 14 : 414-425, 2004
- Oliphant, A., Barker, D.L., Stuelpnagel, J.R., et al. : BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping. *BioTechniques* 32 : S56-S61, 2002
- 宮川 卓, 徳永勝士: 疾患感受性遺伝子の探索の実際: ゲノムワイド関連解析を中心として. *最新医学* 62 : 2134-2141, 2007
- Grant, S.F., Steinlich, S., Nentwich, U., et al. : SNP genotyping on a genome-wide amplified DOP-PCR template. *Nucleic Acids Res.* 30 : e125, 2002
- Jordan, B., Johdan, B., Charest, A., et al. : Genome complexity reduction for SNP genotyping analysis. *Proc. Natl. Acad. Sci. USA* 99 : 2942-2947, 2002
- Miyagawa, T., et al. : Appropriate data cleaning methods for genome-wide association study. *J. Hum. Genet.* (投稿中)
- Voight, B.F., Pritchard, J.K. : Confounding from cryptic relatedness in case-control association studies. *PLoS Genet.* 1 : e32, 2005

# Processed Pseudogenes and Their Functional Resurrection in the Human and Mouse Genomes

Hiroaki Sakai, *National Institute of Agrobiological Sciences, Kannondai, Tsukuba, Japan*

Takeshi Itoh, *National Institute of Agrobiological Sciences, Kannondai, Tsukuba, Japan*

Takashi Gojobori, *National Institute of Genetics, Yata, Mishima, Japan*

The genomes of human and mouse contain thousands of pseudogenes. Processed pseudogenes, which were derived through retrotransposition, were recognized as 'dead on arrival' elements in the genomes, and they have been thought to have no contribution to genome evolution. However, several recent studies have shown that a significant portion of processed pseudogenes were transcribed and were able to function as noncoding ribonucleic acid (RNA) genes. Since the transcription mechanism and functions of transcribed processed pseudogenes still remain unknown, it is of great importance to elucidate the biological roles of processed pseudogenes by both approaches of experiments and bioinformatics.

## Introduction

Recently, analyses of the fully sequenced genomes of human and other mammals have revealed that their genomes possess a large number of pseudogenes. Although pseudogenes show high sequence similarities to functional genes, their ability of encoding functional proteins appears to have been destroyed. Pseudogenes are classified into two categories, as defined by the mechanisms of their molecular generation. One mechanism is that nonprocessed pseudogenes are derived by genomic duplication or by unequal crossing-over. The other mechanism is retrotransposition, which generates copies of functional genes mainly with mediation by LINE1 that is a member of the LINE (long interspersed nuclear element) retrotransposon family. Copies of genes derived by retrotransposition are referred to as 'retrotransposed copies' or 'retrocopies'. Since retrotransposed copies are reverse-transcribed elements that are derived from transcribed messenger ribonucleic acids (mRNAs), they do not contain the transcriptional regulatory elements of the progenitor genes

(parental genes), neither do they contain the introns of the parental genes. In addition, frequent 5'-end truncations are observed because of incomplete reverse transcription of mRNAs by LINE1 reverse transcriptase. Thus, retrotransposed copies, which are referred to as processed pseudogenes, are recognized as nonfunctional elements. Indeed, in many cases, processed pseudogenes have accumulated a number of mutations that have made their coding sequences nonfunctional. However, recent genome-wide studies of gene expression and individual processed pseudogenes have shown that a fraction of processed pseudogenes in mammals are transcribed possibly with some biological functions (Hirotune *et al.*, 2003; Harrison *et al.*, 2005; Zheng *et al.*, 2007). Such paradoxical phenomena provide an opportunity to reconsider the biological significance of retrotransposed pseudogenes. **See also:** Pseudogenes and their Evolution

In this article, referring to important studies of processed pseudogenes, we describe various aspects of the contribution of processed pseudogenes to genome evolution, such as the rate of emergence, transcriptional activity and biological function. Given the current observations of functional processed pseudogenes, the term 'pseudo' sometimes gives rise to confusion. Thus, we also discuss how to define processed pseudogenes.

## Number and Rate of Emergence of Processed Pseudogenes

It is much easier to find processed pseudogenes than non-processed pseudogenes in the genome sequences, due to the

Advanced article

Article Contents

- Introduction
- Number and Rate of Emergence of Processed Pseudogenes
- Transcriptional Activity of Processed Pseudogenes
- Biological Function of Transcribed Processed Pseudogenes
- Future Prospects for the Study of Processed Pseudogenes

Online posting date: 15<sup>th</sup> July 2008

ELS subject area: Evolution and Diversity of Life

### How to cite:

Sakai, Hiroaki; Itoh, Takeshi; and, Gojobori, Takashi (July 2008) Processed Pseudogenes and Their Functional Resurrection in the Human and Mouse Genomes. In: Encyclopedia of Life Sciences (ELS). John Wiley & Sons, Ltd: Chichester.

DOI: 10.1002/9780470015902.a0021000



mechanisms of their generation. As noted earlier, since processed pseudogenes are derived from reverse-transcribed mRNAs of functional genes, they show high sequence similarities to their parental genes except introns. In addition, various sequence features of functional genes, such as 3'-end polyadenine tails, short direct repeats in flanking regions and 5'-end truncations, can be observed in addition to the inferred original coding capacity. Disabling alterations of the coding capacities in nucleotide sequences of retrotransposed copies, such as premature stop codons and frame-shifts, are convincing evidence that they have lost functionality. However, Torrents *et al.* (2003) showed that half of the human processed pseudogenes observed did not have any frame disruptions. They, therefore, determined the intensity of natural selection by calculating the ratio of the number of nonsynonymous substitutions per site ( $d_N$ ) to the number of synonymous substitutions per site ( $d_S$ ) between processed pseudogenes and their parental genes. In the comparison of a pair of closely related genes, it is known that purifying selection may be operating when the  $d_N/d_S$  value is lower than 1.0, whereas positive selection may be detected when the  $d_N/d_S$  value is greater than 1.0. In the case of neutral evolution, the numbers of synonymous and nonsynonymous substitutions per site are virtually even, since nucleotide substitutions are expected to take place at each site equally. When a processed pseudogene is compared with its parental gene, it is expected that the processed pseudogene evolves neutrally, whereas the parental gene is under purifying selection. Thus, the  $d_N/d_S$  ratio between a processed pseudogene and its parental gene is expected to become larger compared to that between two functional genes. For both cases, of course, the ratios do not become greater than 1.0 unless positive selection is operating.

The numbers of processed pseudogenes estimated to be in the human and mouse genomes differ between different research groups, mainly because of different criteria adopted (Ohshima *et al.*, 2003; Torrents *et al.*, 2003; Zhang *et al.*, 2003; Sakai *et al.*, 2007; Table 1). One of the main differences in their criteria was sequence coverage of each processed

pseudogene against its parental gene. For example, since Ohshima *et al.* used the most stringent criteria (90%), the estimated number of processed pseudogenes is relatively small. In contrast, Torrents *et al.* did not apply any threshold. Thus, the estimated number of processed pseudogenes is relatively large. Another difference is the gene sets used in their analyses. Torrents *et al.* and Zhang *et al.* used a large set of protein sequences including computationally predicted genes, whereas Ohshima *et al.* and Sakai *et al.* used only genes having transcriptional evidence such as identified complementary deoxyribonucleic acids (cDNAs) and expressed sequence tags (ESTs) (Table 1). However, when we use the same coverage (70%) and the same data sets (including all computationally predicted genes), we obtain that the number of processed pseudogenes in the human genome is approximately 7000 as the conservative estimate.

If processed pseudogenes evolve neutrally so that their appearance and disappearance take place in a random fashion over time, an exponential decay of the distribution of the processed pseudogenes should be observed along the time. The divergence times of the processed pseudogenes since their appearance can be estimated under the assumption of molecular clock by calculating the synonymous distances ( $d_S$ ) in which nucleotide substitutions were virtually neutral for both pseudogene and functional gene pairs. In fact, the age distribution of the mouse processed pseudogenes followed the expected distribution, whereas that of the human processed pseudogenes manifested a unimodal distribution (Sakai *et al.*, 2007). The estimated date at the peak of the distribution for the human processed pseudogenes has been controversial. An estimation of the peak, as supported by some studies, is approximately 40 million years ago (MYA), which corresponds almost exactly to the time of the explosion of LINE1 retrotransposons (Ohshima *et al.*, 2003). However, another study estimated the peak to be more recent, approximately 30 MYA (Sakai *et al.*, 2007). This discrepancy may be partially due to differences in the datasets as well as the calibrated rates of neutral substitutions used. When the same

**Table 1** Gene sets used to detect processed pseudogenes and the numbers of processed pseudogenes in the human and mouse genomes

Gene set	Ohshima <i>et al.</i>	Torrents <i>et al.</i>	Zheng <i>et al.</i>	Sakai <i>et al.</i>
Genes with transcripts	Used	Used	Used	Used
Possible noncoding RNA genes with transcripts	Not used	Not used	Not used	Used
Computationally predicted genes	Not used	Used	Used	Not used
Number of predicted processed pseudogenes in the human genome	~3700	~13 500	~7800	~7300 (~6900 <sup>a</sup> )
Number of predicted processed pseudogenes in the mouse genome	NA	NA	~4500	~6200 (~6000 <sup>a</sup> )

<sup>a</sup>Number of processed pseudogenes which could be aligned with protein sequences.

Note: NA, not available.



rate of neutral evolution is used, these two estimated dates of the peak become similar to each other (Sakai *et al.*, 2007). See also: Pseudogenes: Age

When we assume an exponential decay of processed pseudogenes, the rate of emergence of processed pseudogenes in the mouse genome was estimated to be approximately 1.6% per gene per million years (Sakai *et al.*, 2007). Although the number of processed pseudogenes in the human genome declines during recent several tens of millions years, the emergence rate of processed pseudogenes was estimated to be approximately 2% per gene per million years before the decline (Sakai *et al.*, 2007). These estimates of human and mouse are markedly higher than the predicted rate of gene duplication in the human genome of 0.9%. Moreover, the processed pseudogenes appear to be, in many cases, relinquished from the genomes, with a half-life time of approximately 10 million years, whereas the half-life time of duplicates was estimated to be only 7.5 million years (Sakai *et al.*, 2007). The observation that the half-life time of mouse processed pseudogenes is shorter than that of human processed pseudogenes may be due to the fact that the mouse has the higher rates of nucleotide substitution, insertion and deletion than the human.

These results suggest that quantitative contributions of processed pseudogenes to genome diversity may be greater than the contribution made by duplicated genes, although the similarity of the tempo and of the mode of evolution between processed pseudogenes and duplicated genes suggest that both resources produce a great deal of junk sequence and produce only a few genes with biologically important functions.

## Transcriptional Activity of Processed Pseudogenes

Processed pseudogenes are, in principle, nonfunctional elements because they do not usually have appropriate functional promoters. They are derived from transcribed mRNAs by reverse transcription that is followed by reintegration of the cDNA into genomic DNA, meaning that they lost the functional promoters of their parental genes. Thus, retrotransposed copies have been recognized as 'dead on arrival' elements. However, recent studies on the transcriptional activity of human pseudogenes have demonstrated that a proportion of processed pseudogenes were actually transcribed (Harrison *et al.*, 2005; Zheng *et al.*, 2007). Indeed, according to Harrison *et al.* (2005), over 5% of the processed pseudogenes were transcribed. This estimation was based on the sequence similarities between processed pseudogenes and a comprehensive set of expressed sequences such as cDNAs, ESTs and Unigene consensus sequences.

Recently, various large-scale experimental techniques such as cap analysis of gene expression (CAGE), rapid amplification of cDNA ends (RACE) and tiling arrays have been developed to conduct comprehensive investigation of gene expression (Zheng *et al.*, 2007). These techniques have

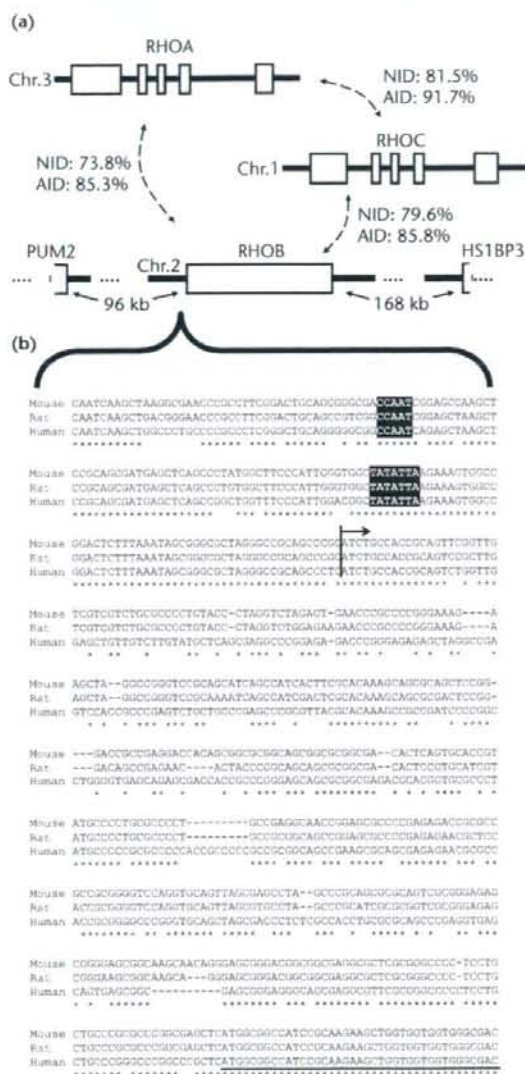
unveiled transcriptome landscapes of some eukaryotes and have been also used to validate transcriptional activity of predicted pseudogenic loci. Zheng *et al.* (2007) showed that approximately 15% of the processed pseudogenes annotated within ENCODE (ENCyclopedia Of DNA Elements) were transcriptionally active. Interestingly, they also estimated the number of transcribed nonprocessed pseudogenes to be approximately 26%, which is a much larger proportion than that of processed pseudogenes. Nonprocessed pseudogenes are derived by genomic duplication or by unequal crossing-over. Since flanking regions of their parental genes are also duplicated, nonprocessed pseudogenes often retain the functional promoters of the parental genes. Thus, it is likely that nonprocessed pseudogenes have more chances of being transcribed compared to processed pseudogenes. Then, it is of particular interest to know how processed pseudogenes are transcribed.

Transcribed processed pseudogenes tend to be observed in proximal regions or in intronic regions of functional genes, suggesting that those processed pseudogenes are occasionally transcribed using the functional promoters of preexisting genes (Harrison *et al.*, 2005). Even so, about half of the transcribed pseudogenes are located far from the surrounding functional genes. For example, the *RHOB* (*ras* homologue gene family, member B) gene, which is a well-investigated tumour suppressor gene of the Rho GTPase (guanine triphosphatase) family, was found to be a transcribed processed pseudogene (Sakai *et al.*, 2007). Along with *RHOB*, two genes, *RHOA* and *RHOC*, are members of the Rho GTPase family. *RHOA* and *RHOC* have multiple exons and the positions of their splice sites are conserved. *RHOB*, however, has only one exon, suggesting that *RHOB* was derived by retrotransposition. A recent study has shown that an ancestral gene of *RHOA* and *RHOC* was retrotransposed to a long intergenic region (Sakai *et al.*, 2007). Genes in the upstream and downstream regions of *RHOB* are located approximately 96 and 168 kb away from *RHOB*, respectively. *RHOB* is conserved between the mouse and rat and, in particular, 300 bp flanking the 5'-end are also well conserved (Figure 1). Two characteristic transcription initiation motifs such as TATA box-like sequence, TATATTA and CAAT box, CCAAT, are conserved within this region. A processed pseudogene may gain a unique functional promoter if some transposable elements carrying promoter-like sequences are integrated into the proximity of the processed pseudogene. In the case of *RHOB*, however, there are no relics of transposable elements (Figure 1). Thus, *RHOB* appears to have been integrated near promoter-like elements by chance or to have replaced a preexisting gene, so that they eventually obtained the constitutive function.

## Biological Function of Transcribed Processed Pseudogenes

Transcription has been demonstrated for some individual pseudogenes. A study by Hirotsune *et al.* (2003) investigating the functional makorin 1 pseudogene 1 (*Mkrnl-p1*)





**Figure 1** (a) Genomic structures of and sequence identities between Rho family members. NID stands for the nucleotide identity and AID for the amino acid identity. (b) Nucleotide alignment of the 5'-flanking regions of the RHOB genes obtained from human, mouse and rat. The arrow indicates the transcription start site. Two cis-elements for the transcription initiation are highlighted. Protein-coding regions are underlined. The alignment was constructed by Clustal W (ver. 1.83).

described breakthrough phenomena of the transcribed pseudogene. Firstly, *Mkrnl1-p1* is a processed pseudogene carrying frame disruptions and truncations in the coding sequence compared to that of its parental gene, *Makorin1*. Furthermore, *Mkrnl1-p1* is truncated at the 3'-end region,

whereas the 5'-end region is conserved. Since processed pseudogenes tend to be truncated at their 5'-end regions because of incomplete reverse transcription, this truncation does not seem to be due to incomplete reverse transcription of *Makorin1* mRNA, but is probably due to mutations accumulating in the initial state of *Mkrnl1-p1*. Thus, *Mkrnl1-p1* no longer possesses the ability of encoding a protein sequence. Secondly, for the mouse, *Mkrnl1-p1* functions as a noncoding RNA gene that is necessary for the stability of *Makorin1* mRNAs. The 5'-end region of *Makorin1* mRNA contains an element that facilitates the degradation of the mRNA. *Mkrnl1-p1* mRNA functions as a competitive inhibitor that prevents binding between *Makorin1* mRNA and the degrading factor, leading to the stability of the *Makorin1* mRNA. Transgenic mice mutants, in which expression of *Mkrnl1-p1* was suppressed, show severe defects in development (Hirotsumi *et al.*, 2003). This function of *Mkrnl1-p1* is still controversial, but the transcription of *Mkrnl1-p1* has been demonstrated not only in the mouse but also in rat and in human (Kaneko *et al.*, 2006).

Another transcribed pseudogene, the NOS (nitric oxide synthase) pseudogene, acts as an antisense RNA gene on its parental gene (Korneev *et al.*, 1999). NOS and its pseudogene, pseudo-NOS, are expressed in the central nervous system of snail *Lymnaea stagnalis*. The pseudo-NOS mRNA forms an RNA-RNA duplex with the NOS mRNA so that it prevents the translation of the NOS mRNA. This antisense transcription of pseudo-NOS is considered to result from duplication of the NOS gene that is followed by inversion. The pseudo-NOS gene is a nonprocessed pseudogene, meaning that it was derived by genomic duplication or by unequal crossing-over. Thus, it is possible that processed pseudogenes could also function as antisense RNA genes, although no case has yet been observed. The earlier studies have indicated a possible biological function of transcribed processed pseudogenes; they function as novel noncoding RNA genes.

The retrotransposition mechanism can generate functional genes that retain intact protein coding sequences, and such genes are called retrogenes, retroposed genes or processed genes. The *RHOB* gene, mentioned earlier, is considered as a typical retrogene. Comparison of the protein-coding sequence of *RHOB* with those of *RHOA* and *RHOC* shows that the C-terminal regions differ. In fact, *RHOB* is three amino acid residues longer than both *RHOA* and *RHOC*, and those variations in the C-terminal regions are considered to contribute to the functional divergence among members of the *RHO* family. Genome-wide investigation of retrogenes has revealed that there are about 100 retrogenes in the human and mouse genomes (Emerson *et al.*, 2004). Functionality of these retrogenes was validated by sequence and statistical analyses. Evidence of expression was derived from sequence similarities between retrogenes and a comprehensive set of known ESTs. A significantly low  $d_N/d_S$  ratio ( $< 0.5$ ) indicated that evolutionary selective constraint existed in the lineage of the retrogene. One major feature of the retrogenes is that the X-chromosome has generated a significantly higher number of retrogenes compared to the autosomes (Emerson *et al.*, 2004). This phenomenon



is considered to be related to male development. Indeed, most retrogenes derived from the X-chromosome, including some genes involved in male germline function, showed testis specific expression. Moreover, many mouse retrogenes showed testis specific expression (Sakai *et al.*, 2007). Since the parental genes of nonfunctional processed pseudogenes are distributed evenly among all chromosomes, retrotransposition events themselves should occur randomly. Therefore, natural selection appears to act on fixation and maintenance of the retrogenes on the autosomes derived from functional genes on the X-chromosome. The X-chromosome has also recruited a large number of retrogenes from the autosomes (Emerson *et al.*, 2004). Many X-linked retrogenes show significantly less expression in female-specific tissues, compared to male-specific tissues. It is, therefore, assumed that retrogenes that originated from the autosomes and that are now X-linked may confer beneficial effects if they are hemizygous in males. A comparative analysis of the autosomal retrogenes derived from the X-chromosome, as well as the retrogenes on the X-chromosome derived from autosomes, has shown that many retrotransposition events date to before the divergence of human and mouse (Emerson *et al.*, 2004). In contrast, transcribed processed pseudogenes are less conserved between human and mouse (Harrison *et al.*, 2005), suggesting that natural selection acts in both directions and that retrogenes play an important role in sex development in mammals.

## Future Prospects for the Study of Processed Pseudogenes

A recent comprehensive study of gene expression has shown that pseudogene transcriptome landscapes in the human genome are more complicated than expected (Zheng *et al.*, 2007). Few processed pseudogenes are proven to have biological functions, but the number is likely to increase because potential transcribed pseudogenes may have escaped from detection because of low levels of expression or because of tissue-specific expression. In addition, since recently emerged retrotransposed copies showed very high sequence similarities to their parental genes, it is difficult to unambiguously validate the expression of such retrotransposed copies based on sequence similarities between retrotransposed copies and various kinds of expressed sequences, such as cDNAs and ESTs. Thus, further investigation of the transcriptional activity of processed pseudogenes would be required. Note that the existence of transcriptional activity does not necessarily mean biological relevance. As mentioned earlier, about half of the transcribed processed pseudogenes resided in the proximal regions of functional genes (Harrison *et al.*, 2005). This indicates that transcripts of such processed pseudogenes were just coproducts of the adjacent functional genes and do not have any particular functions. In addition, only 5% of the transcribed processed pseudogenes were conserved between human and mouse, whereas approximately 40% of the total processed pseudogenes and

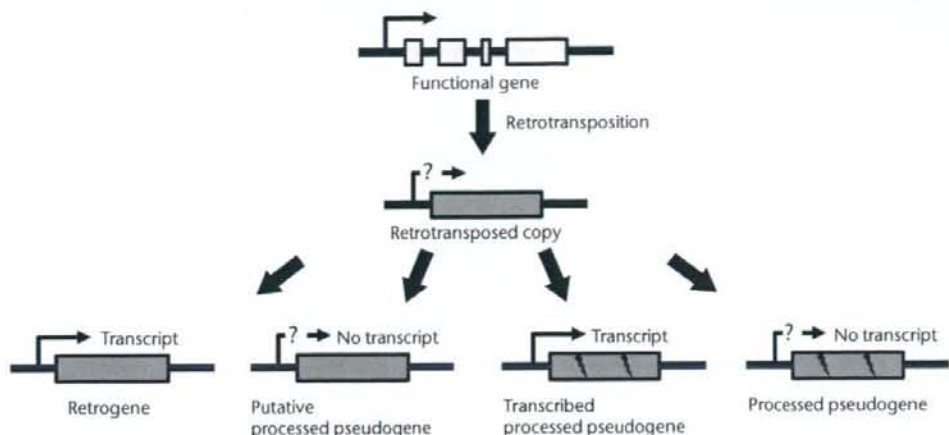
20% of the retrogenes were conserved between human and mouse (Emerson *et al.*, 2004; Harrison *et al.*, 2005). This indicates that they are only transcribed because they emerged recently and that they might eventually lose transcriptional activity after the accumulation of mutations. However, we cannot exclude the possibility that transcribed processed pseudogenes specific to a certain lineage might provide novel biological functions for the host. Thus, the next challenge for the study of transcribed processed pseudogenes is to demonstrate the biological significance of individual transcribed processed pseudogenes.

Proteome assays might be useful to show the functional relevance of transcribed pseudogenes, especially of retrotransposed copies with intact protein-coding sequences. Since transcribed mRNAs, such as *Mkrnl-1* and NOS pseudogene, might form complexes with proteins or with other mRNAs, interactome analysis will also be useful. A more straightforward approach is mutagenesis analysis, as shown for *Mkrnl-1* (Hirotsumi *et al.*, 2003). Phenotypic changes, caused by deletion mutants, could directly demonstrate the biological activities of the transcribed pseudogenes. Note however, that such experimental studies are extremely difficult to conduct for human pseudogenes. Thus, comprehensive analysis of processed pseudogenes in species other than human is of great importance.

Comparative analysis could enable us not only to facilitate experimental validation of the functionality of the transcribed processed pseudogenes but also to understand the evolutionary trajectory of the processed pseudogenes. In particular, comparative analysis of fully sequenced organisms provides sequence information of not only processed pseudogenes but also of flanking regions. Statistical measurement of selective constraint, such as calculation of the  $d_N/d_S$  ratio, might be useful to evaluate whether selective constraint is found widely in distantly related species or is found in some specific lineages. Sequence comparison of flanking regions might provide an inference of when and how processed pseudogenes were initially transcribed. As stated earlier, a pseudogene can restore an ability to be transcribed if transposable elements that carry promoter-like sequences are inserted into the 5'-flanking region of the pseudogene. If such cases are found, it means that those pseudogenes have been resurrected at some point in time.

Since the late 1970s, when the term 'pseudogene' was originally used, numerous investigations have described unusual pseudogenes, i.e. they are transcribed and have biological functions. Consequently, for such cases, the term 'pseudo' is confusing. Pseudogenes were recognized as gene-like elements that showed high sequence similarities to parental genes, but which had disrupted coding sequences. Thus, processed pseudogenes can be defined as retrotransposed copies that have accumulated mutations and consequently have lost coding potential (Figure 2). However, there is a possibility that some processed pseudogenes could gain new functions as noncoding RNA genes or possibly as completely novel protein-coding genes. A term 'exapted pseudogenes' is proposed to separate functional pseudogenes





**Figure 2** Classification of a retrotransposed copy into four categories: retrogene, putative processed pseudogene, transcribed processed pseudogene and processed pseudogene. Light grey boxes indicate exons of a functional gene and dark grey boxes indicate retrotransposed copies. The jagged arrows denote substitutions that collapse protein-coding regions of retrotransposed copies.

from dead pseudogenes, although it is quite difficult to demonstrate the nonfunctionality of pseudogenes. It is better to deal with retrogenes separately from processed pseudogenes (Figure 2). If no transcripts are found for those retrotransposed copies without any frame disruptions, it seems reasonable to classify them as putative processed pseudogenes, since only a fraction of the retrotransposed copies have been proven to be retrogenes (Figure 2). As noted earlier, the extent of selective constraint would be useful information to classify such retrotransposed copies.

Considering that the rate of emergence of retrotransposed copies is approximately 2% per gene per year in the human and mouse genomes, which is higher than that of duplicated genes, we cannot ignore the contribution of retrotransposed copies to genome evolution (Sakai *et al.*, 2007). Currently almost all of the retrotransposed copies are recognized as processed pseudogenes and a few of the processed pseudogenes have gained functions. However, the remaining processed pseudogenes might contain bona fide functional pseudogenes. Further experimental validation and evolutionary analysis will fully elucidate the biological significance of processed pseudogenes.

## References

- Emerson JJ, Kaessmann H, Betran E and Long M (2004) Extensive gene traffic on the mammalian X chromosome. *Science* **303**: 537–540.
- Harrison PM, Zheng D, Zhang Z, Carriero N and Gerstein M (2005) Transcribed processed pseudogenes in the human genome: an intermediate form of expressed retrosequence lacking protein-coding ability. *Nucleic Acids Research* **33**: 2374–2383.
- Hirotsune S, Yoshida N, Chen A *et al.* (2003) An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene. *Nature* **423**: 91–96.

- Kaneko S, Aki I, Tsuda K *et al.* (2006) Origin and evolution of processed pseudogenes that stabilize functional *Makorin1* mRNAs in mice, primates and other mammals. *Genetics* **172**: 2421–2429.
- Korneev SA, Park JH and O'Shea M (1999) Neuronal expression of neural nitric oxide synthase (nNOS) protein is suppressed by an antisense RNA transcribed from an NOS pseudogene. *Journal of Neuroscience* **19**: 7711–7720.
- Ohshima K, Hattori M, Yada T *et al.* (2003) Whole-genome screening indicates a possible burst of formation of processed pseudogenes and Alu repeats by particular L1 subfamilies in ancestral primates. *Genome Biology* **4**: R74.
- Sakai H, Koyanagi KO, Imanishi T, Itoh T and Gojobori T (2007) Frequent emergence and functional resurrection of processed pseudogenes in the human and mouse genomes. *Gene* **389**: 196–203.
- Torrents D, Suyama M, Zdobnov E and Bork P (2003) A genome-wide survey of human pseudogenes. *Genome Research* **13**: 2559–2567.
- Zhang Z, Harrison PM, Liu Y and Gerstein M (2003) Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome. *Genome Research* **13**: 2541–2558.
- Zheng D, Frankish A, Baertsch R *et al.* (2007) Pseudogenes in the ENCODE regions: consensus annotation, analysis of transcription, and evolution. *Genome Research* **17**: 839–851.

## Further Reading

- Esnault C, Maestre J and Heidmann T (2000) Human LINE retrotransposons generate processed pseudogenes. *Nature Genetics* **24**: 363–367.
- Mighell AJ, Smith NR, Robinson PA and Markham AF (2000) Vertebrate pseudogenes. *FEBS Letters* **468**: 109–114.
- Zheng D and Gerstein MB (2007) The ambiguous boundary between genes and pseudogenes: the dead rise up, or do they? *Trends in Genetics* **23**: 219–224.

患者さんの質問に答える

# 慢性肝疾患診療

改訂2版

監修 松崎晴司 / 東京医科大学 第一病棟 消化器内科 教授  
宜保行雄 / 立保消化器科・内科クリニック 院長



南山堂



## Q6

最近口内炎がよく出来ます。肝臓が悪いことと関係があるのでしょうか？

患者背景 55歳 男性. AST 55 IU/L, ALT 85 IU/L, HCV 抗体 (+)



C型肝炎ウイルスの持続感染者の方には、治りにくい口内炎が出来ることがよくあります。扁平苔癬という名前の病気で、長い間炎症が続き、自然に治ることはほとんどありません。無症状のこともあります。食事の際に口の中の粘膜が“しみる”という特徴的な症状です。扁平苔癬は、やがて口腔癌（口の中の癌）に変化することもありますので、必ず専門医での診察が必要となります。なお、C型慢性肝炎で扁平苔癬を合併している方が、インターフェロン治療を受ける場合は、扁平苔癬が悪化しインターフェロン治療が行えない場合があります。したがって、インターフェロン治療を受ける予定の方は、口の中に自覚症状がなくても専門医での診察を受けておくようにしましょう。

## AのPOINT

### 1. 肝外病変とは

肝外病変

C型肝炎ウイルス (HCV) は、肝障害のみならず種々の臓器障害の原因にもなっていることが明らかにされている。HCVは肝臓以外の疾患を引き起こし（肝外病変と呼ぶ）、その代表的な疾患としてクリオグロブリン血症、膜性増殖性糸球体腎炎、晩発性皮膚ポルフィリン症、シェーグレン症候群、悪性リンパ腫、筋炎、心筋障害、扁平苔癬、口腔癌、糖尿病、間質性肺炎、モーレン角膜潰瘍、関節リウマチ、慢性甲状腺炎などがある。HCVキャリア（HCV持続感染者）が少なくとも1つ以上の肝外病変を有する率は、retrospective studyでは74% (1,202/1,604人)、prospective studyでは38% (122/321人)と報告されている。

HCVキャリア

肝外病変についての知識は医療従事者の中でも必ずしも十分に認識されているとはいえ、また引き起こされる病態も多岐にわたるため、臓器別診療体制の傾向が強い医療施設では、各診療科間の十分な連携のもとでHCVキャリアの経過観察が必要である。

### 2. 扁平苔癬とHCV

扁平苔癬

慢性の角化異常を伴う炎症性疾患である扁平苔癬とHCV感染に関する報告は多数認められる（表6-1）。その感染率に差があるのは、各国でのHCV感染率の地域差、人種あるいは研究対象の選択の差等に影響していると考えられる。扁平苔癬の発症には、肝疾患の程度やウイルス側因子（HCV RNA量やHCV genotype）は関与しない。皮膚あるいは口腔粘膜上皮内において、HCV増殖が証明

表 6-1 扁平苔癬患者における HCV 感染率

国	研究者	調査年	対象人数 (口腔扁平苔癬 / 皮膚扁平苔癬 / 口腔皮膚 扁平苔癬のみ)	LP 陽性		
				HCV 抗体陽性率	HCV RNA 陽性率 (%)	コントロール HCV 抗体陽性率 (%)
日本	Nagao et al.	1995	45 (45/0/0)	62	60	コントロールなし
	Tanei et al.	1995	45 (28/8/9)	37.8	未検行	6.7% (3 of 45)
イタリア	Divano et al.	1992	46 (0/46/0)	32.6	未検行	コントロールなし
	Rebora et al.	1992	50 (50/0/0) (group 1 : without CLD)	4	未検行	コントロールなし
			29 (29/0/0) (group 2 : with CLD)	65	未検行	コントロールなし
			46 (46/0/0) (group 3 : with or without CLD)	24	未検行	コントロールなし
	Gandolfo et al.	1994	105 (105/0/0)	9.5	未検行	コントロールなし
スペイン	Carrozo et al.	1996	70 (70/0/0)	27.1	21.4	4.3% (3 of 70)
	Mignogna et al.	1998	263 (263/0/0)	28.8	未検行	3% (3 of 100)
	Mignogna et al.	2002	600 (600/0/0)	27.5	未検行	コントロールなし
	Bagen et al.	1994	187 (187/0/0)	15	未検行	コントロールなし
	Sanchez-Perez et al.	1998	78 (22/22/34)	20	16.7	2.4% (2 of 82)
フランス	Bagen et al.	1998	100 (100/0/0)	23	未検行	5% (5 of 100)
	Cribier et al.	1994	52 (4/48/0)	3.8	未検行	2.6% (3 of 112)
	Dupin et al.	1997	102 (102/0/0)	4.9	未検行	4.5% (14 of 306)
イギリス	Dupond et al.	1998	28 (28/0/0)	28.6	17.9	コントロールなし
	Ingafou et al.	1998	55 (55/0/0)	0	未検行	0% (0 of 110)
スコットランド	Tucker et al.	1999	45 (13/32/0)	0	0	3.1% (1 of 32)
	Roy et al.	2000	6 (6/0/0)	0	0	コントロールなし
USA	Bellman et al.	1995	30 (0/30/0)	23	16.7	4.8% (2 of 41)
	Chuang et al.	1999	22 (0/22/0)	55	未検行	25% (10 of 40) (control 1) 0.17% (255 of 149,756) (control 2)
	Basird et al.	2001	24 (0/24/0)	17	未検行	5% (1 of 20)
	Chenai-Wu et al.	2001	31 (31/0/0)	45	未検行	コントロールなし
	Eisen	2002	195 (195/0/0)	0	未検行	コントロールなし
ドイツ	Imhof et al.	1997	84 (22/82/0)	16	14	1.1% (1 of 87)
	Grote et al.	1998	24 (24/0/0)	4.2	未検行	コントロールなし
オランダ	van der Meij et al.	2000	55 (55/0/0)	0	未検行	コントロールなし
トルコ	Ilter et al.	1998	75 (0/75/0)	0	未検行	0% (0 of 75)
	Kirtak et al.	2000	73 (27/46/0)	6.8	未検行	1.36% (1 of 73)
	Erkek et al.	2001	54 (0/54/0)	12.9	9.3	3.7% (2 of 54)
ネパール	Garg et al.	2002	64 (14/35/15)	0	未検行	0% (不明)
ナイジェリア	Daramola et al.	2002	57 (0/55/2)*	15.8	未検行	25% (6 of 24) (control A) 0% (0 of 24) (control B)
ブラジル	Figueiredo et al.	2002	68 (63/0/5)	8.8	未検行	0.6% (6 of 898)

CLD : 慢性肝炎。control 1 : 乾草。control 2 : ボランティア献血者。control A : 扁平苔癬以外の皮膚疾患。control B : 正常。\*看者が Dr. Daramola に問い合わせた

(Nagao et al. J Gastroenterol Hepatol 2004 ; 19 : 1101-1113)

され (*in situ* hybridization 法あるいは RT-PCR 法), HCV 特異的な T 細胞の関与が証明されている。

HCV 関連の肝外病変に対する治療として、インターフェロンやリバビリンによる治療効果が確認されている。現在までに、効果の認められた疾患として、膜性増殖性糸球体腎炎、クリオグロブリン血症、悪性リンパ腫、晩発性皮膚ポルフィリン症、扁平苔癬、モーレン角膜潰瘍などがある。難治性の HCV 関連の膜性増殖性糸球体腎炎やクリオグロブリン血症に対しては、インターフェロンならびにリバビリン併用療法が有効な治療法であると考えられ、期待される治療であるといえる。しかしながら、必ずしも肝外病変に対するインターフェロン治療が有効な治療とはいえず、逆に肝外症状が悪化し、インターフェロン治療を中止せざるをえないケースも存在することやインターフェロンを中止すると再発することが問題視されている。特にびらん型の口腔扁平苔癬を合併する C 型慢性肝炎例に、

IFN 治療



● 症 状

インターフェロン療法を行うと皮膚や口腔病変が増悪する傾向があり、インターフェロン投与はきわめて慎重に行うことが肝要である。したがって、肝外病変に対するインターフェロン治療には注意が必要である。C型肝炎に対するインターフェロン治療前には、全身のスクリーニングと専門医への紹介を怠ってはならない。

口腔外科専門医

口腔粘膜病変に対する診断は、口腔外科専門医に紹介することを推奨する。ただし、口腔外科専門医名や口腔外科専門医施設は公表されていないのが現状であるが、2005年7月7日付けの口腔外科専門医の指定研修機関は[http://www.jsoms.or.jp/2\\_03\\_3.html](http://www.jsoms.or.jp/2_03_3.html)で閲覧可能である。なお、口腔外科専門医であってもHCVと扁平苔癬について精通しているとは限らない。

〈佐田 通夫〉

# BASIC—LIVER, PANCREAS, AND BILIARY TRACT

## Direct Cytopathic Effects of Particular Hepatitis B Virus Genotypes in Severe Combined Immunodeficiency Transgenic With Urokinase-Type Plasminogen Activator Mouse With Human Hepatocytes

MASAYA SUGIYAMA,\* YASUHIRO TANAKA,\* FUAT KURBANOV,\* ISAO MARUYAMA,† TAKASHI SHIMADA,‡ SATORU TAKAHASHI,§ TOMOYUKI SHIRAI,§ KEISUKE HINO,|| ISAO SAKAIDA,¶ and MASASHI MIZOKAMI\*<sup>¶</sup>

\*Department of Clinical Molecular Informative Medicine, and †Department of Experimental Pathology and Tumor Biology, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan; ‡PhoenixBio Co, Ltd, Higashi-Hiroshima, Japan; †Department of Basic Laboratory Sciences, ‡Department of Gastroenterology and Hepatology, Yamaguchi University Graduate School of Medicine, Yamaguchi, Japan, and †Research Center for Hepatitis & Immunology, Kohnodai Hospital, International Medical Center of Japan, Chiba, Japan

**Background & Aims:** Little is known about the direct cytopathic effect of hepatitis B virus (HBV) and its association with particular viral genotypes or genetic mutations. We investigate HBV genotype-related differences in viral replication, antigen expression, and histopathology in severe combined immunodeficiency transgenic with urokinase-type plasminogen activator mice harboring human hepatocytes. **Methods:** Mice were inoculated with wild-type of different genotype strains (3 for each HBV/A2, B1, and C2) recovered from preinfected-mice sera or patient sera. **Results:** Histologic analysis of mice infected with HBV/C2 for 22–25 weeks showed abundant ground-glass appearance of the hepatocytes and fibrosis in the humanized part of the murine liver owing to the activation of hepatic stellate cells mediated by oxidative stress through transforming growth factor- $\beta$ 1 signaling, whereas neither was observed with HBV/A2 and B1. The HBV-DNA level in sera was the highest in mice infected with HBV/C2 compared with those with HBV/A2 and HBV/B1 ( $10^9$ ,  $10^7$ , and  $10^4$  log copies/mL, respectively,  $P < .05$ ) during 6–8 weeks postinoculation. HB core-related antigen excretion had a similar trend among the genotypes, whereas secretion of HB surface antigen was more pronounced for HBV/A2 followed by HBV/C2 and much less for HBV/B1. Introduction of precore stop-codon mutation in the HBV/B1 caused a significant increase in viral replication, antigen expression, and a histopathologic picture similar to HBV/C2. **Conclusions:** By using a humanized in vivo model, we show that different HBV genotypes and even particular mutations resulted in different virologic and histopathologic outcomes of infection, indicating that particular genetic variants of HBV may be directly cytopathic in immunosuppressive conditions.

With an estimated 420 million chronic carriers, hepatitis B virus (HBV) infection is one of the most prevalent chronic viral infections of human beings. The chronic infection often leads to cirrhosis and/or hepatocellular carcinoma, which is responsible for at least 1 million deaths annually worldwide.<sup>1</sup> The precise mechanism by which chronic viral hepatitis results in hepatocellular carcinoma (HCC) is not known. However, evidence now is available concerning the direct effects of HBV in this process.<sup>2–3</sup> The important issue of a distinct impact of the various HBV genotypes on the virulence has not been addressed directly so far.<sup>4,5</sup>

Genotypes are subdivided further into subgenotypes on the basis of phylogenetic relationships.<sup>6</sup> Evidence for the influence of HBV genotypes/subgenotypes on liver diseases in acute, fulminant, and chronic infection have been reported increasingly.<sup>7–13</sup> Involvement of genetic mutations of HBV in its pathogenesis is another open question. Previous reports have indicated that mutations in basal core promoter, precore/core, envelope, and X coding regions may be associated with HCC.<sup>14</sup> The term *precore mutants* refers to HBV strains with nonsense frameshift or initiation codon mutation in the precore region that prevent translation of hepatitis B e antigen (HBeAg) precursor and are associated with an increase of viral replication via stabilization of the pregenomic encapsidation signal.<sup>15</sup> However, little is known about the histopathologic implication of the mutants. Complexity

**Abbreviations used in this paper:**  $\alpha$ -SMA,  $\alpha$ -smooth muscle actin; PCm, precore stop-codon mutation; HBcAg, antigens related to hepatitis B virus core; HSC, hepatic stellate cell; 8-OHdG, 8-hydroxydeoxyguanosine; PCR, polymerase chain reaction; ROS, reactive oxygen species; TGF- $\beta$ 1, transforming growth factor- $\beta$ 1.

© 2009 by the AGA Institute

0016-5085/09/\$36.00

doi:10.1053/j.gastro.2008.10.048