

# ゲノムレベルでの遺伝子機能情報の統合化と遺伝子システム解明への応用

●松田 秀雄<sup>1)</sup> ◆坊農 秀雅<sup>2)</sup>

1) 大阪大学大学院情報科学研究科 2) 埼玉医科大学ゲノム医学研究センター

## ＜研究の目的と進め方＞

同一もしくは関連した機能を持つ遺伝子ファミリーに属する遺伝子であるかどうかを判定する配列相同性や発現プロファイル類似性の基準についての決定的な方法が未だに確立されておらず、各遺伝子の持つ配列および発現量のデータを体系的に比較・分類する手法の開発が望まれている。数多くの生物種のゲノムから遺伝子に関するデータが大量に得られつつあるが、依然として多くの遺伝子についてその機能が未知となっている。この理由としては、遺伝子の機能に関する情報が体系的に整備されておらず利用しづらい、遺伝子の集合を相互の配列相同性や発現プロファイルの類似性から機能的に関連した遺伝子のグループである遺伝子ファミリーに分類するための手法が確立されていない、などが考えられる。そこで本研究では、遺伝子の機能をより明確に特徴付ける情報の体系的抽出法の開発と、抽出された遺伝子機能情報をゲノムのレベルで体系的に整備する手法の開発を行うことで、遺伝子の相互作用を中心とした遺伝子システムの予測に結びつけることを目指し、次のようなアプローチで研究を行った。

- (1) 遺伝子の機能をより明確に特徴付ける情報の体系的抽出法の開発
- (2) 抽出された遺伝子機能情報をゲノムのレベルで相互に関連付けし、体系的に整備する手法の開発
- (3) 得られた遺伝子機能情報の体系を利用して、機能情報相互の関連性に対してデータマイニング等の統計的な解析を行うことにより、遺伝子の相互作用を中心と

した遺伝子システムの予測に結びつける。

## ＜研究開始時の研究計画＞

1. 配列データや発現データ相互の類似度をもとに、それらを分類するグラフ構造（例えば、階層的クラスタリングによる分類木）を作成する。配列データと発現データとの分類木比較、および異なる実験条件ごとの発現データの分類木の比較により、特徴的な遺伝子（例えば、野生型と変異型とを比較したときに分類木上の位置が大きく異なる遺伝子）を体系的に抽出する手法を開発する。
2. このようにして得られた特徴的な遺伝子セットを相互に、そのアノテーションデータも含めて、関連づけを行う手法を開発する。これにより、複数の実験条件で共通して特徴的な変化がある遺伝子セットの同定、およびその組織特異性や機能分類での関連性（例えば、同じ代謝反応パスウェイ上にマップされているなど）が得られると考えられる。
3. 遺伝子システムの解明に向け、配列解析から得られる情報と公になりつつあるマイクロアレイによる遺伝子発現情報を統合して、主に代謝系をターゲットにして高等真核生物において生物種間比較解析を行い、生物学的な知見を得る。その結果を元に、配列類似性や発現プロファイルの類似性構造のモデルを構築する。さらにこのモデルを遺伝子ファミリー間や生物種間で相互に比較することにより、遺伝子システムの持つ特徴的な性質を明らかにする。

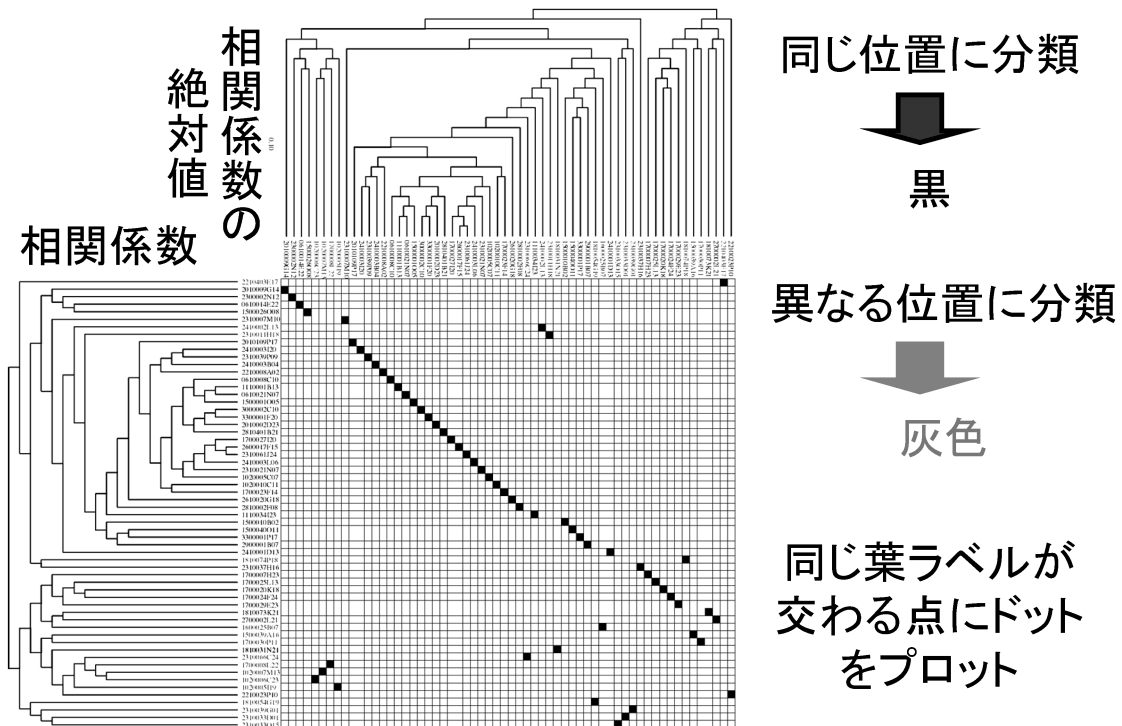


図1 異なる尺度で分類した分類木の比較による特徴的な遺伝子の検出

〈研究期間の成果〉

1.ゲノム上の多数の遺伝子をそのまま分類するのではなく、それらの遺伝子の機能の特徴付ける情報を体系的に抽出し、その情報をもとに比較・解析する方法の開発を目指し、まず、遺伝子を異なる複数の尺度（例えば、配列相同性、発現パターンの相関など）でクラスタリングにより分類し、その結果得られる分類木を比較することで、遺伝子の持つ機能的な特徴の相違を検出できる方法を開発した。例えば、図1は発現パターンの相関係数と、相関係数の絶対値の2つの尺度でクラスタリングにより分類した分類木を比較することで、尺度の違いにより分類の異なる遺伝子を検出した結果を示している。尺度が異なっても分類の変化しない遺伝子は分類木上の位置が変化しないので、この方法で比較すると図1の中央のドットプロットの中で対角線上に並ぶが、分類の変化する遺伝子は対角線から離れた位置にプロットされる。これにより、多数の遺伝子を扱う場合でも、条件の違いにより全体の中での位置付けが大きく変化する遺伝子を容易に検出することが可能となった。

2.配列データと発現データの相互の類似度を研究するためのデータセットとして、まず理化学研究所横浜研究所ゲノム科学総合研究センターにて作製されたマウスcDNAマイクロアレイ（cDNAクローン数約19000個）により得られた、マウスの種々の発生時期や臓器、計49組織での遺伝子発現データのデータベースREAD (Riken Expression Array Database)、ならびにそれらのcDNAクローンの機能アノテーションのデータベースFANTOM-DB(Functional Annotation of Mouse Database)を開発した

(成果公表2., 3., 17., 19.)。機能アノテーションがきっちりとなされたデータセットを用いることで、発現データのクラスタリングにより種々の発生時期や代謝パスウェイに参与している一群の遺伝子を明らかにし、配列データのコンピュータ解析により予測された遺伝子機能との相関を検討した（成果公表2）。とくに、指定した遺伝子とREADデータベース中のすべての遺伝子との発現パターンの類似性を検索（遺伝子発現類似性検索）するツールRINGENE(READ integrates gene expression neighbor)をREAD上に構築し、任意の組織群での発現パターンの類似性によって遺伝子を検出することを実現した（成果公表2., 20.）。

3.数万個に及ぶ大量のタンパク質アミノ酸配列を、相互の配列類似性から、グラフ理論に基づいて効率よくクラスタリングにより分類する方法を開発した（図2、成果公表4.）。この方法を使って、理研で配列決定された約6万個のマウスcDNA配列(FANTOM2配列)を翻訳したアミノ酸配列をクラスタリングして、各クラスターで類似領域を探索することにより、配列モチーフ候補を抽出した。得られたモチーフ候補の中でInterProやPfamなど既存のモチーフデータベースに登録されているものを除くことにより16個の新規モチーフを得ることができた（成果公表9.）。これらのうちの6個はhypothetical proteinにしかモチーフが含まれておらず、機能アノテーションができなかったが、最近になって別のグループの解析により、このうちの2個の機能が同定され、それぞれ、神経の軸索の誘導、RacやCdc42などのRho-family GTPase活性化に関与することがわかった。

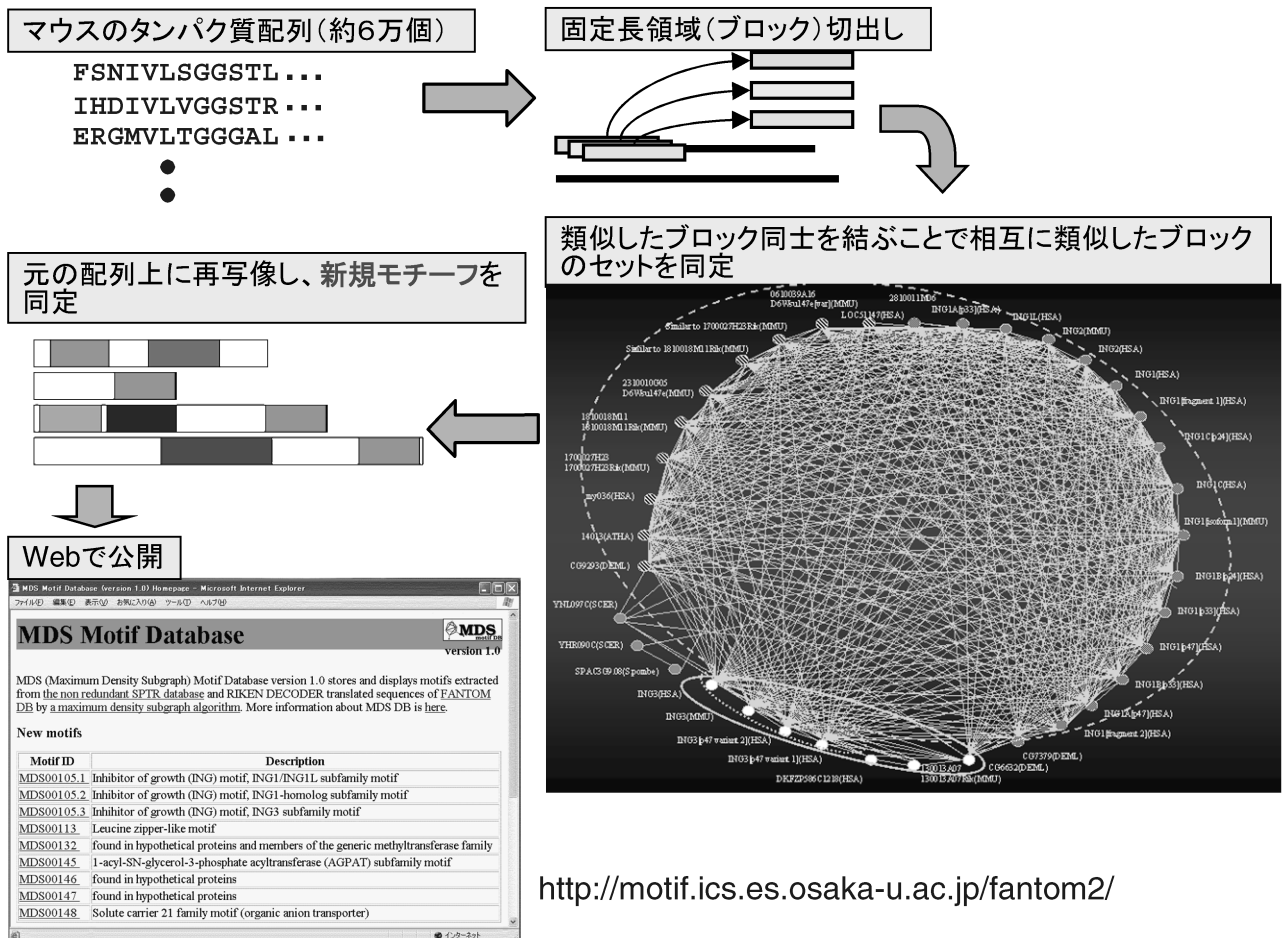


図2 大量のタンパク質配列からの新規モチーフの探索

4. 遺伝子の機能を記述する語彙として研究を始めたころに使われつつあったGene Ontologyを用い、マウス完全長cDNA プロジェクトで解析対象となった60,770個のcDNA 配列に対してコンピュータで付けられたGene Association (現在はGene Annotationと呼ばれるようになってきている) の情報から行う解析に必要なレベルで機能情報を抽出し、可視化を試みた。すなわち、全遺伝子セットを粗視的に見てその特徴を抽出する際には階層的に付けられたGene Ontologyの「根」に近い部分で代表させるGO slimを用いて機能解析に用い、酵素遺伝子をコードした遺伝子セットの解析では酵素番号に関連づけられたものだけを用いることで既存の代謝経路データベースに関連づけることを可能とした (成果公表6., 8.)。理研60kマウスcDNAマイクロアレイを用いてマウスの臓器計20組織に対して遺伝子発現プロファイルを測定したものをを用いて解析を行い、酵素遺伝子をコードした遺伝子のcDNAマイクロアレイによる発現パターンを代謝経路の順に並べるだけでも生物学的に興味深い遺伝子発現プロファイルのサブセットが得られた。なかでも図3に示したマウスのトリプトファン分解経路ではこれまで知られている経路は、上流部分を構成する酵素遺伝子の発現は肝臓や腎臓で高い一方、下流部分では心臓で特に発現が高く、それ以外の代替経路が肝臓や腎臓で働いていることが示唆された。上述のGO slimを用いてマウスの各組織 (肝臓、心臓、大脳など) において組織特異的に発現している遺伝子の特徴を解析した。肝臓や腎臓では酵素遺伝子の割合が組織特異的と判定されたものの中で高く、また心臓ではトランスポーターの割合が高くなって

おり、それぞれの臓器の役割を反映した結果が得られた (成果公表8.)。

5. 代謝反応での酵素をコードしている遺伝子において、EC番号で表現される代謝反応の類似した遺伝子が、代謝反応パスウェイ上で占める位置が近く、かつそれらの遺伝子のゲノム上での位置も近くに来るような場合に、それらの条件を満たす遺伝子セットを検出する方法を開発した (成果公表1., 7.)。この方法を開発した背景としては、生物の持つ一連の代謝反応のネットワークは最初から現在あるような複雑なものであったとは考えにくく、比較的小さなネットワークから出発して、それらをコードしている一群の遺伝子がゲノム上での重複により複製されるとともに、それぞれに起こる独立した変異により異なる機能を持つネットワークへと発展していったのではないかという仮説によっている。大腸菌などのバクテリアのパスウェイとゲノム上で類似した反応の酵素をコードしている遺伝子セットを探索したところ、トリプトファン合成とヒスチジン合成のパスウェイにおいて、類似した反応の酵素遺伝子がゲノム上での近接していることがわかった (図4)。さらに、最近になって、*Streptomyces coelicolor*のゲノムにおいて大腸菌でのtrpF遺伝子 (EC番号5.3.1.24の酵素をコードしている遺伝子) のオーソログが見つからず、この反応をhisA遺伝子の産物が酵素として触媒していることが報告されたが、これはまさに図4で示された遺伝子間の対応関係と一致しており、前述の仮説を間接的に支持しているものと考えられる。

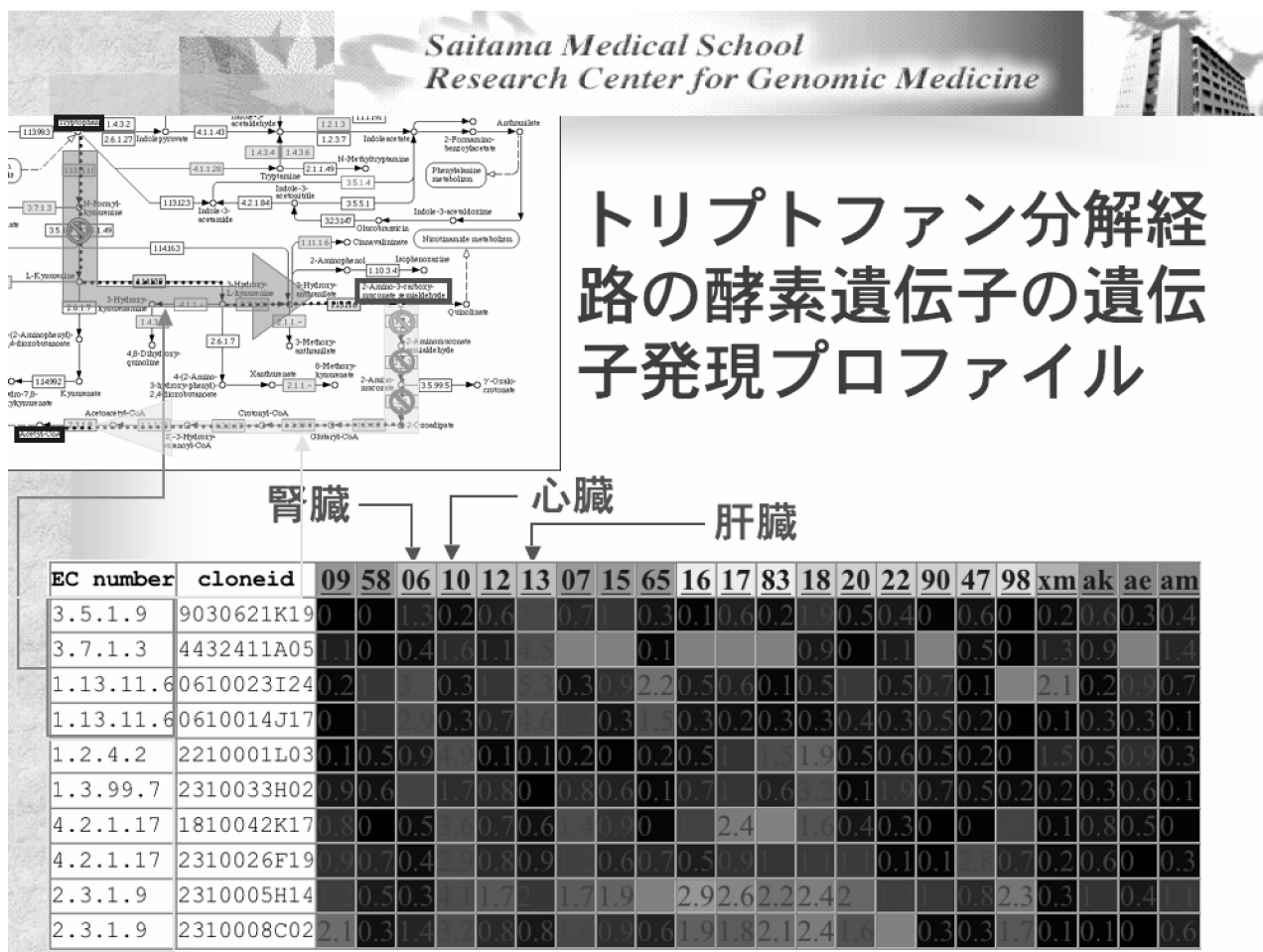
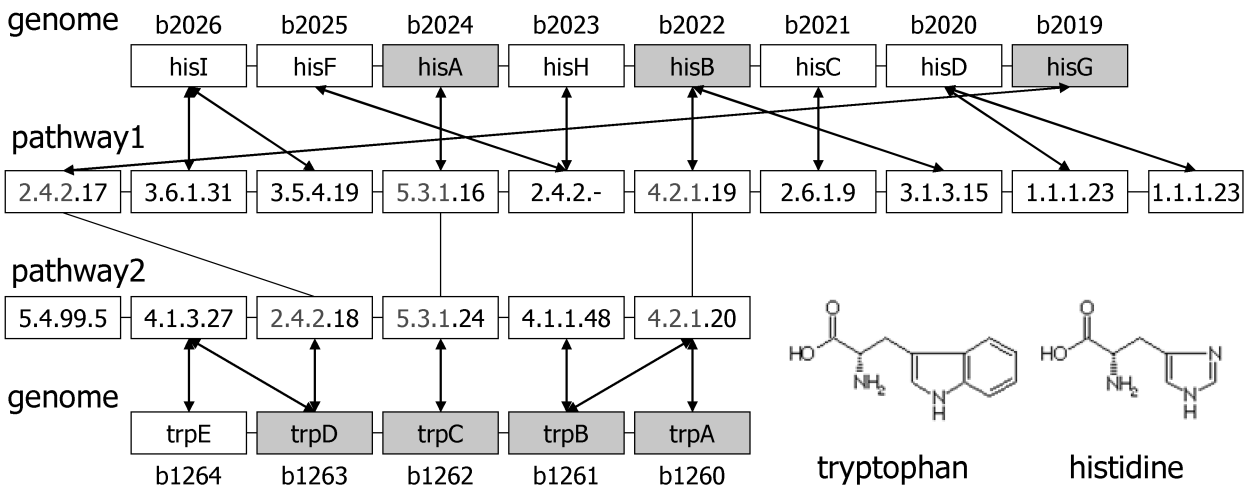


図3 再構築されたマウスのトリプトファン分解経路 (左上) とその経路上の酵素遺伝子の DNA マイクロアレイによる発現プロファイル (下)

### Pathway1: ヒスチジン生合成パスウェイ



### Pathway2: トリプトファン生合成パスウェイ

図4 類似した EC 番号を持つ代謝反応パスウェイの酵素遺伝子がゲノム上でも近接している例(大腸菌ゲノムでの探索結果)

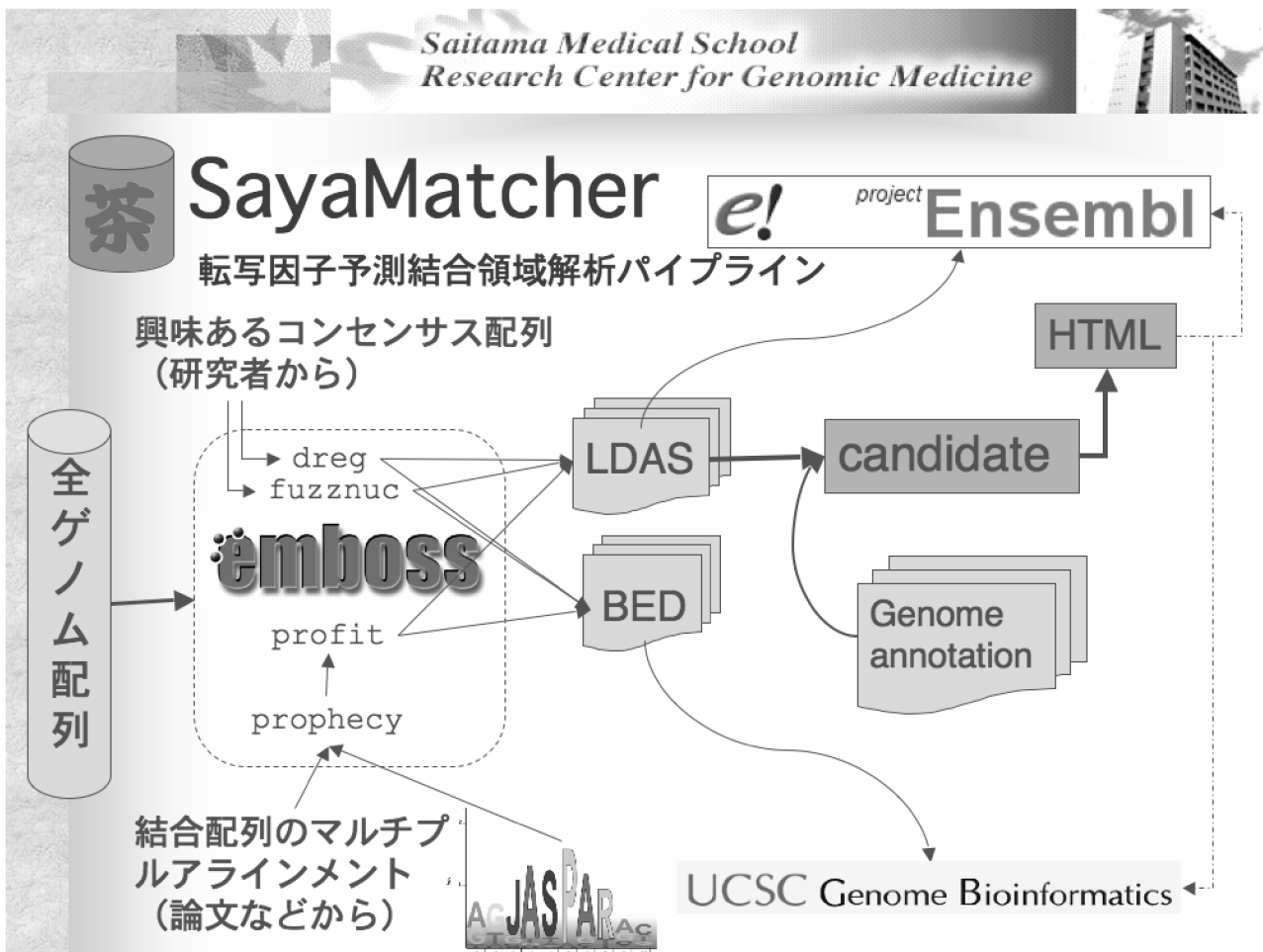


図5 転写因子予測結合領域解析パイプラインSayaMatcherの概略図

6. 解糖系に関わる遺伝子群の発現パターンについて、比較トランスクリプトーム解析の応用例の一つとして、マイクロアレイデータが多くの実験条件や組織で測定され利用可能であったマウスと酵母で比較解析を行った。酵母では一部の偽遺伝子と報告されていたものを除いてほとんどすべての遺伝子がさまざまな実験条件下で似た発現パターンをしているのに対して、マウスでは組織特異的に発現の高い遺伝子群が顕著である一方、同じ機能を果たすアイソザイムの存在とその遺伝子発現制御の重要性を示唆する結果となった(成果公表5.)。

遺伝子発現データの公共データベースとなりつつある、NCBIのGEOやEBIのArrayExpressからマイクロアレイデータを取得し、Ensemblプロジェクトで維持されているゲノムアノテーションと組み合わせることで、ローカルなデータと合わせて解析できる仕組みを構築した。これまで主に取り扱ってきたマウス以外のヒトやラットなどのデータに関しても同様に解析できるようになり、公共のサイトでは実行できないローカルなデータも含めた発現類似性検索ができ、遺伝子発現パターンの類似性に基づく遺伝子機能予測が可能となった。その例として中枢神経胚細胞腫瘍の遺伝子発現プロファイルの解析に応用し、階層的クラスタリングの結果その組織の発現プロファイルは脳由来の腫瘍にもかかわらず脳神経系ではなくむしろ胸腺や心臓、肺などに近いことが示された(成果公表16.)

とくに、これまで結合配列が知られている転写因子に関して、ゲノム中に数百から数十万存在すると期待される転写因子結合配列を複数の検索手段によりゲノムスケールで同定する解析パイプラインSayaMatcher(狭山茶)を構築し、オープンソースのフリーウェアとして研究所のウェブサイト(<http://genome.saitama-med.ac.jp/SayaMatcher/>)で公開した(成果公表15.)。これらの検索方法自体は1.正規表現によるマッチングと2.位置特異的スコア行列を用いたスコアリングといった古典的な方法であるが、一つずつの検索は質問配列が短いため通常よく用いられるBLAST検索では困難な上、複数の生物種のゲノム配列すべてに検索をかけるのは時間と

手間がかかる。常に最新のゲノム配列や探したい任意の配列パターンに対して利用可能とすることでゲノム配列解読後の生物学研究に有用なリソースとなりうる。予測した転写因子結合配列のゲノム上での位置情報はそのままでは役に立たず、既知の遺伝子などの相対的な位置関係などとともに検討することがしばしばである。そういったゲノムアノテーションとともに、予測した転写因子結合配列のゲノム上の場所を既存のゲノムブラウザー上で閲覧可能とするDAS(Distributed Annotation System)サーバーを構築し、ゲノム上に付けられた他のゲノムアノテーションと比較解析することを容易にした(成果公表15., 18.)。

アンドロゲン受容体のゲノムスケールの結合サイトの解析にSayaMatcherを実際に使用した。アンドロゲン受容体はin vitroにおいて15bpからなる回文型ホルモン応答配列(AGAACAAnnTGTTCT)に高結合性を示し、ヒト全ゲノム配列中に完全マッチするものは約180ヶ所存在すると期待される。SayaMatcherによるゲノムスケールの配列解析によってヒトゲノム中に563個の完全マッチする配列を見いだした。そのうちのX染色体上の21個の予測結合部位に関してin vivoで実際にアンドロゲン受容体と結合するかをクロマチン免疫沈降法(ChIP; Chromatin Immunoprecipitation)で検討したところ、2つがアンドロゲン受容体に4-5倍の結合性を示した。in silicoのゲノム配列解析とwet labの実験を組み合わせることにより、新規アンドロゲン応答配列群を網羅的に同定できる可能性が示唆された(成果公表12.)。

7. 遺伝子発現プロファイルからの遺伝子制御ネットワークの推定において、従来のベイジアンネットワークによる方法では、推定すべきネットワークを構成する遺伝子数に比べて、発現プロファイルに含まれる発現サンプル数が少な過ぎることから、有効な推定が困難であるということが指摘されていた。これを改善するために、遺伝子を単位とするネットワークではなく、同じような発現パターンを示す遺伝子の集合をモジュールにまとめ、モジュールを単位とするネットワークを推定するモジュ-

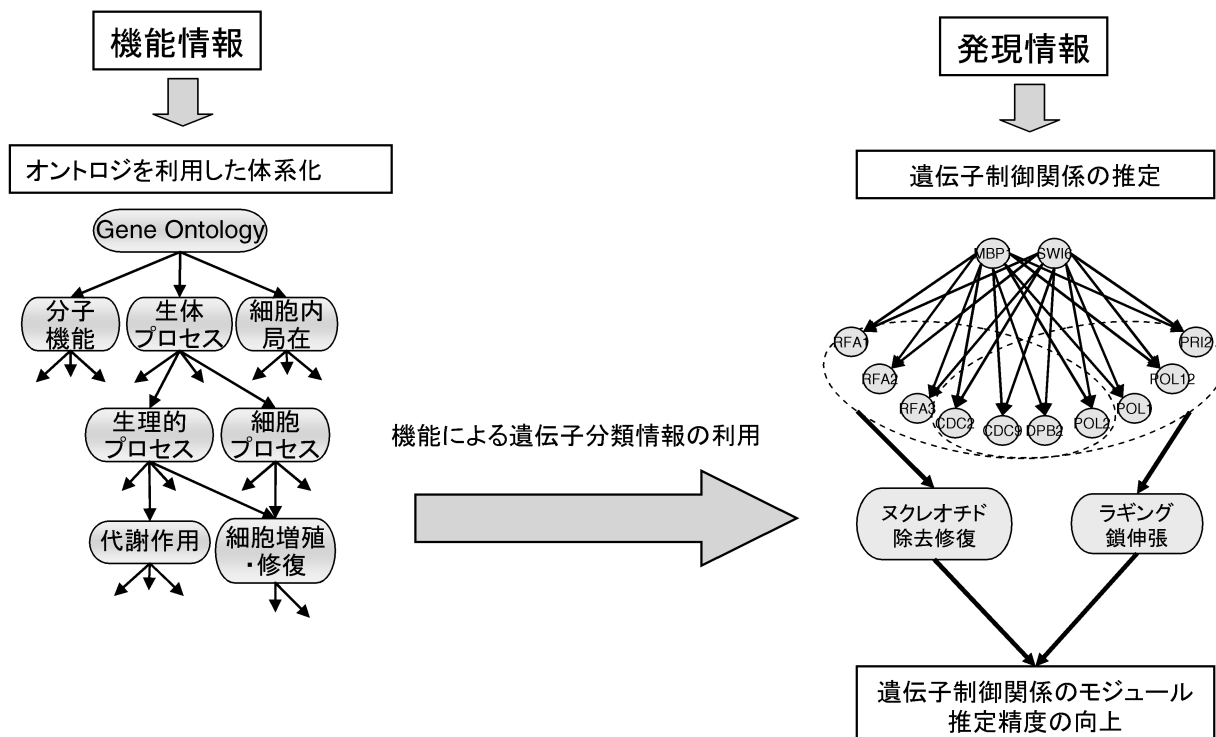


図6 遺伝子機能情報の利用による発現プロファイルからの遺伝子制御ネットワークの推定

ルネットワークという推定方法が開発された。しかし、この方法でも、本来機能的に関係のない遺伝子同士で、発現パターンが偶然類似することによりモジュールにまとめられることがあるため、必ずしもネットワーク推定の改善に結びつかないことがあった。そこで、モジュールを選択する基準として発現パターンが類似しているというだけでなく、Gene Ontologyを利用した遺伝子機能が同一か、またはGene Ontologyの機能階層上で近い所に位置する機能を持っていることも考慮した基準により、モジュールを構成する遺伝子を選択する方法を開発した(図6、成果公表13.)。

#### 〈国内外での成果の位置づけ〉

配列データや、マイクロアレイなどの発現データから一連の遺伝子群の遺伝子機能を推定する研究、とくにヒトのモデル生物としてマウスのマイクロアレイデータから疾病関連遺伝子群の遺伝子機能を推定する研究はいたるところで進められていたが、それらはいずれも配列データや発現データのどちらか一方だけを使っていたり、両方のデータを組み合わせて使う場合でも、対象とする遺伝子の持つ機能アノテーション情報が信頼性が必ずしも信頼性が高いとはいえないことが多く、品質の高い推定の妨げとなっていた。

そこで、本研究では、一貫してFANTOM会議において得られた信頼できるcDNAの機能アノテーション情報を基盤にして、それがなされた理研マウス完全長cDNAの配列およびマイクロアレイのデータを利用した。このようなデータを用いた本研究は、機能関連づけしたのちの生物学的な解釈が容易で、それらの遺伝子をもとにこれまで機能未知であった遺伝子を探すことが可能であるという特徴を持つ。このため、cDNA配列からの新規モチーフの探索では、単に局所的に配列が類似しているという領域ではなく、何らかの機能を果たすことが期待されるモチーフ候補を見つけることができた。実際、新規に見つかったモチーフ候補の中で、探索時には詳細な機能は不明なもの相互作用ドメインの可能性が高いと推定された領域がいくつかあったが、そのうちの2つが後に別のグループにより新規の相互作用ドメインであることが実験的に確かめられている。同様の意味で、本研究課題で作成したデータベースREADでは、遺伝子の発現データがFANTOMの機能アノテーション情報と結びつけられており、発現類似性検索ツールRINGENEの利便性と相まって国内外多くの研究者に利用された。

転写因子予測結合配列は、これまでデータベースが有償であることもあってゲノムブラウザ上にあらかじめ用意されたゲノムアノテーションとしては存在せず、ゲノムスケールの解析が進んでいなかった。既存のフリーウェアを組み合わせることで転写因子結合配列をゲノム中から見出し、DASを使って転写因子結合配列を解析し、遺伝子発現情報と結びつけて解析しようとする試みは世界的に見てもユニークである。負荷分散ソフトウェアを用いることでPCクラスターの計算力をフルに用いて、結果が欲しいときにすぐに計算が実行され結果が得られるようにしたことも特筆すべき特徴である。

SayaMatcherによってコンピュータで予測した転写因子結合予測領域が転写因子が結合して機能するかどうかを、アンドロゲン受容体に関してX染色体上の予測領域21箇所だけに対してではあるが実際にwet labでの実験(定量PCR)によって確かめ、約半数が機能していたことを確認したことはin silicoでのゲノム配列解析がin vivo/in vitroでの実験と連携して遺伝子群の機能解析ができたポ

ストシーケンス時代ならではの解析であり、世界的に見ても先進的である。

#### 〈達成できなかったこと、予想外の困難、その理由〉

本研究では、遺伝子配列間で共通に保存されている多数の配列モチーフ候補の中から、機能的に重要と思われる領域を探索してはいるが、配列の情報だけからそのモチーフの持つ機能を推定することは依然として極めて困難である。これは、モチーフやドメインなどの部分領域を対象とした機能情報は一部がInterProなどのデータベースにあるものの、まだ体系的に整備されているとは言いがたく、現実には個々の配列に関する文献にあたらなくてはならないため、モチーフの機能アノテーションの大きな障害になっているためである。

さらに、組織特異性のシステマティックな分類が困難であった。発現がある・ないというこれまでの実験的な証拠による基準とマイクロアレイによる発現プロファイルが一致しないために、特定の臓器特異性が予想されるほど特異的には現れず(これまで知られている組織特異的な遺伝子がそれほど組織特異的とは判断されず)、組織特異的な遺伝子として選ぶ基準の設定が難しいことが理由としてあげられる。

また、組織特異的な発現を示す遺伝子群や、発現パターンの似た遺伝子群のプロモータ領域の配列解析からは強く保存された配列モチーフを見出すことが困難であったことも予想外の困難な点であった。その理由として、一つには解析をしていた時期にはマウスゲノム配列が公開された直後であったため、そのデータの信頼性の低さということがあげられる。また別の理由として挙げられるのが、完全長cDNA配列を用いていても5'側には予想より多くの時期特異性や組織特異性があることがバリエーションが存在し、どこが正確に転写開始点であるか一意に決めるのが困難であったことがある。これに関してはその後FANTOM3のCAGEやditagを用いた実験によって、実際に転写開始点は多くの遺伝子で多数あることが示されている。

転写因子結合配列の解析において、当初ターゲットとした代謝系での解析から生物学的な知見を得ることは依然として困難で達成できなかった。その理由の一つには、転写因子結合配列を解析するコンピュータを用いた実験系(SayaMatcher)の確立に時間がかかったことがあげられる。ただ、計算そのものは早く終わることができるのでこれはそれほど大きな問題とはならなかったが、別の大きな理由として、転写因子結合配列には疑陽性が多く得られた結果をそのまま解釈することが困難であることがあげられる。また、比較すべきwet labでの実験によるゲノムスケールの転写因子結合領域のデータ(ChIP on chipのデータ)が当初の予想ほど多数公開されず、比較手段を検討している段階で年限となってしまった。

#### 〈今後の課題〉

本研究では、遺伝子の持つ配列・発現・機能データを有機的に結びつけることで種々の特徴的な解析手法とデータベースを開発し、最終的にそれらの手法全体を統合して遺伝子システムの解明を目指してきた(図7)。しかし、それぞれの手法はある程度までは相互に関連付けて利用できるようになったものの、全体が統合されるまでには至っておらず、今後の課題として残されている。また、個別の解析手法に関しては、まず発現パターンのクラスタリング解析から組織特異的な転写ユニットの解析を徹底的に進め、実験的検証を行う。さらにこれまで

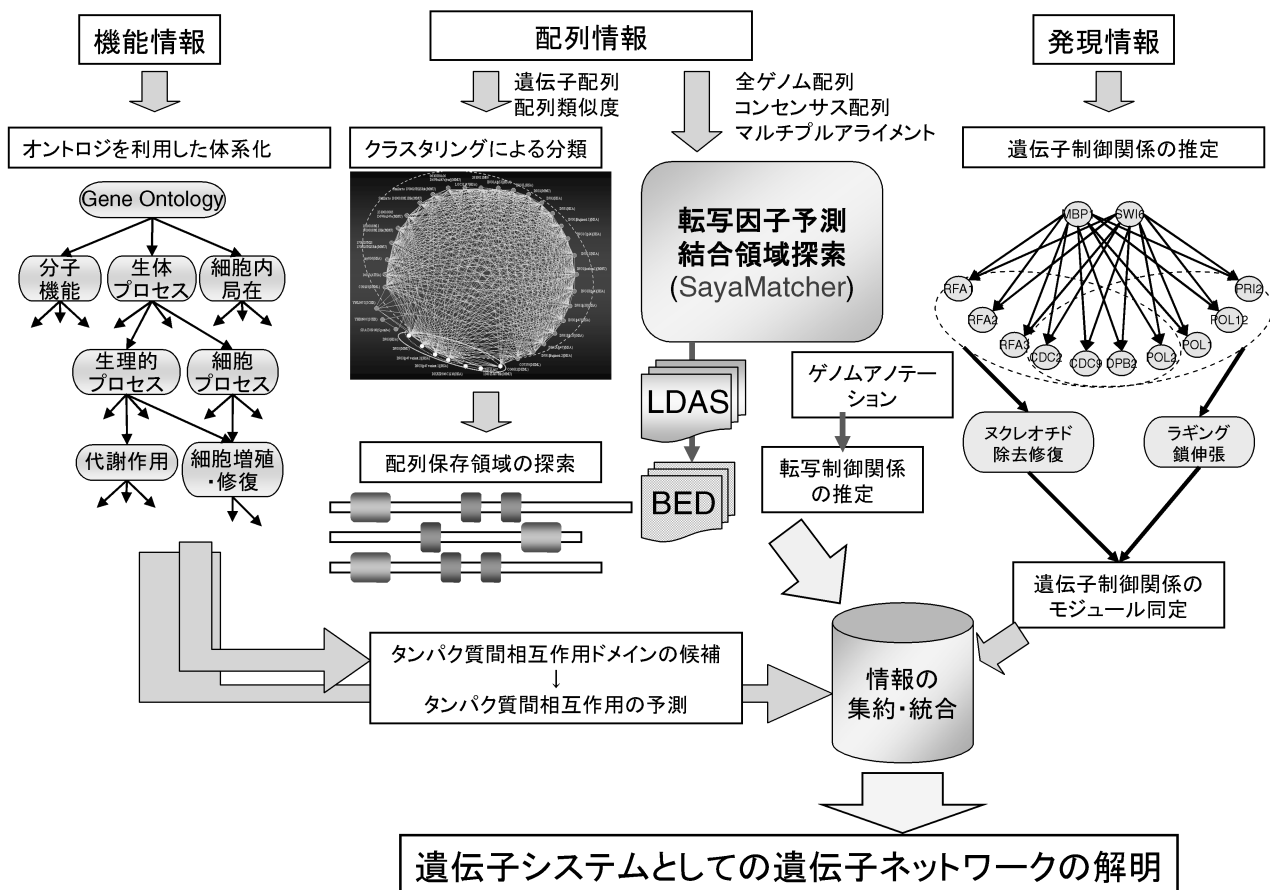


図7 本研究全体の目標イメージ

解析してきたマウス以外の複数の生物種でのマイクロアレイによる発現プロファイルも利用することでそれらのデータの比較解析を行い、すべての生物種間で保存された遺伝子ネットワークを解析する方法論の確立を目指し、実際の問題に適用していくことが今後の課題としてあげられる。

そのためには、遺伝子制御領域に存在している塩基配列モチーフの解析が必要不可欠である。これまで知られている転写因子の結合する塩基配列モチーフと発現パターンのクラスタリング解析によって得られる時期、細胞特異的な塩基配列モチーフを比較解析する方法論を確立し、実際にどの塩基配列モチーフが新規な発見を含むものであるかを見極めていくことも課題としてあげられる。高等真核生物においても複数の生物種のゲノム配列が利用可能となっており、それらの比較解析も今後の課題としてあげられる。複数の生物種間で比較し、対象とする遺伝子配列と相同な部分だけでなく、その近傍の遺伝子や制御因子などのデータ、さらにはそれらの発現についての情報を組み合わせ、相互に関連付けることによる、遺伝子単体ではなくその属するシステムとして比較解析が必要になるであろう。体系的に機能情報を探索する方法論的基礎の開発とそれに基づく実際の解析、そして検証実験が課題としてあげられる。それにはゲノム配列上での遺伝子の配置、クロマチン構造も考慮に入れたうえで転写ユニットを仮定して発現解析を進める必要があるであろう。

〈研究期間の全成果公表リスト〉

- 1) 論文/プロシーディング
- 1. 0202271508
- Tohsato, Y., Matsuda, H., and Hashimoto, A.: An

Application of a Pathway Alignment Method to the Analysis of Metabolic Pathways, *Research Communications in Biochemistry, Cell and Molecular Biology*, 15(3/4), 179-191 (2001).

2. 0202271502  
Bono, H., Kasukawa, T., Hayashizaki, Y., and Okazaki, Y. READ: RIKEN Expression Array Database *Nucleic Acids Research* 30, 211-213 (2002)

3. 0202271505  
Bono, H., Kasukawa, T., Furuno, M., Hayashizaki, Y., and Okazaki, Y. FANTOM DB: database of Functional Annotation of RIKEN Mouse cDNA Clones *Nucleic Acids Research* 30, 116-118 (2002)

4. 0202271504  
Kawaji, H., Schoenbach, C., Matsuo, Y., Kawai, J., Okazaki, Y., Hayashizaki, Y. and Matsuda, H.: Exploration of Novel Motifs derived from Mouse cDNA sequences, *Genome Research*, 12(3), 367-378 (2002).

5. 0210021709  
Bono, H. and Okazaki Y. Functional transcriptomes: comparative analysis of biological pathways and processes in eukaryotes to infer genetic networks among transcripts *Current Opinion in Structural Biology*, 12, 355-361 (2002).

6. 0306271317  
Bono, H., Nikaido, I., Kasukawa, T., Hayashizaki, Y., Okazaki, Y.; RIKEN GER Group; GSL Members: Comprehensive analysis of the mouse metabolome based

- on the transcriptome., *Genome Research*, 13(6b), 1345-1349 (2003)
7. 0303310253  
Miyake, S., Tohsato, Y., Takenaka, Y., Matsuda, H.: A Clustering Method for Comparative Analysis between Genomes and Pathways, 8th International Conference on Database Systems for Advanced Applications (DASFAA 2003), 327-334 (2003)
8. 0306271333  
Bono, H., Yagi, K., Kasukawa, T., Nikaido, I., Tominaga, N., Miki, R., Mizuno, Y., Tomaru, Y., Goto, H., Nitanda, H., Shimizu, D., Makino, H., Morita, T., Fujiyama, J., Sakai, T., Shimoji, T., Hume, D.A., Hayashizaki, Y., Okazaki, Y.; RIKEN GER Group; GSL Members: Systematic expression profiling of the mouse transcriptome using RIKEN cDNA microarrays, *Genome Research*, 13(6b), 1318-1323 (2003)
9. 0307182311  
Kanapin, A., Batalov, S., Davis, M.J., Gough, J., Grimmond, S., Kawaji, H., Magrane, M., Matsuda, H., Schoenbach, C., Teasdale, R.D., RIKEN GER Group and GSL Members, and Yuan, Z.: Mouse Proteome Analysis, *Genome Research*, 13(6b), 1335-1344 (2003).
10. 0403130142  
Kasukawa, T., Bono, H., Hayashizaki, Y., Okazaki, Y. and Matsuda, H.: MaXML: Mouse Annotation XML, *In Silico Biology*, 4(3), 7-15 (2004).
11. 0309021446  
Kawaji, H., Takenaka, Y., and Matsuda, H.: Graph-based Clustering for Finding Distant Relationships in a Large Set of Protein Sequences, *Bioinformatics*, 20(2), 243-252 (2004).
12. 0411301746  
Horie-Inoue, K., Bono, H., Okazaki, Y. and Inoue, S.: Identification and functional analysis of consensus androgen response elements in human prostate cancer cells. *Biochem Biophys Res Commun.* 325(4), 1312-1317 (2004)
13. 0408241817  
瀧浩平、寺本礼仁、竹中要一、松田秀雄：遺伝子の機能分類を利用した遺伝子制御ネットワーク推定手法, *情報科学技術レターズ*, 3, 23-24 (2004).
14. 0408241842  
Nakamura, Y., Itoh, T., Matsuda, H., and Gojobori, T.: Biased Biological Functions of Horizontally Transferred Genes in Prokaryotic Genomes, *Nature Genetics*, 36(7), 760-766 (2004).
15. 0601231335  
Bono, H.: SayaMatcher: genome scale organization and systematic analysis of nuclear receptor response elements, *Gene*, 364, 74-78 (2005)
16. 0601311726  
Bono, H. and Okazaki, Y.: The study of metabolic pathways in tumors based on the transcriptome, *Seminars in Cancer Biology*, 15(4), 290-299 (2005)
- 2) データベース/ソフトウェア
17. 0210021658  
READ (Riken Expression Array Database)  
<http://READ.gsc.riken.jp/>
18. 0602061742  
SayaMatcher  
<http://genome.saitama-med.ac.jp/SayaMatcher/>
- 3) 特許など
- なし
- 4) その他
19. 0303231742  
坊農秀雅、粕川雄也、古野正朗、林崎良英、岡崎康司  
FANTOM-DB: マウスcDNA機能アノテーションデータベース 生化学, 75, 149-152 (2003)
20. 0302082319  
坊農秀雅、中尾光輝 マイクロアレイ・DNAチップデータを活用する 蛋白質 核酸 酵素, 48, 167-172 (2003)