

200500145B

厚生労働科学研究費補助金

ヒトゲノム・再生医療等研究事業

サル完全長cDNAの配列決定とヒト遺伝子との比較解析および  
配列情報に基づくcDNAアレイ作製と応用に関する研究

平成15年度～17年度 総合研究報告書

主任研究者 橋本雄之

平成18(2006)年3月

# 目 次

## I. 総合研究報告

サル完全長 cDNA の配列決定とヒト遺伝子との比較解析および配列  
情報に基づく cDNA アレイ作製と応用に関する研究 ..... 1  
主任研究者: 橋本 雄之 国立感染症研究所遺伝子資源室長 (H15・16 年度)  
病原体ゲノム解析研究センター客員研究員 (H17 年度)

## II. 研究成果の刊行に関する一覧表 ..... 11

## III. 研究成果の刊行物・別刷 ..... 13

平成15-17年度

厚生科学研究費補助金（ヒトゲノム・再生医療等研究事業）総合研究報告書

サル完全長 cDNA の配列決定とヒト遺伝子との比較解析および配列  
情報に基づく cDNA アレイ作製と応用に関する研究

課題番号：H15-ゲノム-003

主任研究者：橋本 雄之 国立感染症研究所遺伝子資源室長（H15-16年度）

病原体ゲノム解析研究センター客員研究員(H17年度)

サル脳 各部の7組織及び精巣、肝臓からオリゴキャップ法により作製した完全長cDNAに富むライブラリーDNAから分離し、部分塩基配列を決定したクローンについて、相同性検索によりヒト参照遺伝子(cDNA)配列に相当するものについて、その全長配列を決定し、カニクイザル参照遺伝子セットを作製することを目標とした。これまでに9,000以上のクローンについて決定（一部、台湾アカデミアシニカ分子生物学研究所との共同研究）し、公共データベースに登録した。合計で、63,570の5'ESTおよび9,407の全長cDNAの配列を得た。9,407の全長配列のうち、7,569cDNAには5,280のヒト標準遺伝子と相同関係があった。その結果、3,067のヒトマカクorthologous遺伝子の整列配列が得られた。

1) ヒトゲノム配列で遺伝子領域として注釈されていない、タンパク質をコードする400以上のカニクイザルcDNAを見いだした。これらをもとに種特異的遺伝子の探索を行った。2) 塩基置換と構造上の変化の両方で、トランスクリプトの制御領域(5' UTR)にヒトとカニクイザルで相違が見られ、そのレベルは組織に依存していた。3) カニクイザルのトランスクリプトームで600以上Nc(non-coding)RNAを見いだした。

チンパンジー精巣ライブラリーから分離したcDNAの全長配列決定を行い、ヒトと蛋白コード領域のほぼ一致する1,077配列について、16.8% (181/1077) に5' UTRの変化が見られた。これはカニクイザルに比べて低い率であるが、カニクイザルと同様、精巣において、同じ遺伝子産物を持つ遺伝子の発現が種特異的に変化しているものと思われる。

チンパンジー脳・皮膚由来全長cDNA87個について、ヒト相同配列との比較を行ったところ、塩基置換率は全体、5' -UTR, CDS, 3' -UTR、アミノ酸でそれぞれ0.73, 0.87, 0.59, 0.93, 0.58%であった。これに塩基挿入・欠失(indel)を考慮に入れると、この種差は増加し、全体では1.36%になった。

Data Baseに登録されている約1,500個の疾患遺伝子に相補的なカニクイザルおよびチンパンジー遺伝子を収集し、ヒトや他の哺乳類の相補遺伝子と比較解析した。ヒトの病因となるアミノ酸置換をもつものがカニクイザルで8遺伝子(9アミノ酸置換)、チンパンジーで27遺伝子(29アミノ酸置換)みつかった。これらのアミノ酸置換の64%はカニクイザル、チンパンジーともに疾病型をもっており、チンパンジーからヒトに進化する過程でヒト正常型が獲得され、ヒトへの進化に寄与したことが予想される。

さらに、ヒトで肥満に関わるといわれるアドレナリン受容体β3遺伝子(ADRB3)のArg64Trp変異に関して、変化型は他の哺乳類にはみられないので、そのexon 1をチンパンジー20個体のゲノムDNAについて調べたところ、この変異に関してヘテロ接合が3個体がみいだされた。したがって、アルギニンのトリプトファンへの置換はヒトではじめて生じたのではなくヒト・チンパンジー系統で生じたものと推定された。

カニクイザル、チンパンジー、ヒトの遺伝子の転写開始部位をゲノムにマップし、プ

ロモーター部位の同定を行い、チンパンジーについては400プロモーターを、カニクイザルについては300プロモーターを同定した。また、これらのプロモーターのクローニングを行った。

全長決定クローンからヒトRefSeq対応クローン約6,000を選別し、cDNAインサートのPCR増幅を行った。また、3'末端配列を軸に各クローンにユニークな配列を選択し、当初として約1,600クロンのオリゴマー(30mer)を設計した。それをもとに、オリゴチップを作製し、その性能をみるために、カニクイザル各種組織での発現を比較して、有意な結果が得られることを確認した。

分担研究者

楠田 潤 独立法人医薬基盤研究所  
遺伝子資源研究リーダー

平井 百樹 東京女子医科大学  
統合医学インスティテュート 教授

菅野 純夫 東京大学大学院  
新領域創成科学研究科 教授

#### A. 研究目的

ヒトゲノム解析プロジェクトの進展で、ヒトゲノム全塩基配列が利用可能となり、ヒト完全長cDNAやESTの配列情報を基にヒト遺伝子約22,000が国際データベースに記載されている。また、ゲノム情報に基づく網羅的遺伝子機能解明が転写物、タンパク質レベルで進められているが、十分な展開を見せるに至っていない。従って、ゲノムの遺伝子部分の確定とその機能解明のために、分離した完全長cDNAクローンとその配列情報を研究の材料とすることは依然として必要である。

本研究ではヒトに近いモデル動物としてのカニクイザル完全長cDNAのカatalogに基づくヒトゲノム遺伝子領域解析をめざして、1)カニクイザル脳7領域、精巣および肝臓から、オリゴキャップ法により作製したライブラリー由来cDNAクローンをもとに、ヒト参照遺伝子(RefSeq)配列に相当するものを検索し、その全長塩基配列を決定し、ヒトとの配列比較解析を行う。チンパンジーは配列一致率が高すぎて意味のある差異の解析ができない例が多いので、カニクイザルを主とするが遺伝子によって適宜チンパンジーcDNAも比較

に加えて、種特異的に出現(欠失)している遺伝子や種特異的進化を示す遺伝子の探索など、遺伝子進化の異同を解析し、ヒトの進化的成り立ちを遺伝子レベルで追究する。特に、ヒトの遺伝疾患として記載のあるもの約1600についてサルcDNAとの配列比較により進化的差異を解析し、種による遺伝子機能変化と病因との関係を探る。それらをもとに、疾病遺伝子cDNAの全長配列に疾病との関連データを付加したデータベース作りを行うとともに、cDNAクローンのバンクでの運用を可能にする。2)上記で分離したカニクイザルcDNAの配列情報に基づき、2万種の遺伝子に対応するオリゴヌクレオチドマイクロアレイ作製を試み、脳部位での特定遺伝子発現の差異や肝臓での薬剤応答遺伝子発現変動を解析することなどへの利用のためバンクからcDNAアレイで供給できるようにする。

これによってこれまで不明だったヒトの疾病に関連する遺伝子の同定から機能解明に向けた研究、また、既に関係の知れたものも含めて遺伝子の機能変化とその疾病の成因との関係を解析する研究、さらに、診断、治療に結び付く研究の発展に資することを目的とする。

#### B. 研究方法

1)サル脳各部の7組織(頭頂葉、側頭葉、前頭葉、後頭葉、脳幹、小脳皮質、延髄、)及び精巣、肝臓からオリゴキャップ法により作製した完全長cDNAに富むライブラリーから分離し、部分塩基配列を決定

したクローンについて、相同性検索によりヒト参照遺伝子(cDNA)配列に相当するものについて、約1万個の全長塩基配列を決定(一部、台湾アカデミアシニカ分子生物学研究所との共同研究)し、カニクイザル参照遺伝子セットを作製する。蛋白コード領域を中心としてヒトとの配列比較を行い、遺伝子進化の異同を解析し、種の独自性との関連を追究する。特に、ヒト疾患遺伝子に対応する配列を検索し、対応するサルcDNA全長配列決定を行い、ヒトで病気を引き起こす遺伝子の変化と進化過程での変化の比較解析を進め、配列の差と病因との関係を探る。また、疾病遺伝子cDNAの全長配列に疾病との関連データを付加したデータベース作りを行うとともに、cDNAクローンを用いて機能解析ができるようにヒト疾患関連遺伝子cDNAも新たにそろえ、カニクイザルcDNAとセットで疾病解析用遺伝子バンクの確立に努める。(橋本、楠田、菅野)

2) チンパンジー組織が入手できたので、新たに精巢cDNAライブラリーを作製してcDNA数(約1万)を増やして、5'端配列を決定する。そのデータをふまえて、上記の配列比較に適当なチンパンジーcDNAを加えることにより、スプライシング変異や挿入・欠失などの構造変異、塩基(アミノ酸)置換率、非同義置換数と同義置換数の比などを調べ、種間での配列保存度を比較することにより、種特異的に出現(欠失)している遺伝子や種特異的進化を示す遺伝子の探索を行う。また、チンパンジー大脳、心筋、腎臓、肝臓などの組織よりtotal RNAを抽出し、ヒトならびにカニクイザルのマイクロアレイを用いた解析のための試料とする。(平井、橋本)

3) カニクイザルcDNAの配列情報をもとに、オリゴヌクレオチドを設計し、特異性が高く、感度のよいオリゴヌクレオチドマイクロアレイを試作する。最終的に、約2万種

の遺伝子に対応するオリゴヌクレオチドマイクロアレイ作製を試み、脳部位での特定遺伝子発現の差異や肝臓での薬剤に応答した遺伝子発現変動の解析など、その利用法を探りバンクから供給できるようにする。

(菅野、橋本)

#### (倫理面への配慮)

ヒトcDNAライブラリー原材料は医療機関に属する医師により、本人または遺族の書面による許諾を得て、採取されたもので、コード化され個人識別情報は研究者に示されない。さらに、作製に当たって複数の材料を使用するなどの手段で保護を行う形を取った。

カニクイザルは国立感染症研究所サル需給調整委員会及び動物実験委員会で審議されたもので、殺処分を行わざるをえなくなったものを霊長類センター研究員との共同研究という形で分与を受けた。チンパンジー組織は三和科学所有で自然死したのものから、同所倫理委での議を経て分与を受けた。

### C. 研究結果

1) サル脳 各部の7組織(頭頂葉、側頭葉、前頭葉、後頭葉、脳幹、小脳皮質、延髄)及び精巢、肝臓からオリゴキャップ法により作製した完全長cDNAに富むライブラリーDNAから分離し、部分塩基配列を決定したクローンについて、相同性検索によりヒト参照遺伝子(cDNA)配列に相当するものについて、その全長配列を決定し、カニクイザル参照遺伝子セットを作製することを目標とした。これまでに9,000クローン以上について決定(一部、台湾アカデミアシニカ分子生物学研究所との共同研究)し、公共データベースに登録した。

合計で、63,570の5'ESTおよび9,407の全長cDNAの配列が得られた。9,407の全長配列のうち、7,569cDNAsには5,280のヒト標準遺伝子と相同関係があった。その結果、

3,067ペアのヒトマカク orthologous 遺伝子の整列配列が得られた。そうしたヒトマカク配列比較で、2,672(87%)ペアで2つの系列の間にアミノ酸置換がみられた。同義のサイトで平均的置換率は5.7%であった。こうした結果をデータベース化し、QFbase と名付けて、インターネットに公開した。

<http://genbank.nibio.go.jp/gbank/qfbase/index.html>

(1) 全長配列のヒト遺伝子配列との比較解析の結果、脳で発現している遺伝子約2,000についてみると、他の組織の値と比べて、 $K_s$  (同義置換率) はほぼ6%と同じなのに対し、 $K_a$  (非同義置換率) は脳で1%と他組織に比べて1/2になっていることが分かった。脳では平均的に遺伝子の変化が少なくなっているとみられる。ヒトではカニクイザル、さらにチンパンジーに比べて、平均的にタンパク質の配列がより保存されていることも示された。ヒト脳の高度の機能を維持するために他の遺伝子産物との相互作用の維持の必要性などの機能的制約があつて、遺伝子(タンパク質配列)の進化速度が他に比較して遅くなっていると考えられる。

(2) ヒトゲノム配列で意味のある遺伝子領域として注釈されてない、400以上のカニクイザルのタンパク質をコードするcDNAを見いだした。これらの隠れた遺伝子の大部分は低発現レベルのためにヒトの現在のゲノムアノテーションでは特定されていないものと思われる。

(3) ヌクレオチド置換と構造上の変化の両方で、トランスクリプトの制御領域(5' UTR)にヒトとカニクイザルで相違が見られ、そのレベルは組織に依存していた。また、5' 制御領域がタンパク質コーディング領域のように選択圧の対象であることが示された。

(4) カニクイザルのトランスクリプト

ムで600以上Nc(non-coding)RNAを見いだした。これらのある程度は実験的レベルで生じたかもしれないが、ある程度のものは重要な機能を持つものと思われる。そのことは近縁種での系列比較で推論することができる。

(5) カニクイザルトランスクリプトーム(上記2)から種特異的転写物を探索し、候補7クローンを得ている。1例として、*TMEM30(CDC50)* C遺伝子についてヒトチンパンジー分岐後に転写のスプライス部位の塩基置換で引き起こされた転写の中途終了により、ヒトでは正常遺伝子産物が見られないことを見出した。

上記3)のように、ヒト相同遺伝子をもつカニクイザルcDNAについて、ORFがほぼ一致する精巣由来、脳由来ペアについて、その5' UTRを比較すると、配列がアラインできないような領域が存在し、その領域の相同配列はヒトゲノム配列には存在するが、通常のヒト転写物にみられない。そうしたエキソン領域をスタートエキソンとするなどの変化型5' UTRが精巣cDNAでは27%の割合で存在し、脳由来cDNAでは14%であった。そうしたエキソン領域をスタートエキソンとするなどの変化型5' UTRがチンパンジーの精巣転写物にも高率で見られるかどうか、チンパンジー精巣cDNAの全長配列解析を行った。

2) チンパンジー精巣よりオリゴキャップ法でcDNAライブラリーを作製し、約8,000クローンを分離した。その5' 末端配列を決定し、約6,000の引き続き解析する価値のある配列を得た。韓国KRIBBとの共同研究で、そのcDNAクローンのインサートの全長配列決定を行い、2,178配列を得た。それらをヒトゲノム配列データベース、ヒトcDNAデータベースを用いて比較解析した結果、ヒトRefSeq遺伝子に相同性のある配列1,802個が得られた。

ヒトゲノム配列、登録cDNAに当たらないcDNAは58個で、蛋白をコードしない、ヒトゲノムにエクソン様に配列する、ncRNAと見られるcDNAクローンが27個見られた。蛋白コード領域のほぼ一致する1,077配列について、UCSCヒトゲノム配列上にBlat方式でマップし、5' UTR領域エクソンに注目すると、148個(13.7%)でスタートエクソン(ex1)がヒトのトランスクリプトと異なり、31個(2.9%)でex1は一致するが、Alternative splicingの結果と思われるエクソン配列の変化が見られた。他2個にヒトゲノム配列にない配列をex1にもつものがあり、全体で、16.8%(181/1077)に5' UTRの変化が見られた。これはカニクイザルに比べて低い率であるが、精巣以外のcDNAクローンでは4.5%(6/134)であることからして、カニクイザルと同様、精巣において、同じ遺伝子産物を持つ遺伝子の発現が種特異的に変化しているものと思われる。昨年度報告したカニクイザルcDNAの5' UTRの変化したクローン169個のうち、55個がチンパンジーにもヒトオルソログとしてみられたが、うち、31個が同様に5' UTRの変化が見られたことから、チンパンジー、カニクイザル共通にヒトとは違った変化をしていることが窺える。

チンパンジーの完全長cDNAライブラリーを作製し、約1万5千クローンから得られた5' 端配列について、3,771種のヒト既知mRNAに対応する配列が得られた。脳と皮膚から得た完全長cDNA 87個を選び、その全長配列と対応するヒトmRNA配列(RefSeq)とをヒトゲノム配列上にそれぞれをマッピングした。そして、従来行っていた置換率算定に加え、exon-intron構造、転写開始点・終了点の位置の比較など、大規模な遺伝子構造における種差の有無について比較解析を行った。置換率は全体、5' -UTR, CDS, 3' -UTR, アミノ酸でそれぞれ

0.73, 0.87, 0.59, 0.93, 0.58%であった。これに塩基挿入・欠失(indel)を考慮に入れると、この種差は増加する。種間差は、indelを加算すると1.36%になり、従来の置換のみの種間差よりも大きな値となっている。

2) 遺伝子の機能と霊長類の進化との関係を知るため、Human Gene Mutation Data Baseに登録されている約1,500個の疾患遺伝子に相補的なカニクイザルおよびチンパンジー遺伝子を収集し、ヒトや他の哺乳類の相補遺伝子と比較解析した。カニクイザル全長cDNA配列および公開されているチンパンジーゲノム配列中にはそれぞれ300個及び1025個の疾患遺伝子ホモログが見いだされた。そのうち、ヒトの病因となるアミノ酸置換をもつものがカニクイザルで8遺伝子(9アミノ酸置換)、チンパンジーで27遺伝子(29アミノ酸置換)みつけた。これらのアミノ酸置換の64%はカニクイザル、チンパンジーともに疾病型となっており、チンパンジーからヒトに進化する過程でヒト正常型が獲得され、ヒトへの進化に寄与したことが予想される。

さらに、心血管疾患に関連する6遺伝子の多型について、チンパンジー20個体についてダイレクトシーケンシングにより比較解析を行った。特にアドレナリン受容体β3遺伝子(ADRB3)のArg64Trp変異に関して、変化型は他の哺乳類にはみられないので、そのexon 1をチンパンジー20個体のゲノムDNAについて調べたところ、この変異に関してヘテロ接合が3個体のみ見いだされた。したがって、アルギニンのトリプトファンへの置換はヒトではじめて生じたのではなくヒト・チンパンジー系統で生じたものと推定された。

4) これまで得られた配列データを、チンパンジー、カニクイザル等のゲノム配列にマップし、チンパンジーでは約400の、カニクイザルでは約300の遺伝子について、転

写開始部位についてマップすることができた。これらの転写開始部位を含む約1,000塩基のゲノム領域をチンパンジーおよびカニクイザルのゲノムDNAからクローン化をおこない、その活性をヒト培養細胞HEK293で検討したところ、ほとんどすべてのものにプロモーター活性が認められた。

上記カニクイザル cDNA 配列情報をもとに、オリゴヌクレオチドを設計し、ヒト参照遺伝子に対応する約2,000の遺伝子についてオリゴヌクレオチドマイクロアレイ作製を試み、脳部位での特定遺伝子発現の差異や肝臓での薬剤に応答した遺伝子発現変動の解析を可能とした。

5) ヒト標準遺伝子に相当する約8,000個の遺伝子 cDNA クローンを得ているので、この cDNA 塩基配列データベースに基づき、外部委託して1,600クローンに対応するオリゴマー搭載 DNA チップを試作した。カニクイザル脳に対して、精巣または脾臓のトータル RNA、さらに、処理と未処理肝臓 RNA を用いて、遺伝子発現解析した結果、精巣で高く発現している遺伝子群、肝臓処理で発現の上がる遺伝子群など、有意なデータが得られる結果を得た。シグナルの実測値が低いなどの問題点もあり、遺伝子発現の網羅的解析においては、発現定量性の正確さが重要な問題であるので、さらに、完全長 cDNA をプローブとして用いることによりシグナルの強度化を計り、発現量の定量性を改善するため、約6,000クローンについてPCR増幅して、チップに張り付ける方式で cDNA マイクロアレイを作製した。

#### D. 考察

カニクイザルはニホンザルと同属で、性格がより穏やかで扱いやすい為、対人薬の開発などに実験動物として用いられているが、網羅的遺伝子解析は行われていなかった。ゲノムからヒト化の過程を明らかにするには

近縁の霊長類での比較解析が不可欠であり、特に、霊長類としての特徴が明確になる旧世界ザルでの解析が重要である。カニクイザルはマウスなどよりもはるかにヒトに近縁であり、各遺伝子構造はヒトとほぼ同じであると考えられている。脳は、体の中で最もバラエティーに富んだ遺伝子が発現しており、精巣では臓器特異的に発現して新規に分離される cDNA 種が多いと期待された。さらに、カニクイザル cDNA ライブラリーを用いることにより、ヒトよりも新鮮なサンプルが入手でき、ヒトの cDNA ライブラリーでは作製の過程で失われる可能性のある希少な発現の遺伝子を発見することが期待できる。

ヒト遺伝子配列と相同性のあるカニクイザルおよびチンパンジー cDNA クローンについて、ORF がほぼ一致する精巣由来クローンの 5' UTR の変化率が高いことは精巣において、同じ遺伝子産物を持つ遺伝子の発現が種特異的に変化しているものと思われる。また、カニクイザル cDNA の 5'UTR の変化したクローン 169 個のうち、55 個がチンパンジーにもヒトオルソログとしてみられたが、うち、31 個が同様に 5'UTR の変化が見られたことからチンパンジー、カニクイザル共通にヒトとは違った変化をしている遺伝子の存在が窺える。さらに、ヒト RefSeq 遺伝子に対応するヒト cDNA14,628 の半数以上 (52%) が異なったプロモーターに制御されて転写されている事が報告されている (Genome Res. 16:55-65, 2005)。逆に7,000ほどは1つのプロモーターしか記載されていないことになる。上記のカニクイザルおよびチンパンジーのスタートエクソンがヒトの RefSeq 遺伝子のものと異なる例を示したが、カニクイザル 124 個のうち、29 個、チンパンジー 148 個のうち、33 個がそれらに該当した。ヒトではサルで使われているプロモーター配列もしくは活性がなくなった



か、あるいはそのプロモーターに由来する転写物がまだ未発見であると推測される。いずれにしても、サル の cDNA 全長配列を利用して、ヒトの遺伝子で一層の数のものが2つ以上のプロモーターを使って、転写を行っている可能性が示された。

今後の方向として、

(1) 人類に共通する悩みである疾病はヒト集団に分布する多型の一つで、実はその原因となる変異が祖先である非ヒト霊長類から受け継がれている例が多数示されるようになった。われわれはカニクイザルおよびチンパンジーよりヒト疾患に関連する遺伝子のホモログを単離、収集してヒト遺伝子と比較し、その祖先型を明らかにすることで治療方法の開発や予防に利用することをめざす。ヒトゲノム上にある22,000個の標準遺伝子のうち疾患の原因もしくは発症に関係することが現在まで明らかにされている遺伝子は約1,500個である。ヒトに最も近縁なチンパンジーと医学薬学実験に広く用いられているカニクイザルに注目し、複数の組織からオリゴキャップ法で完全長cDNAクローンを作成し、その中から疾患遺伝子ホモログを含むクローンを選別し、ヒト、チンパンジー、カニクイザルの疾患関連遺伝子のセットをつくり相互に比較することで疾患の原因となる変異の祖先型の確定を効率的に行う。

(3) cDNAクローンを固定したDNAチップを

## F. 健康危険情報

特になし

## G. 研究発表

### 1. 論文発表

1) Sakate R., Hida, M., Sugano, S., Hayasaka, I., Shimohira, N., Yanagi, S., Suto, Y., Osada, N., Hashimoto, K., Hirai, M.: Analysis of 5'-end sequences of

作製し、霊長類を使用した新薬の効力確認実験や毒性試験等に利用する。またカニクイザルES細胞から各臓器細胞への分化誘導が可能になり、カニクイザルcDNAチップで多数の遺伝子の発現パターンを調べ、目的細胞への分化を確認することを試みる。

## E. 結論

カニクイザル完全長 cDNA ライブラリー由来クローン計9万個の部分塩基配列を決定し、cDNAクローン約1万個の全長塩基配列を決定した。カニクイザルまたはチンパンジーcDNAの全長配列のヒト相同配列との比較により、遺伝子進化からヒトの進化を探ること、遺伝子の差異から疾病の成立要因を探ることの有用性を示した。ヒト、マカク、チンパンジーの配列比較により、塩基置換率の程度、5' UTRの構造変化、ヒト疾病遺伝子変位部位の祖先型で新知見を得た。カニクイザル種特異的遺伝子候補7クローンを得ている。

発現解析についてはカニクイザルcDNAチップ作製の基盤を作り基礎実験を行ったが、実例を示すまでには至らなかった。

chimpanzee cDNAs

Genome Research 13:1022-1026, 2003.

2) Iwashita, S., Osada, N., Itoh, T., Sezaki, M., Oshima, K., Hashimoto, E., Kitagawa, Y., Takahashi, I., Masui, T., Hashimoto, K., and Makalowski, W.: A transposable element-mediated gene divergence that directly produces a novel type protein including the

- endonuclease domain of RTE-1: gene duplication of bovine *bcnt*  
Mol. Biol. Evol. 20(9): 1556-1563, 2003.
- 3) Felix M. Mesak, F.M. , Osada, N., Hashimoto, K., Liu, Q.Y. and Ng, C.E. :  
Molecular cloning, genomic characterization and over-expression of a novel gene, XRR1, identified from human colorectal cancer cell HCT116<sup>clone2\_XRR</sup> and macaque testis  
BMC Genomics 4:32, 2003.
- 4) Arai M, Yokosuka O, Chiba T, Imazeki F, Kato M, Hashida J, Ueda Y, Sugano S, Hashimoto K, Saisho H, Takiguchi M, Seki N. :  
Gene expression profiling reveals the mechanism and pathophysiology of mouse liver regeneration.  
J Biol Chem. 278(32):29813-29818, 2003.
- 5) Han B, Dong Z, Liu Y, Chen Q, Hashimoto K, Zhang JT. :  
Regulation of constitutive expression of mouse PTEN by the 5'-untranslated region.  
Oncogene 22(34):5325-5337, 2003.
- 6) Suzuki, Y., Sugano, S. :  
Construction of a full-length enriched and a 5'-end enriched cDNA library using the oligo-capping method  
Methods in Molecular Biology 221:73-91, 2003.
- 7)坂手龍一、数藤由美子、平井百樹 :  
ヒトとチンパンジーの遺伝子構成の異同を探る。  
バイオサイエンスとインダストリー 61: NO.12, 11-16 (2003).
- 8) K. Nagao, N. Takenaka, M. Hirai and S. Kawamura: Coupling and decoupling of evolutionary mode between X- and Y-chromosomal red-green opsin genes in owl monkeys. Gene (2005) 352:82-91
- 9) Imanishi T, Itoh T, Suzuki Y, O'Donovan C, Fukuchi S, Koyanagi KO, Barrero RA, Tamura T, Yamaguchi-Kabata Y, Tanino M, Yura K, Miyazaki S, Ikeo K, Homma K, Kasprzyk A, Nishikawa T, Hirakawa M, Thierry-Mieg J, Thierry-Mieg D, Ashurst J, Jia L, Nakao M, Thomas MA, Mulder N, Karavidopoulou Y, Jin L, Kim S, Yasuda T, Lenhard B, Eveno E, Suzuki Y, Yamasaki C, Takeda JI, Gough C, Hilton P, Fujii Y, Sakai H, Tanaka S, Amid C, Bellgard M, Bonaldo Md M, Bono H, Bromberg SK, Brookes AJ, Bruford E, Carninci P, Chelala C, Couillault C, Souza SJ, Debily MA, Devignes MD, Dubchak I, Endo T, Estreicher A, Eyraas E, Fukami-Kobayashi K, R Gopinath G, Graudens E, Hahn Y, Han M, Han ZG, Hanada K, Hanaoka H, Harada E, Hashimoto K, Hinz U, Hirai M, Hishiki T, Hopkinson I, Imbeaud S, Inoko H, Kanapin A, Kaneko Y, Kasukawa T, Kelso J, Kersey P, Kikuno R, Kimura K, Korn B, Kuryshev V, Makalowska I, Makino T, Mano S, Mariage-Samson R, Mashima J, Matsuda H, Mewes HW, Minoshima S, Nagai K, Nagasaki H, Nagata N, Nigam R, Ogasawara O, Ohara O, Ohtsubo M, Okada N, Okido T, Oota S, Ota M, Ota T, Otsuki T, Piatier-Tonneau D, Poustka A, Ren SX, Saitou N, Sakai K, Sakamoto S, Sakate R, Schupp I, Servant F, Sherry S, Shiba R, Shimizu N, Shimoyama M, Simpson AJ, Soares B, Steward C, Suwa M, Suzuki M, Takahashi A, Tamiya G, Tanaka H, Taylor T, Terwilliger JD, Unneberg P, Veeramachaneni V, Watanabe S, Wilming L, Yasuda N, Yoo HS, Stodolsky M, Makalowski W, Go M, Nakai K, Takagi T, Kanehisa M, Sakaki Y, Quackenbush J, Okazaki Y, Hayashizaki Y,

Hide W, Chakraborty R, Nishikawa K, Sugawara H, Tateno Y, Chen Z, Oishi M, Tonellato P, Apweiler R, Okubo K, Wagner L, Wiemann S, Strausberg RL, Isogai T, Auffray C, Nomura N, Gojobori T, Sugano S.:

Integrative Annotation of 21,037 Human Genes Validated by Full-Length cDNA Clones.

PLoS Biology 2:6: 0856-0875, 2004.

10) Osada, N., Hirata, M., Tanuma, R., Kusuda, J., Hida, M., Suzuki, Y., Sugano, S., Gojobori, T., Shen, C.-K. J., Wu, C.-I., Hashimoto, K.:

Substitution rate and structural divergence of 5' UTR evolution:

Comparative analysis between human and cynomolgus monkey cDNAs.

Mol. Biol. Evol. 22(10):1976-1982(2005)

11) Yamashita R, Suzuki Y, Wakaguri H, Tsuritani K, Nakai K, Sugano S. DBTSS: DataBase of Human Transcription Start Sites, progress report 2006. Nucleic Acids Res. 34: D86-D89, 2006.

12) Kimura K, Wakamatsu A, Suzuki Y, Ota T, Nishikawa T, Yamashita R, Yamamoto J, Sekine M, Tsuritani K, Wakaguri H, Ishii S, Sugiyama T, Saito K, Isono Y, Irie R, Kushida N, Yoneyama T, Otsuka R, Kanda K, Yokoi T, Kondo H, Wagatsuma M, Murakawa K, Ishida S, Ishibashi T, Takahashi-Fujii A, Tanase T, Nagai K, Kikuchi H, Nakai K, Isogai T, Sugano S. Diversification of transcriptional modulation: Large-scale identification and characterization of putative alternative promoters of human genes. Genome Res. 16: 55-65, 2006.

## 2. 学会発表

### 国内学会

1)坂手龍一 肥田宗友 長田直樹、数籐由美子、菅野純夫、早坂郁夫、橋本雄之、平井百樹：チンパンジー遺伝子配列データベース 第19回日本霊長類学会大会、仙台、2003年6月

2)楠田潤、長田直樹、田沼玲子、平田誠、橋本雄之：カニクイザルにおける神経疾患関連遺伝子のスプライス変異体の解析 第26回日本分子生物学会年会、神戸、2003年12月

3)橋本雄之：サル完全長cDNAの分離とヒト相同遺伝子との比較解析 国立遺伝学研究所共同研究集会「霊長類の遺伝学的研究」

三島、2004年2月27-28日

4) 坂手龍一、今西 規、平井百樹、橋本雄之、五條堀 孝：

ヒトとチンパンジーのcDNA配列の比較解析から探るヒト遺伝子の進化

日本遺伝学第76回 大会、吹田、2004年9月

5) 橋本雄之、長田直樹、平田 誠、田沼玲子、楠田潤：カニクイザル精巢cDNAとヒト相同遺伝子との配列比較-28%の5' UTRに変化

第27回日本分子生物学会年会、神戸、2004年12月

6)楠田 潤、長田直樹、田沼玲子、平田 誠、平井百樹、橋本雄之

第27回日本分子生物学会年会、神戸、2004年12月

7) 楠田 潤、田沼玲子、平田 誠、高橋一朗、亀岡洋祐、橋本雄之 長田直樹

非ヒト霊長類におけるヒト疾患遺伝子ホモログの網羅的解析

第28回日本分子生物学会年会、福岡、2005年12月

8) 長田直樹、田沼玲子、平田 誠、高橋  
一朗、亀岡洋祐、橋本雄之、楠田 潤：  
カニクイザルcDNAの解析とデータベース構  
築

第28回日本分子生物学会年会、福岡、2005  
年12月

国際学会

1) Felix M. Mesak, 1 Katsuyuki  
Hashimoto, 2 Naoki Osada, 3 and Cheng E.  
Ng1 : Overexpression of a newly cloned and  
characterized gene XRR1, demonstrates a  
lack of nuclear KU-86 immunostaining.  
HUPO 2nd Annual & IUBMB XIX Joint World  
Congress, Montreal, October , 2003.

2) Y. Suto , M.H. Park, D.H. Seo, M.  
Hammond, E. Smart, R. Matsuoka, M. Hirai,  
T. Juji and Y. Ishikawa : Fiber  
fluorescence in situ hybridization  
(FISH) analysis of Rhesus (Rh) blood  
group locus.

Human Genome Variation Society  
Scientific & Annual General Meeting  
Toronto, Canada, October, 2004

3) Naoki Osada, Makoto Hirata, Jun Kusuda,  
Reiko Tanuma, Katsuyuki Hashimoto:  
Rate and pattern of 5' UTR  
evolution: comparative analysis between  
human and cynomolgus monkey.

International Workshop on Encoding  
Information in DNA sequences. Feb, 2005  
Okinawa.

4) Osada, N., Hirata, M., Tanuma, R., Kusuda  
, J., Hashimoto, K. : Cynomolgus monkey  
cDNA libraires: rate and Pattern of 5' UTR  
evolution in primets. Hman Genome  
Meeting 2005, Kyoto, Apr. 18-21, 2005.

5) R. Sakate, Y. Suto, T. Imanishi, I.  
Hayasaka, T. Gojobori, K. Hashimoto and  
M. Hirai: Comparative analysis of  
chimpanzee full-length cDNAs mapped onto  
humangenome unveils potentially large  
inter-species divergence in gene loci.  
55<sup>th</sup> Annual Meeting of American Society of  
Human Genetics 2005 (October 25-29, 2005,  
Salt Lake City, Utah, U. S. A.

H. 知的財産権の出願・登録状況  
なし

## 研究成果の刊行に関する一覧表

- 1) Arai M, Yokosuka O, Chiba T, Imazeki F, Kato M, Hashida J, Ueda Y, Sugano S,  
Hashimoto K, Saisho H, Takiguchi M, Seki N. :  
Gene expression profiling reveals the mechanism and pathophysiology of mouse liver  
regeneration.  
J Biological Chemistry 278(32):29813-29818, 2003.
- 2) Han B, Dong Z, Liu Y, Chen Q, Hashimoto K, Zhang JT. :  
Regulation of constitutive expression of mouse PTEN by the 5'-untranslated region  
Oncogene 22(34):5325-5337, 2003.
- 3) Mesak F.M., Osada, N., Hashimoto, K., Liu, Q.Y. and Ng, C.E. :  
Molecular cloning, genomic characterization and over-expression of a novel gene,  
XRRAl, identified from human colorectal cancer cell HCT116clone2\_XRR and macaque testis  
BMC Genomics 4: 32, 1-17, 2003.
- 4) Suzuki, Y., Sugano, S. :  
Construction of a full-length enriched and a 5'-end enriched cDNA library  
using the oligo-capping method  
Methods in Molecular Biology 221:73-91, 2003.
- 5) Sakate R., Hida, M., Sugano, S., Hayasaka, I., Shimohira, N., Yanagi, S., Suto, Y.,  
Osada, N., Hashimoto, K., Hirai, M. : Analysis of 5'-end sequences of chimpanzee cDNAs  
Genome Research 13:1022-1026, 2003.
- 6) 坂手龍一、数藤由美子、平井百樹 :  
ヒトとチンパンジーの遺伝子構成の異同を探る.  
バイオサイエンスとインダストリー 61: NO.12, 11-16 (2003).
- 7) Imanishi, T. et al. (Hashimoto, K., Hirai, M. and Sugano, S.) :  
Integrative Annotation of 21,037 Human Genes Validated by Full-Length cDNA Clones.  
PLoS Biology 2;6: 0856-0875, 2004.
- 8) K. Nagao, N. Takenaka, M. Hirai and S. Kawamura :  
Coupling and decoupling of evolutionary mode between X- and Y-chromosomal red-green op  
sin genes in owl monkeys.  
Gene (2005) 352-:82-91

- 9) Osada, N., Hirata, M., Tanuma, R., Kusuda, J., Hida, M., Suzuki, Y, Sugano, S., Gojobori, T., Shen, C.-K. J., Wu, C.-I., Hashimoto, K. :  
Substitution rate and structural divergence of 5' UTR evolution: Comparative analysis between human and cynomolgus monkey cDNAs.  
Mol. Biol. Evol. 22(10):1976-1982(2005)
- 10) Yamashita R, Suzuki Y, Wakaguri H, Tsuritani K, Nakai K, Sugano S. :  
DBTSS: DataBase of Human Transcription Start Sites, progress report 2006.  
Nucleic Acids Res. 34: D86-D89, 2006.
- 11) Kimura K, Wakamatsu A, Suzuki Y, Ota T, Nishikawa T, Yamashita R, Yamamoto J, Sekine M, Tsuritani K, Wakaguri H, Ishii S, Sugiyama T, Saito K, Isono Y, Irie R, Kushida N, Yoneyama T, Otsuka R, Kanda K, Yokoi T, Kondo H, Wagatsuma M, Murakawa K, Ishida S, Ishibashi T, Takahashi-Fujii A, Tanase T, Nagai K, Kikuchi H, Nakai K, Isogai T, Sugano S. :  
Diversification of transcriptional modulation: Large-scale identification and characterization of putative alternative promoters of human genes.  
Genome Research 16: 55-65, 2006.

## Gene Expression Profiling Reveals the Mechanism and Pathophysiology of Mouse Liver Regeneration\*<sup>§</sup>

Received for publication, December 11, 2002, and in revised form, May 12, 2003  
Published, JBC Papers in Press, June 2, 2003, DOI 10.1074/jbc.M212648200

Makoto Arai<sup>‡</sup>§, Osamu Yokosuka<sup>‡</sup>¶, Tetsuhiro Chiba<sup>‡</sup>, Fumio Imazeki<sup>‡</sup>, Masaki Kato<sup>§</sup>, Junya Hashida<sup>||</sup>, Youji Ueda<sup>||</sup>, Sumio Sugano<sup>\*\*</sup>, Katsuyuki Hashimoto<sup>‡‡</sup>, Hiromitsu Saisho<sup>‡</sup>, Masaki Takiguchi<sup>§§</sup>, and Naohiko Seki<sup>§</sup>

From the Departments of <sup>‡</sup>Medicine and Clinical Oncology, <sup>§</sup>Functional Genomics, and <sup>§§</sup>Biochemistry and Genetics, Graduate School of Medicine, Chiba University, Chiba 260-8670, Japan, the <sup>||</sup>Department of Life Science Products, Hitachi Software Engineering Co., Ltd., Yokohama, 230-0045, Japan, and <sup>\*\*</sup>Department of Virology, Institute of Medical Sciences, University of Tokyo, Minato-ku, Tokyo 108-8639, Japan, and the <sup>‡‡</sup>Division of Genetic Resources, National Institute of Infectious Diseases, Shinjuku-ku, Tokyo 162-8640, Japan

**Comprehensive analysis of the changes in gene expression during liver regeneration was carried out by using an in-house microarray composed of 2,304 distinct mouse liver cDNA clones. Mice were subjected to partial two-thirds hepatectomy, and changes in mRNA levels were monitored up to 48 h. Of the 2,304 genes analyzed, 496 genes showed expression levels measurable at all time points after the partial hepatectomy. 317 genes were up- or down-regulated 2-fold or more at least at one time point during liver regeneration and were classified into eight clusters based on their expression patterns. With a more stringent cut-off value of  $\pm 2$  S.D., 68 genes were listed and were classified into five clusters. In these two analyses with different clustering criteria, functionally categorized genes showed similar cluster distributions. Genes involved in protein synthesis and posttranslational processing were significantly enriched in the cluster characterized by rapid gene activation and subsequent persistence. This suggests the importance of modulating the efficiency of protein supply and/or altering the composition of protein population from the early phase of hepatocyte proliferation. Genes for two major liver functions, *i.e.* plasma protein secretion and intermediate metabolism were enriched in distinct clusters exhibiting the features of gradual gene activation and sustained repression, respectively. Therefore, these genes are differentially regulated during the regeneration, possibly leading to changes in the flow of amino acids and energy from enzyme proteins to plasma proteins in their synthesis. Thus, clustering analysis of expression patterns of functionally classified genes gave insights into mechanism and pathophysiology of liver regeneration.**

genes responsible for liver regeneration. Cytokines such as interleukin-6 and tumor necrosis factor  $\alpha$ , hormones and growth factors including insulin, norepinephrine, hepatocyte growth factor, and epidermal growth factor, and a number of transcription factors have been shown to be involved in liver regeneration (1–3). While administration of these proteins and/or overexpression of their genes induced the rapid regeneration, disruption of the genes in mice resulted in severe impairment of the regeneration (1–3). Despite the clarification of the importance of these genes for liver regeneration, the complex mechanism for the regeneration by the interplay of many factors remains to be investigated.

The expression patterns of many genes associated with liver regeneration were discussed previously (4). However, a simultaneous and comprehensive analysis of the expression profiles of these genes has been difficult because of technical limitation. Recently, DNA microarray technology has been developed and shown to be a powerful tool for analyzing the expression profiles of many genes at one time (5). By using this technique, Su *et al.* (6) revealed gene expression profiles during the priming phase of liver regeneration up to 4 h after partial hepatectomy (PHx),<sup>1</sup> and demonstrated the changes in expression of 185 genes. After the priming phase, hepatocytes proliferation starts at the periportal area of the lobular architecture and then proceeds to the perivenular areas (1, 7). Since the peaks of DNA synthesis in hepatocytes after PHx are around 24–40 h (1, 3), here we analyzed the changes in gene expression up to 48 h by using an in-house microarray harboring mouse liver cDNAs. 317 genes were shown to be up- or down-regulated during the course following PHx, and their clustering analysis revealed striking features of expression profiles of functionally classified genes, giving insights into roles and regulation of genes in liver regeneration.

Liver is unique in the ability to regenerate rapidly even in adulthood. A number of studies have been done to reveal the

### EXPERIMENTAL PROCEDURES

**Mice and mRNA Preparation**—Male C57BL/6 mice aged 5–7 weeks were obtained from CLEA Japan, Inc. (Tokyo, Japan) and were subjected to partial two-thirds PHx under ether anesthesia as described (8). Two mice were sacrificed at each time point (2, 6, 12, 24, and 48 h after PHx), and livers were resected. Liver sections removed by original hepatectomy operation of respective mice were used as a control. Total RNA was prepared from livers using TRIZOL reagent (Invitrogen), and subjected to isolation of poly(A)<sup>+</sup> RNA using the Oligotex-dT30 mRNA purification kit (TaKaRa Shuzo Co., Kyoto, Japan) according to the manufacturer's instructions.

\* This work was supported in part by grants-in-aid from the Ministry of Education, Culture, Science, Sports and Technology of Japan, DNA Chip Research Inc., and Hitachi Software Engineering Co., Ltd. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

<sup>§</sup> The on-line version of the article (available at <http://www.jbc.org>) contains supplemental Table SI.

<sup>¶</sup> To whom correspondence should be addressed: Dept. of Medicine and Clinical Oncology (K1), Graduate School of Medicine, Chiba University, Inohana 1-8-1, Chiba 260-8670, Japan. Tel.: 81-43-226-2083; Fax: 81-43-226-2088; E-mail: yokosuka@med.m.chiba-u.ac.jp.

<sup>1</sup> The abbreviations used are: PHx, partial hepatectomy; Hsp, heat-shock protein; HNF, hepatocyte nuclear factor; HFH, HNF-3/fork head homolog.

**Preparation of the cDNA Microarray**—A cDNA microarray chip, consisting of 2,304 cDNA (1,504 known genes and 800 unknown genes) was made as described previously (5, 9, 10). 2,304 unique clones were selected from about 9,000 sequenced clones in an oligo-capped cDNA library (11) of the mouse liver. PCR-amplified cDNA products were mixed with nitrocellulose in dimethyl sulfoxide and then spotted onto carbodiimide-coated glass slides using robotics SPBIO-2000 (Hitachi Software Engineering Co., Yokohama, Japan).

**Microarray Analysis**—Fluorescent cDNA probes (Cy3- or Cy5-labeled) were prepared from 2  $\mu$ g of poly(A)<sup>+</sup> RNA. Hybridization and fluorescence detection were performed essentially as described previously (9). Images were analyzed with Quant Array (GSI Lumonics, Nepean, Canada) and DNASIS Array (Hitachi Software Engineering) according to the manufacturer's instructions. The intensities of four areas between diagonal spots were used as the background for each spot. The mean and S.D. of background levels were calculated, and the genes whose intensities were less than mean plus 2 S.D. of background levels at any of the time points were excluded from further analysis. The Cy5/Cy3 ratios of all spots on the microarray were normalized by dividing them by the median. Averaged data of the two animals were subjected to statistical analysis.

**Northern Blot Analysis**—20  $\mu$ g of total RNA were electrophoresed through a 1% agarose-formaldehyde gel, transferred to a nitrocellulose membrane (Amersham Biosciences) overnight and cross-linked with irradiation. Probes were generated using Megaprime DNA labeling system (Amersham Biosciences) and [ $\alpha$ -<sup>32</sup>P]dCTP. Probes and blots were hybridized in Rapid-Hyb buffer (Amersham Biosciences). Experiments were performed twice and quantified using BAS 2000 (Fuji Photo Film Co., Tokyo, Japan).

**Statistical Analysis and Annotation of Gene Function**—The normalized Cy5/Cy3 ratios in the microarray analysis were log<sub>2</sub>-transformed and used to classify the patterns of serial changes of the gene expression. Genes whose expression levels varied at least 2-fold or by 2 S.D. at any of the time points were subjected to hierarchical clustering analysis, using the algorithm of Euclid and Ward in GeneMaths software (Applied Maths BVBA, Sint-Martens-Latem, Belgium).

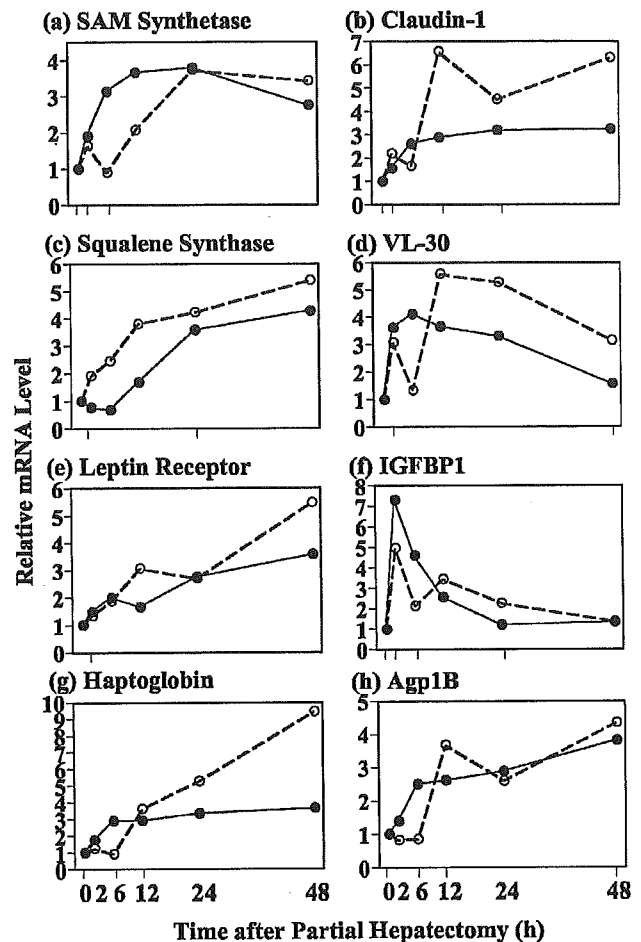
The molecular functions of the genes were assigned referring to GENE ONTOLOGY™ ([www.geneontology.org/](http://www.geneontology.org/)) and GeneCards™ ([bioinfo.weizmann.ac.il/cards/](http://bioinfo.weizmann.ac.il/cards/)).

To examine statistical significance for frequencies of genes of each functional group in each cluster, values of the other groups in the relevant cluster, and values of the other clusters in the relevant group, were each combined, and the resultant combined values were compared with the relevant value with Fisher's exact test by computing for the 2  $\times$  2 table with the statistical program StatView (SAS Institute Inc., Cary, NC).

## RESULTS AND DISCUSSION

**Application of the cDNA Microarray Analysis for Detection of Changes in Gene Expression during Liver Regeneration**—To examine the sequential changes in gene expression during liver regeneration comprehensively, we performed cDNA microarray analysis. We selected the PHx to cause the regeneration, because the initiation time of the regeneration is very clear in this model and because effects on gene expression by tissue injury or inflammation observed in other models using chemical compounds such as carbon tetrachloride or acetyl aminofluorene are minimal in the remnant intact liver in the PHx model. Poly(A)<sup>+</sup> RNAs derived from the liver at 2, 6, 12, 24, or 48 h after PHx and the control liver in duplicate were subjected to Cy3 and Cy5 labeling, respectively, coupled with cDNA synthesis. Both cDNAs were mixed in an equal amount, and hybridized with a microarray. We used an in-house microarray (9) harboring 2,304 mouse liver cDNA clones, facilitating efficient detection of changes in gene expression in the target organ. Out of the 2,304 genes analyzed, we regarded expression of 496 genes as meaningful, because their expression levels were higher than background levels in all microarray experiments. 317 genes were up- or down-regulated 2-fold or more at least at one time point during the regeneration.

To determine the validity of results obtained by the microarray analysis, 10 randomly selected genes were subjected to Northern blot analysis. While mRNAs for two genes (glutathi-



**FIG. 1. Comparison of results obtained with microarray and Northern blot analysis for changes in mRNA levels during liver regeneration.** mRNA levels at various time points after PHx relative to the level of normal liver in microarray analysis (open circles) and Northern blot analysis (closed circles) are shown for following mRNAs: a, S-adenosylmethionine (SAM) synthetase; b, claudin-1; c, squalene synthase; d, virus-like retro-element (VL-30); e, leptin receptor; f, insulin-like growth factor-binding protein 1 (IGFBP1); g, haptoglobin, and h,  $\alpha$ -1 acid glycoprotein 1B (Agp1B).

one peroxidase and *itih-4*) were under the detectable level in Northern analysis (data not shown), mRNA levels of the other eight genes (*S*-adenosylmethionine synthetase, claudin-1, squalene synthase, virus-like retro-element, leptin receptor, insulin-like growth factor-binding protein 1, haptoglobin, and  $\alpha$ -1 acid glycoprotein 1B) were almost comparable between microarray and Northern analysis (Fig. 1), generally verifying the competency of the microarray analysis for detection of changes in gene expression after PHx. However, some discrepancy was present between the results of microarray and Northern experiments. Following analyses on microarray data were done with gene numbers large enough for statistical treatment.

**Serial Changes in Expression Levels of the Genes in Relation to Their Functions**—The 317 genes with the changes in their expression levels after PHx were subjected to clustering analysis based on expression patterns and were classified into eight clusters (Table SI (see Supplemental Material) and Fig. 2A). Time course of changes in mRNA levels in each cluster can be roughly characterized as follows (Fig. 2B): cluster A, rapid increase and persistence; cluster B, transient up-regulation and gradual decrease; cluster C, gradual increase; cluster D, transient down-regulation and rebound; cluster E, rapid in-



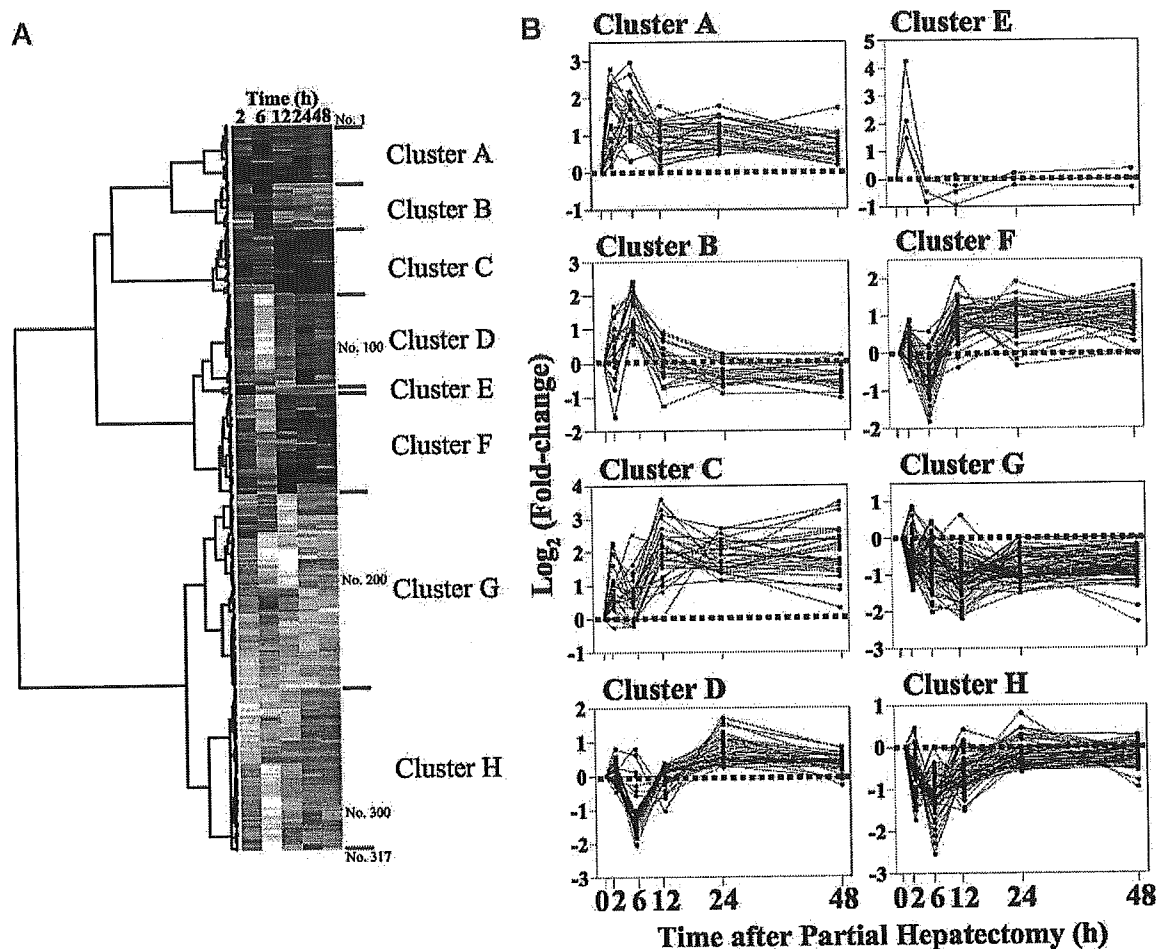


FIG. 2. Gene expression profiles during liver regeneration. A, cluster analysis of genes with expression levels that changed during liver regeneration. mRNA levels were assessed at 2, 6, 12, 24, and 48 h after PHx by the microarray analysis. A total of 317 genes whose intensities varied 2-fold or over at least at one time point during the regeneration were subjected to hierarchical clustering analysis. Time points are represented by columns, and genes in rows. Black, gray, and white represent the higher, equal, and lower mRNA level relative to that of control liver. These 317 genes were classified into eight clusters (Clusters A–H) by using GeneMaths software. B, the expression patterns of genes in each cluster. mRNA levels of each gene at various time points after PHx relative to the control mRNA level (dotted line) in normal liver are shown.

crease and rapid return; cluster F, transient down-regulation and gradual increase; cluster G, general decrease; cluster H, transient down-regulation and persistence.

We also categorized the 317 genes based on their functions, referring to GENE ONTOLOGY™ and GeneCards™, and classified them into 11 groups (Table SI in Supplemental Material and Fig. 3): cell growth and maintenance, cytoskeleton and cell membrane, metabolism, secreted proteins, signal transduction, transcription and processing, translation and processing, mitochondrion, others, and expressed sequence tags (EST). Frequencies of these functionally classified genes in each cluster are shown in Fig. 3, with statistical significance evaluated by Fisher's exact test for high and low frequencies.

Hitherto, we have used the 2-/0.5-fold changes in expression levels as cut-off values for seemingly reliable changes, which were exhibited by 17, 29, 23, 23, and 15% (amounting to 317 genes in total) of intensity-measurable 496 genes at the time points of 2, 6, 12, 24, and 48 h, respectively. The 2-/0.5-fold changes corresponded with 1.41/-1.12, 1.51/-0.75, 1.24/-0.97, 1.12/-1.34, and 1.21/-1.32 S.D. at each time point. We also tested a more stringent cut-off value of  $\pm 2$  S.D., to verify the cluster-function relationship (Fig. 4). The numbers of genes that exceeded this cut-off value occupied 3.8, 4.8, 5.6, 3.2, and 5.0% (amounting to 68 genes in total) of the 496 genes at each

time point. The 68 genes were classified into five clusters (a–e) by hierarchical clustering analysis. Their features are as follows: cluster a, rapid increase and rapid return; cluster b, rapid increase and persistence; cluster c, transient up-regulation and gradual decrease; cluster d, gradual increase; cluster e, general decrease. Again, frequencies of functionally categorized genes in these clusters are represented with their statistical significance (Fig. 4, right panels). Together with the results of Fig. 3, high frequencies were reproducibly observed in following gene distributions: genes for the category "translation and processing" in the cluster "rapid increase and persistence" (clusters A and b in Figs. 3 and 4, respectively); "secreted proteins" in "gradual increase" (clusters C and F and d); "metabolism" in "general decrease" (clusters G and e). We will discuss these characteristic gene distributions below, mainly based on the gene cluster classification in Table SI in Supplemental Material and Figs. 2 and 3.

**Activation of Genes in the Category "Translation and Processing" in the Early Stage of Liver Regeneration**—Among genes in the cluster A with properties of rapid activation and following persistence, significantly frequent were members of the category "translation and processing" (Fig. 3), which are a ribosomal protein (number 19 in Table SI in Supplemental Material), elongation factors (numbers 16 and 17 in Table SI in

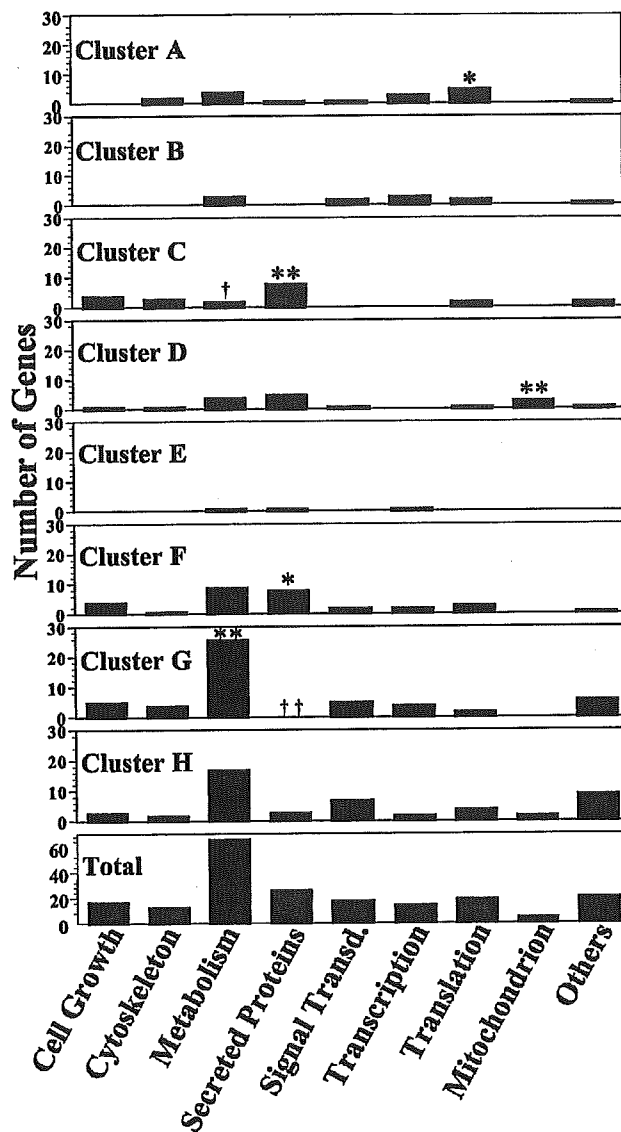


Fig. 3. Distribution of functionally categorized genes in each cluster. For classification of genes based on their molecular functions, see "Results and Discussion." Frequencies of genes of each functional category are shown. Significantly high (\*\*,  $p < 0.01$ ; \*,  $p < 0.05$ ) and low (†,  $p < 0.01$ ; †,  $p < 0.05$ ) frequencies were evaluated by Fisher's exact test.

Supplemental Material) and components of protein transport machineries (numbers 1 and 24 in Table SI in Supplemental Material). This suggests that machinery for protein synthesis, folding, and transport were prepared at the early stage of liver regeneration, presumably to synthesize and deliver proteins of altered population with modified efficiencies. Importance of adequate protein synthesis and processing for cells to pass the restriction point and enter the S phase has been repeatedly noted. When the cellular ribosome content decreases, the translation of CLN3 mRNA is inhibited, presumably leading to growth arrest (12). Involvement of translation factors in growth regulation has been most dramatically exemplified by the fact that overexpression of the mRNA cap binding protein (eIF4E) causes deregulated cell growth and malignant transformation of rodent and human cells (13). Importance of co-translational and posttranslational protein quality control in cell cycle progression has been also stressed. For example,

accumulation of misfolded proteins in endoplasmic reticulum inhibits cyclin D1 translation and results in cell cycle arrest (14). After PHx, most of the hepatocytes in the residual liver should participate in one or two proliferative events theoretically, and the peak of DNA synthesis is about at 24–40 h (1, 3). Most of hepatocytes have to pass the restriction point and enter the S phase before this period, apparently being reflected by early activation of genes for protein-supplying machineries.

**Differential Regulation of Genes for Categories "Metabolism" and "Secreted Proteins"**—Genes for two major live functions, i.e. intermediary metabolism and plasma protein secretion exhibited distinctive cluster assignment. The category "metabolism" is mainly composed of enzymes involved in hepatic intermediary metabolism of compounds such as sugars, lipids, amino acids, nucleotides, steroids, and xenobiotics. Genes of this category were significantly enriched in the cluster G, which exhibited general features of gene repression: a majority of mRNAs were under the control level even at 24 and 48 h after PHx (Fig. 2B). Seemingly, metabolic enzymes in the liver have catalytic capacities sufficient to compensate a two-thirds loss of the liver by PHx, presumably being aided also by prolongation of half-lives of the enzyme proteins during the regeneration. In fact, the degradation rate of protein decreases in the regenerating liver (15).

On the other hand, the category "secreted proteins" contains plasma proteins as a large majority and a small number of proteins involved in local liver function. Genes of this category were enriched in the clusters C and F, which are characterized by gradual increases in mRNA levels despite down-regulation, if any, in the early periods. Most of mRNAs were above the control level at 24 and 48 h following PHx (Fig. 2B). This may reflect at least two aspects in regulation of plasma protein levels. One is induction of acute phase proteins such as fibrinogen subunits (numbers 53 and 65 in Table SI in Supplemental Material),  $\alpha$ -1 acid glycoprotein 1B (number 70 in Table SI in Supplemental Material), C-reactive protein (number 137 in Table SI in Supplemental Material) and complement components (number 151 in Table SI in Supplemental Material), in response to systemic injury following PHx. The other is supplementation of constitutive plasma proteins that are under constant turnover: the two-thirds PHx seems to impose up-regulation of plasma protein synthesis on the one-third remnant liver.

Thus, syntheses of hepatic enzymes and plasma proteins seem to be regulated in the opposite direction. Amino acids and energy saved by decreased synthesis of enzyme proteins are likely to be turned toward plasma protein synthesis during liver regeneration.

**Remarkable Genes with Altered Expression Levels during Liver Regeneration**—Claudin-1 (number 66 in Table SI in Supplemental Material) in the cluster C is an integral membrane protein of tight junctions (16). Previously, a related factor claudin-3 was reported to be up-regulated during liver regeneration (17). Behavior of tight junctions during cell division is potentially very interesting. 55.11 binding protein (number 63 in Table SI in Supplemental Material, also designated TRAP2) in the cluster C is a regulatory component of the 26S proteasome and binds to the p55 tumor necrosis factor receptor (18), possibly allowing direct regulation of the proteasome by the receptor (19). Heat shock protein (Hsp) 84-1 (number 64 in Table SI in Supplemental Material, also designated Hsp84 and Hsp90) in the cluster C was previously shown to be induced during liver regeneration (20), and was recently reported to participate also in tumor necrosis factor-induced activation of I $\kappa$ B kinase (21). Hepatocyte nuclear factor (HNF)-3 $\beta$  (number 116 in Table SI in Supplemental Material, also designated Foxa2)

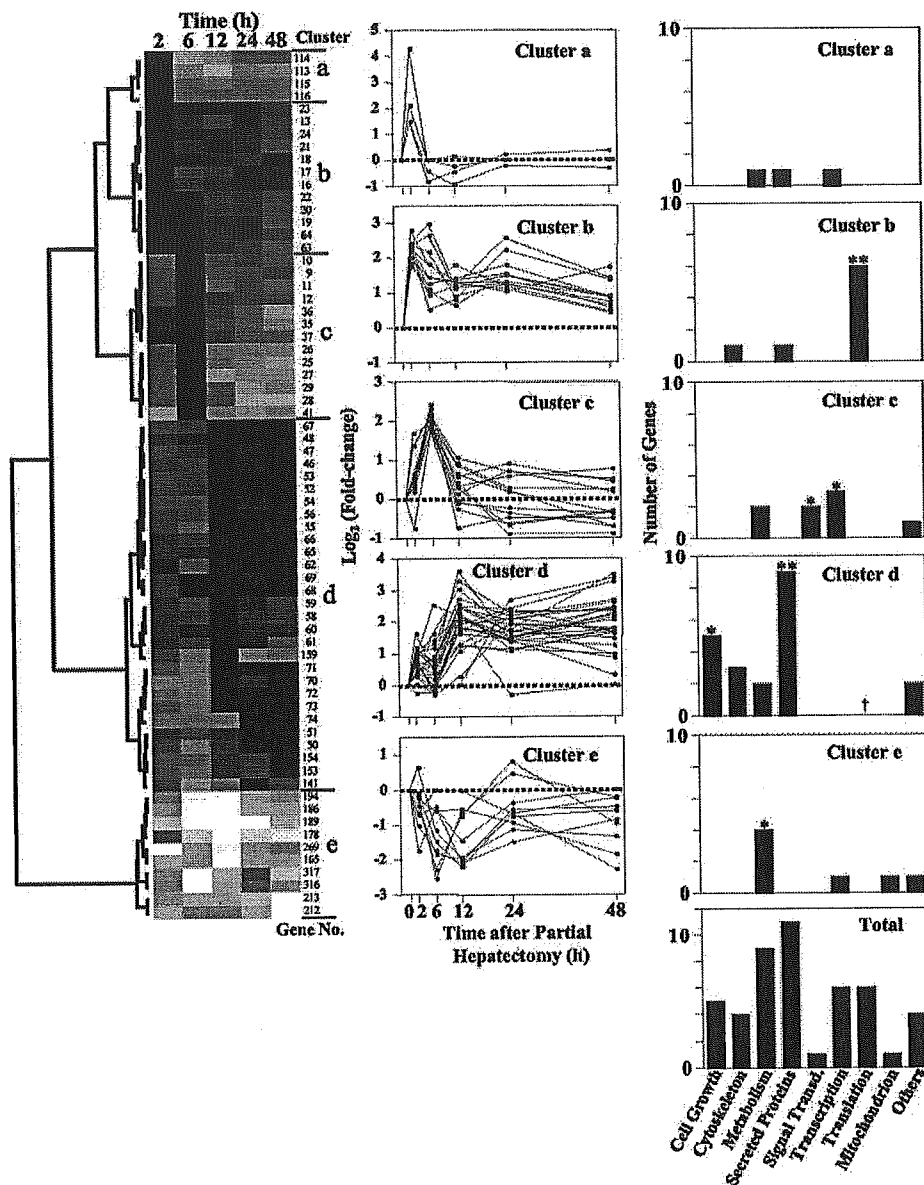


FIG. 4. Verification of the cluster distribution of functionally categorized genes with a cut-off value of  $\pm 2$  S.D. A total of 68 genes whose mRNA levels were outside  $\pm 2$  S.D. at least at one time point were subjected to the clustering analysis, and the resultant five clusters (a-e) were characterized for serial changes in mRNA levels of each gene and for frequencies of functionally categorized genes, in the same way as in Figs. 2 and 3. Significantly high (\*\*,  $p < 0.01$ ; \*,  $p < 0.05$ ) and low ( $\dagger$ ,  $p < 0.05$ ) frequencies were evaluated by Fisher's exact test.

was among the four members of the cluster E, which exhibited a striking expression profile of rapid transient augmentation. HNF-3 $\alpha$ , -3 $\beta$ , and -3 $\gamma$  are members of the winged helix/*fork head* transcription factor family and regulate a number of hepatocyte-specific genes (22, 23). Another related member HNF-3/*fork head* homolog (HFH)-11B (Foxm1) was shown to be up-regulated during the regeneration (24). In HFH-11B-overexpressing transgenic mice, the onset of hepatocyte DNA replication and mitosis during liver regeneration was accelerated (25). Thus, the winged helix/*fork head* transcription factor family including HNF-3 $\beta$  seems to play important roles in liver regeneration. Quiescin Q6 (number 153 in Table SI in Supplemental Material) in the cluster F was originally identified as a protein strongly induced when fibroblasts enter reversible quiescence, suggesting a role in growth regulation (26). Recently, this protein was shown to be a sulfhydryl oxidase that catalyzes

formation of disulfide bonds in reduced proteins (27). Quiescin Q6 was detected in the endoplasmic reticulum and Golgi apparatus as well as outside the cell and seems to be involved in oxidative folding of various proteins including secreted ones (26). Consistently, Quiescin Q6 resides in the cluster F that harbors a number of secreted proteins, the candidate substrates of the enzyme. The observation that mRNA level of Quiescin Q6 is still increasing at 48 h after PHx (Table SI in Supplemental Material) is also concordant with the notion that this enzyme might play a role in growth arrest of the regenerating liver, and it remains to be investigated.

In conclusion, the present comprehensive study with cDNA microarray analysis and following statistical analysis revealed remarkable features of changes in gene expression during liver regeneration after PHx. It is plausible that early activation of genes for protein synthesis and processing is essential for hepa-

toocytes to enter the S phase. Prolonged activation of plasma protein genes and repression of metabolic enzyme genes suggest dynamic changes in the flow of amino acids and energy in protein synthesis. A number of remarkable individual genes were also identified. These results shed light on mechanism and pathophysiology of the liver regeneration.

*Acknowledgments*—We thank M. Tagawa, K. Fukai, T. Kanda, M. Otsuka, and our colleagues for suggestions, help, and discussion.

## REFERENCES

1. Michalopoulos, G. K., and DeFrances, M. C. (1997) *Science* **276**, 60–66
2. Yamada, Y., Kirillova, I., Peschon, J. J., and Fausto, N. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94**, 1441–1446
3. Fausto, N. (2000) *J. Hepatol.* **32**, Suppl. 1, 19–31
4. Taub, R. (1996) *FASEB J.* **10**, 413–427
5. DeRisi, J. L., Iyer, V. R., and Brown, P. O. (1997) *Science* **278**, 680–686
6. Su, A. I., Guidotti, L. G., Pezacki, J. P., Chisari, F. V., and Schultz, P. G. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99**, 11181–11186
7. Assy, N., and Minuk, G. Y. (1997) *J. Hepatol.* **26**, 945–952
8. Higgins, G. M., and Andersen, R. M. (1931) *Arch. Pathol.* **12**, 186–202
9. Yoshikawa, T., Nagasugi, Y., Azuma, T., Kato, M., Sugano, S., Hashimoto, K., Masuho, Y., Muramatsu, M., and Seki, N. (2000) *Biochem. Biophys. Res. Commun.* **275**, 532–537
10. Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. (1995) *Science* **270**, 467–470
11. Suzuki, Y., Yoshitomo-Nakagawa, K., Maruyama, K., Suyama, A., and Sugano, S. (1997) *Gene (Amst.)* **200**, 149–156
12. Polymenis, M., and Schmidt, E. V. (1999) *Curr. Opin. Genet. Dev.* **9**, 76–81
13. Abid, M. R., Li, Y., Anthony, C., and De Benedetti, A. (1999) *J. Biol. Chem.* **274**, 35991–35998
14. Brewer, J. W., Hendershot, L. M., Sherr, C. J., and Diehl, J. A. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96**, 8505–8510
15. Swick, R. W., and Ip, M. M. (1974) *J. Biol. Chem.* **249**, 6836–6841
16. Furuse, M., Fujita, K., Hiiragi, T., Fujimoto, K., and Tsukita, S. (1998) *J. Cell Biol.* **141**, 1539–1550
17. Takaki, Y., Hirai, S., Manabe, N., Izumi, Y., Hirose, T., Nakaya, M., Suzuki, A., Mizuno, K., Akimoto, K., Tsukita, S., Shuin, T., and Ohno, S. (2001) *Cell Tissue Res.* **305**, 399–409
18. Boldin, M. P., Mett, I. L., and Wallach, D. (1995) *FEBS Lett.* **367**, 39–44
19. Tsurumi, C., Shimizu, Y., Saeki, M., Kato, S., Demartino, G. N., Slaughter, C. A., Fujimuro, M., Yokosawa, H., Yamasaki, M., Hendil, K. B., Toh-e, A., Tanabashi, N., and Tanaka, K. (1996) *Eur. J. Biochem.* **239**, 912–921
20. Carr, B. I., Huang, T. H., Buzin, C. H., and Itakura, K. (1986) *Cancer Res.* **46**, 5106–5111
21. Chen, G., Cao, P., and Goeddel, D. V. (2002) *Mol. Cell* **9**, 401–410
22. Lai, E., Prezioso, V. R., Smith, E., Litvin, O., Costa, R. H., and Darnell, J. E., Jr. (1990) *Genes Dev.* **4**, 1427–1436
23. Lai, E., Prezioso, V. R., Tao, W. F., Chen, W. S., and Darnell, J. E., Jr. (1991) *Genes Dev.* **5**, 416–427
24. Ye, H., Kelly, T. F., Samadani, U., Lim, L., Rubio, S., Overdier, D. G., Roebuck, K. A., and Costa, R. H. (1997) *Mol. Cell. Biol.* **17**, 1626–1641
25. Ye, H., Holterman, A. X., Yoo, K. W., Franks, R. R., and Costa, R. H. (1999) *Mol. Cell. Biol.* **19**, 8570–8580
26. Thorpe, C., Hooper, K. L., Rajc, S., Glynn, N. M., Burnside, J., Turi, G. K., and Coppock, D. L. (2002) *Arch. Biochem. Biophys.* **405**, 1–12
27. Benayoun, B., Esnard-Feve, A., Castella, S., Courty, Y., and Esnard, F. (2001) *J. Biol. Chem.* **276**, 13830–13837