

厚生労働科学研究費補助金

トキシコゲノミクス研究事業

ヒト末梢血中リンパ球を用いたトキシコゲノミクス基盤研究(H17-トキシコ一般-006)

平成18年度 総括・分担研究報告書

主任研究者 藤村 昭夫

平成 19(2007)年 4 月

目 次

I. 総括研究報告

ヒト末梢血中リンパ球を用いたトキシコゲノミクス基盤研究	1
藤村 昭夫	

II. 分担研究報告

1. “マイクロアレー実施・データ解析・精度管理”に関する研究	3
大島 康雄	
別紙1 遺伝子解析研究倫理審査委員会設置規程	24
別紙2 遺伝子解析研究許可申請書	26
別紙3 遺伝子解析研究許可決定通知書	39
2. バイオインフォマティクス	40
篠原 歩	

III. 研究成果の刊行に関する一覧表	42
---------------------	----

V. 研究成果の刊行物・別刷	添付
----------------	----

厚生労働科学研究費補助金（トキシコゲノミクス研究事業）
総括研究報告書

ヒト末梢血中リンパ球を用いたトキシコゲノミクス基盤研究

主任研究者 藤村昭夫 自治医科大学臨床薬理学 教授

研究要旨

腎障害有りおよび腎障害無しの薬物を EBV 不死化 B リンパ球に曝露させ、遺伝子発現解析を Affymetrix GeneChip System により行なった。このデータを基に、腎障害未知の薬物について、腎障害性を予測するクラスプレディクションを行なう手法を開発している。現在のところ、コンピューターへ腎障害の有無情報を与えずに発現データのみを入力し、その遺伝子発現を誘導した薬物の腎障害性を予測すると、80%以上の正確性で、腎障害性が予測できた。一方、発現データを基に薬物による腎障害メカニズムへアプローチして、その機序を明らかにし、さらに、治療への応用モデルを構築し論文として報告するとともに、特許申請を行なった。

A. 研究目的

創薬時には、前臨床研究によって安全とされる化合物が選ばれるが、臨床開発時には、さらにその安全性を確認する必要がある。この臨床研究でもトキシコゲノミクスは有力な研究手段となるが、その際、患者さんから、より負担の少ない方法で臨床検体を採取する必要がある。そのような検体として、ヒト末梢血中有核細胞、特にリンパ球が考えられる。そこで本研究では、1) マイクロアレーデータ解析および精度管理方法の開発、2) 腎障害性薬物と腎障害が知られていない薬物を曝露後の、両群間で有意に発現量に差のある遺伝子の探索、さらに、この有意な変化を示す遺伝子が腎障害性を予測する上で有用か否かの評価、3) 両群間で発現量に差のある遺伝子情報を基にした腎毒性メカニズムの探索、などを行い、臨床検体を用いるトキシコゲノミクス研究を行なう上で必要な基礎的なデータを蓄積する事を目的とする。

B. 方法

本研究では以下のような項目につき検討した。

- 1) マイクロアレーデータ解析および精度管理方法の開発。---多検体のデータ管理に有用なパラメータをアジレントバイオアナライザー2100 やアフィメトリクス GCOS ソフトウェアから出力されるファイルより抽出し、効率的に処理する方法を開発した。
- 2) 腎障害性薬物と腎障害が知られていない薬物を曝露後の、両群間で有意に発現量に差のある遺伝子の探索。さらに、この有意な変化を示す遺伝子が腎障害性を予測する上で有用か否かの評価。---様々なパラメータを用いた実験を行ない、最適のパラメータを見いだした。これらの遺伝子リスト

やパラメータが腎障害性の予測に有用か否かを評価した。

- 3) 両群間で発現量に差のある遺伝子情報を基にした腎毒性メカニズムの探索。---As2O3 曝露後に発現が増加する HMOX-1 および一般的に腎毒性の知られている薬物により相対的に発現が抑制される AIF-1 の機能解析を行なった。

以上のそれぞれの具体的な方法は分担研究者大島の報告書へ記載した。

C. 結果・D. 考案

- 1) リンパ球を用いたトキシコゲノミクス研究の基盤整備については、予定していた薬物曝露遺伝子発現実験を H17 年度中に終了し、H18 年度中に基礎的なデータ解析を終了した。2) 腎障害性薬物の評価をリンパ球によって行う際に用いる新しい DNA チップを作成する場合に有用な遺伝子の選別とその遺伝子リストの有用性の評価については、3 年間で曝露予定の薬物 160 を全て用いて薬物曝露遺伝子発現実験を平成 18 年度中に終了した。新たな DNA チップへ搭載するトランスクリプトの選別を平成 18 年度中に行なった。平成 19 年度よりその新たなチップに搭載される遺伝子リストの有用性を評価する。3) 薬物性腎障害の機序解明にリンパ球を用いることの妥当性の検討については、平成 17 年度中に薬物曝露遺伝子発現実験を終了した。平成 18 年度中に解析を行なったが、遺伝子発現の共通性と言う視点からはその機序へ迫る事は困難と考えられた。

E. 結論

これまでのところ当初の予定通りに研究は進行している。平成 19 年度は選択された遺伝子を用いた

DNA チップの有用性を評価する予定である。

F. 健康危険情報

なし

G. 研究発表

トキシコゲノミクス研究論文 3 報（下記）、学会発表 1 件（下記）、特許申請 2 件（1 件申請済み（下記）、1 件準備中）、臨床薬理研究振興財団研究奨励賞 1 件

〈論文〉

[1] H. Ando, Y. Oshima, H. Yanagihara, Y Hayashi, T Takamura, S Kaneko, and A. Fujimura, Profile of rhythmic gene expression in the livers of obese diabetic KK-A(y) mice. *Biochem Biophys Res Commun* 346 (2006) 1297-302

[2] A. Sasaki, Y. Oshima, S. Kishimoto and A. Fujimura, Individual Differences in Gene Expression in Primary Cultured Renal Cortex Cells Derived from Japanese Subjects *IP SJ Transactions on Bioinformatics* 47 (2006) 67-72

[3] A. Sasaki, Y. Oshima, and A. Fujimura, An approach to elucidate potential mechanism of renal toxicity of arsenic trioxide *Experimental Hematology* 35 (2007) 252-262

〈学会〉

臨床薬理学会 2006 年度、年会 シンポジウム S11 「ゲノム解析による副作用の予測」（12 月 1 日（金曜日）15:00～17:10 第 1 会場（本館 5 階コンコード A+B））、座長：藤村 昭夫（自治医科大学 臨床薬理学部門）、長尾 拓（内閣府 食品安全委員会）

H. 知的財産権の出願・登録状況（予定を含む）

[1] 出願番号：特願 2007-47053, 発明の名称：腫瘍治療における三酸化二ヒ素の副作用低減剤及び解毒剤

[2] 出願準備中：リンパ球を用いた腎毒性の予測

厚生労働科学研究費補助金（トキシコゲノミクス研究事業）
分担研究報告書

”マイクロアレー実施・データ解析・精度管理”に関する研究

分担研究者 大島康雄 自治医科大学臨床薬理学 助手
研究協力者 津田英利 自治医科大学臨床薬理学 ポストドクター

研究要旨

我々は萌芽の先端医療技術推進研究事業「トキシコゲノミクス」（平成 14—16 年度）において患者の手術検体から腎細胞を培養し、これをトキシコゲノミクス研究に用いる手法を確立した。新薬の臨床開発時に患者から検体を採取する場合、できる限り負担の少ない方法で臨床検体を採取する必要がある。そのような検体として、ヒト末梢血中有核細胞がまず考えられる。そこで本研究ではこれまでの研究成果を生かして、以下の項目を検討した。1) 多検体のサンプルの状態の評価方法の開発。2) リンパ球に腎障害性薬物と腎障害が知られていない薬物を曝露させた後の、両群間で有意に発現量に差のある遺伝子の探索。さらに、この有意な変化を示す遺伝子が腎障害性を予測する上で有用か否かの評価。3) メカニズムの探索。4) これまで我々が得た遺伝子発現データの評価をするための外部のデータベースとの発現データの比較。5) 臨床応用する際の問題点の探索。結果として、1) 多検体の状態を効率的に評価する手順を開発した。2) 2005 年度と併せて添付文書上副作用として急性腎不全などの腎障害が記載されている薬物とそのような記載のない薬物を 80 種類ずつ選び出し、合計 160 種類曝露・遺伝子発現実験を行った。データの評価 1) で開発した方法を用いて、クオリティが問題ないと評価された 125 検体について、腎障害性の有無を予測する方法を開発した。様々な機序による腎障害を含めて 80%を超える正確性でその腎障害性が予測できた。3) メカニズムの探索としては薬物を絞り込んで、白血病治療薬 As₂O₃ の腎毒性を解析し、酸化ストレスがその毒性機序と関連している事が推定された。さらに、その阻害薬を使用する事で毒性が軽減され、かつ薬効はほとんど抑制されない事が示された。その有効性は臨床的にも評価される価値があるものと考えられた。4) 我々が発現解析を行なった薬物未曝露の発現データと公共のデータベース CGAP に登録されている発現データベースの情報を比較したところ、有意に発現がある遺伝子に相関が見られ、両データが適切な情報を含んでいる事が推定された。5) 臨床応用を行なう上で必要な基礎的な検討に関する臨床試験計画が倫理審査を経て承認され、少数例ではあるが被験者のエントリーが開始された。

リンパ球に腎障害性薬物と腎障害が知られていない薬物を曝露させ、両群間で有意に発現量に差のある遺伝子を用いることにより、計算機に対して腎障害性の有無の情報を入力せずに発現情報を入力する事で、80%を超える正確性をもって腎障害性を予測する事ができた。これらの遺伝子リストなどの情報は、腎障害性を予測するための新たな DNA チップを開発する上で重要な情報を提供するものと考えられる。

A. 研究目的

我々はマイクロアレーテクノロジーを毒性予測に応用する「トキシコゲノミクス」研究事業により、プライマリーヒト組織を用いた薬物による腎障害について研究を行ってきた。このトキシコゲノミクス研究は、将来、臨床開発時にヒトを対象として、薬物が投与された際の末梢血液中の有核細胞を用いて行うこととなる。本研究ではその前提として必要と思われる基礎検討を行う。平成 17 年度は 1) 検体処理方法についての基礎調査、2) リンパ球を用いた遺伝子発現解析研究、3) 薬物による腎障害の発現を、リンパ球という毒性発現とは異なる組織で予測する科学的根拠の考察を行った。

平成 18 年度は 1) 多検体のサンプルの状態の評価方法の開発。2) リンパ球に腎障害性薬物と腎障害が知られていない薬物を曝露させた後の、両群間で有意に発現量に差のある遺伝子の探索。また、この有意な変化を示す遺伝子が腎障害性を予測する上で有用か否かの評価。3) 腎毒性薬物の毒性メカニズムの探索。4) これまで我々が得た遺伝子発現データの評価をするための外部のデータベースとの発現データの比較。5) 臨床応用する際の問題点の探索。

1) 多サンプルの状態の評価方法の開発
---rRNA[28S/18S]比, Percent Present, 3'/5'比

RNA のスペクトロフォトメータによる検体の計測はその定量以外に OD260/280 比による検体の質の情報を提供する。それは、有機溶媒や蛋白質・核酸断片等のコンタミネーションにより値や波形が変化するため、これら夾雑物の有無・ある場合はその程度の評価が可能と考えられる。しかし、それだけでは GeneChip による発現解析には、OD260/280 比の評価では十分ではない事が知られてきている。我々はその他の評価方法をさらに組み合わせてより厳格なサンプルの質の管理を行ってきたが、多サンプルの評価を行うのに適した形ではそれぞれの測定機器がデータを提供しないため、全て手作業で機器からのプリントアウトをチェックする必要があった。様々な機器から出力される情報を集計しレビューするうえで、より定型的手順を開発し質の評価の効率化を図ることを目的として、2100 Bioanalyzer や GeneChip の出力データを処理するスクリプトを開発した。

2) 発現解析-クラスプレディクション

本研究プロジェクトにおいてリンパ球に腎障害性薬物と腎障害が知られていない薬物を曝露させ、両群間で有意に発現量に差のある遺伝子の探索を行ってきた。2006 年度内に予定の曝露実験を終了したため、二群間で有意な変化を示す遺伝子を選び、これが腎障害性を予測する上で有用か否かを評価した。この腎障害性を予測することができるのであればこれらの遺伝子情報が新たな DNA CHIP を作成する上で有用と考えられた。

3) 毒性メカニズムの解析

As₂O₃ 毒性メカニズムの解析

毒性メカニズムにアプローチする上で、様々な毒性の機序による、様々な薬物を一度に解析する方法は適していないと思われたため、今回は多くのデータの中で、arsenic trioxide 曝露後の発現実験により得られたデータについて着目した。本研究では薬物曝露後、HMOX1 遺伝子が有意に発現が上昇すると考えられた事を発展させ、その機序へアプローチし、さらにその毒性メカニズムを抑制する事で治療的な可能性も探索した。

AIF-1(Allograft inflammatory factor-1) 毒性メカニズムとの関連の解析

AIF-1 は移植後の心臓、肝臓、腎臓等の組織で発現が見られると共に、T 細胞やマクロファージ等でも細胞刺激因子により発現が見られる遺伝子である。血管内皮細胞に対して細胞増殖性を上げる効果が認められているが、B 細胞自身への影響は検討されていない。そこで、腎障害無しの薬物群で有意に発現上昇が見られた AIF-1 の B 細胞お

よび腎細胞 (293) への影響を遺伝子導入を行い、過剰発現することにより検討を行った。

4) 発現データと外部データの比較

これまでの発現データの質を評価する上で、RNA の質を評価方法を行ってきた。しかし、内部的なデータの状態の評価でしかなく、何らかのアプローチが求められる事があった。これに対応するために、蓄積した我々の発現データと、外部のリソースと比較しその類似性を評価する事とした。本プロジェクトで解析対象とした日本人由来腎組織の遺伝子発現データで、高発現と評価したものと、CGAP, Cancer genome Anatomy Project の EST および SAGE リソースに存在しているデータ類似性を検討し、その信頼性を評価した。

5) 臨床試験

トキシコゲノミクス研究において、臨床の現場で何らかの疾患治療のために薬物を投与される患者を対象として、末梢血リンパ球の遺伝子発現を解析するというアプローチを目指して、基礎研究を行ってきた。本セクションでは臨床の現場で遺伝子発現研究を行う上での実施上の問題点を明らかにする事を目的として、比較的侵襲の少ない方法で、少人数の患者を対象として、実際に検体を得て解析する事とした。

B. 方法

1) rRNA [28S/18S]比

OD260/280 比は、有機溶媒や蛋白質・核酸断片等のコンタミネーションにより値や波形が変化するため、これら夾雑物の有無・ある場合はその程度の評価が可能と考えられる。薬物曝露実験、RNA 純化は前年度報告の通り行なった。各種薬物を曝露した後、NanoDrop スペクトロメータ ND-10 (米国、NanoDrop 社)により OD260/280 比を確認し、1.8 以上のサンプルのみを選択し、その後の解析に用いた。

我々の使用しているヒト由来の HEV 細胞は多くの他の細胞株と同様に、total-RNA で抽出した場合、mRNA より rRNA が多く抽出される。電気泳動上主要なピークを示す rRNA である 18S と 28S の分解性の違いを基に RNA が適切な環境下で保存・抽出された物が何らかの分解を受けやすい環境におかれた物かが評価できると考えられる。OD260/280 比が 1.8 以上であったサンプルについて、Agilent Bioanalyzer 2100 (以下 Bioanalyzer 2100 とします。Agilent Technologies) を用いて 28S/18S 比を評価した。多検体について情報をサマライズするには以下のようにデータを扱った。

Bioanalyzer は他の多くのデータとともに 18S/28S 値を、*Results.csv ファイルに Export する。出力されたファイルは以下のような形式となっている<表 1>。

表 1 Bioanalyzer の出力ファイル (一部)

```
Data File Name,2100
expert_EukaryoteTotal RNA
Nano_DE34903227_2005-11-24_15-20-35.xad
Data File Path,C:\Program
Files\Agilent\2100 bioanalyzer\2100
expert\Data\2005-11-24
Date Created,"Thursday, November
24, 2005 6:20:35 AM"
Date Last Modified,"Thursday,
November 24, 2005 6:45:18 AM"
Version Created,B.01.02.SI136
Version Last Modified,B.01.02.SI136

Assay Name,Eukaryote Total RNA Nano
Assay Path,C:\Program
Files\Agilent\2100 bioanalyzer\2100
expert\assays\electrophoresis\RNA
Assay Title,Eukaryote Total RNA
Nano
Assay Version,1.4

Number of Samples Run,12

Sample Name,#17-31

Peak Table
Time corrected area,Area,Aligned
Migration Time [s],Peak Height,Peak
Width,% of Total
3.5,0.8,22.50,2.4,0.8,0.0
5.3,1.5,26.79,3.7,1.0,0.0
28.0,11.8,41.10,26.0,2.0,0.0
5.2,2.3,43.54,1.9,1.9,0.0
2.1,1.0,44.32,1.9,0.6,0.0
2.6,1.2,45.15,2.3,0.8,0.0
55.2,27.1,47.78,24.9,4.1,0.0

Fragment Table
Name,Start Time [s],End Time
[s],Area,% of total Area
18S,39.84,42.03,27.3,22.1
28S,45.54,49.68,51.2,41.5

Overall Results:
RNA Area,123.4,
RNA Concentration,463,ng/μl
rRNA Ratio [28s / 18s],1.9,
S/N ratio,0.0229,
```

Bioanalyzer 2100 が出力した*Results.csv ファイルを、作業を行う計算機の一つのワーキングディレクトリへ全てコピーした。そして、以下のスクリプトを実行した。この様な情報が本プロジェクト分の

みでも 200 を超える回数繰り返し出力された。一方黄色に背景を着色した情報が今回の評価に必要な情報である。一回のアッセイで最大でも 12 検体処理できる。一方、複数回に分けておこなわれた情報を 1 ファイルで出力する事はできず、この *Results.csv ファイルは結果、複数のファイルとなる。これらのファイルの中の Sample Name と rRNA Ratio [28s / 18s] の情報を順次抽出する。単純な作業であるが、巨大なファイルの中のごく一部のデータを抽出する作業が数百回繰り返される事になるため、この集計を半自動化するための手順を開発した<表 2>。

表 2 データ処理スクリプト 1

```
$ awk '/rRNA/' *Results.csv >
rRNA.txt
$ awk '/Sample Name/' *Results.csv
> SampleName.txt
```

これらの結果生成されるファイルは Bioanalyzer の実行 OS が採用している改行コード (LF-CR)を含んでいるため、これを UNIX 改行コードである LF へ変換した<表 3>。

表 3 スクリプト実行結果ファイル 1 (一部)

```
<rRNA.txt>
rRNA Ratio [28s / 18s],0.0,
rRNA Ratio [28s / 18s],1.6,
rRNA Ratio [28s / 18s],1.8,
rRNA Ratio [28s / 18s],1.7,
rRNA Ratio [28s / 18s],2.2,
rRNA Ratio [28s / 18s],3.2,...

<SampleName.txt>
Sample Name,#18-5
Sample Name,#18-6
Sample Name,#18-7
Sample Name,#18-8
Sample Name,#18-9
Sample Name,#18-10...
```

複数ファイルになっているこれら 2 出力ファイルを以下の通りマージした<表 4>。

表 4 データ処理スクリプト 2

```
$ paste *.txt > output.txt
```

表5 スクリプト実行結果ファイル2 (一部)

```
<output.txt>
Sample Name,#18-5 rRNA Ratio [28s / 18s],0.0,
Sample Name,#18-6 rRNA Ratio [28s / 18s],1.6,
Sample Name,#18-7 rRNA Ratio [28s / 18s],1.8,
Sample Name,#18-8 rRNA Ratio [28s / 18s],1.7,
Sample Name,#18-9 rRNA Ratio [28s / 18s],2.2,
Sample Name,#18-10          rRNA Ratio [28s / 18s],3.2,...
```

この結果出力される output.txt の改行コードを引き続き処理を行なう OS に適した物へ変換 (我々は CR へ変換) して、汎用表計算ソフトや統計ソフトにより開くことができた<表5>。

2) Percent Present, 3'/5'比

OD260/280 比および rRNA 28S/18S 比による QC パスしたサンプルは昨年度までの報告にある通りの方法でラベリングを行ない、HG-U133 チップ (Affymetrix 社) へハイブリダイゼーションを行ない、GeneChip System (Affymetrix 社) で検出し、GCOS ソフトウェア (Affymetrix 社) で初期の解析を行なった。HG-U133 チップを用いたシステムは RNA を始め、用いられたサンプルの状態を評価するために有用なパラメータを RPT(report)ファイルとして出力する事ができる。我々はこのデータを多数集計するため、効率の良い手順を見いだしたのでこれを以下に記載する。2005-2006 年の2年間で曝露した結果のデータは216のチップデータとなった。(同じサンプルの繰り返しスキャンや、再曝露などを除き、採血条件の検定のためのサンプルと除くと、腎障害あり無し予測と関連して使用するデータは142サンプルとなる)それぞれに対して1つのRPTファイルが出力されたので report file 全てを一つのワーキングディレクトリへコピーした。GCOS ソフトウェアの仕様のため Expression Analysis Settings の Scaling の項目で、Target Signal に設定した値が記載される行があるが、Report file 内の TGT Value は、User Defined のオプションが選択されている場合は、Scale Factor で設定するで、TGT Value の値が記載されない。一方、All Probe Sets または Selected Probe Sets のオプションが選択された場合は、必ず Target Signal に設定値がレポートファイル内に記載される。機器の設定が一定でないと TGT Value の値がある Report と 値がない Report が存在する。まず、TGT Value の記述のある行をすべて除いたファイルを以下の手順で処理した<表6><表7>。

下の表に Report ファイルの一部を示す。背景が黄色に着色されている部分が今回抽出すべき対象の情報である。

表6 データ処理スクリプト3

```
$ paste *.rpt > sum.txt
```

このデータ処理スクリプトを実行する事で、216あったファイルが1つのファイルへマージされた。

表7 Report ファイル (一部)

Report Type:	Expression Report
Date:	05:11PM 10/19/2005...
Filename:	#17-11.CHP
Probe Array Type:	HG-U133A
Algorithm:	Statistical
Probe Pair Thr:	8
Controls:	Antisense Central-
	Avg: 8102 Count: 9...
The following data represents probe sets that exceed the probe pair threshold and are not called "No Call".	
Total Probe Sets:	22283
Number Present:	8240 37.0%
Number Absent:	13777 61.8%
Number Marginal:	266 1.2%
Average Signal (P):	106.0
Average Signal (A):	2.8
Average Signal (M):	9.6
Average Signal (All):	41.1...
Housekeeping Controls:	
Probe Set	Sig(5')
Det(5')	Sig(M')
Det(M')	Sig(3')
Det(3')	Sig(all)
	Sig(3'/5')
AFFX-HUMGAPDH/M33197	311.1 P
367.6 P	913.0 P
530.59	2.93
AFFX-HSAC07/X00351	69.0 P
345.8 P	2417.3 P
944.05	35.01
AFFX-M27830 2.6	A 19.2 P
4.0 A	8.59 1.58

表8 データ処理スクリプト3の実行結果 (一部)

```

Report Type:      Expression
Report
Report Type:      Expression
Report
Report Type:      Expression
Report
...
Date:              07:54PM
02/15/2006
Date:              07:54PM
02/15/2006
Date:              07:54PM
02/15/2006
...
Filename:          #8-29A.CHP
Filename:          #8-3 A.CHP
Filename:          #8-4 A.CHP
...
Number Present:   10607 47.6%
Number Present:   6887 30.4%
Number Present:   10537 47.3%
...
AFFX-HUMGAPDH/M33197 1542.3
P 1895.4 P
1559.5 P 1665.73
1.01
AFFX-HUMGAPDH/M33197 1513.6
P 1316.2 P
1297.4 P 1375.74
0.86
AFFX-HUMGAPDH/M33197 1455.5
P 1537.8 P
1462.2 P 1485.13
1.00
...
AFFX-HSAC07/X00351 1479.4
P 1688.7 P
1415.5 P 1527.83
0.96
AFFX-HSAC07/X00351 1597.2
P 1852.3 P
1585.8 P 1678.43
0.99
AFFX-HSAC07/X00351 1483.5
P 1728.8 P
1408.9 P 1540.41
0.95

```

マージ後のファイルは、元ファイルの各行ごとにまとまる(表8)。このファイルは改行コードがCR+LFなので、UNIX改行コードであるLFへ変換した。
次に以下の処理スクリプトを順次実行した(表9～16)。

表9 データ処理スクリプト4

```
$ awk '/Number Present/' sum.txt > present.txt
```

表10 データ処理スクリプト4の実行結果 (一部)

```

Number Present:   9888 44.4%
Number Present:   8240 37.0%
Number Present:   8276 37.1%
Number Present:   8631 38.7%
Number Present:   7486 33.6%
Number Present:   7823 35.1%
Number Present:   2047 9.2%
Number Present:   701 3.1%
Number Present:   6611 29.7%
Number Present:   1186 5.3%
...

```

表11 データ処理スクリプト5

```
$ awk '/AFFX-HSA/' sum.txt > HSA.txt
```

表12 データ処理スクリプト5の実行結果 (一部)

```

AFFX-HSAC07/X00351 121.7 P
467.4 P 1557.6 P
715.58 12.79
AFFX-HSAC07/X00351 69.0 P
345.8 P 2417.3 P
944.05 35.01
AFFX-HSAC07/X00351 71.3 P
347.8 P 2373.3 P
930.80 33.28
AFFX-HSAC07/X00351 78.9 P
380.1 P 2639.1 P
1032.69 33.45
AFFX-HSAC07/X00351 98.8 P
294.6 P 2056.3 P
816.59 20.82
AFFX-HSAC07/X00351 184.8 P
407.5 P 2927.4 P
1173.23 15.84
AFFX-HSAC07/X00351 1.7 A
10.7 P 1114.7 P
375.71 654.21
AFFX-HSAC07/X00351 1.7 A
7.3 P 804.7 P
271.25 467.79
AFFX-HSAC07/X00351 45.4 P
256.8 P 1143.0 P
481.74 25.16
AFFX-HSAC07/X00351 0.4 A
8.2 P 1132.7 P
380.46 2521.32
...

```

表 13 データ処理スクリプト 6

```
$ awk '/AFFX-HUMGAPDH/' sum.txt >
GAPDH.txt
```

表 14 データ処理スクリプト 7 の実行結果 (一部)

AFFX-HUMGAPDH/M33197	289.6 P		
323.9 P	803.0 P		
472.19	2.77		
AFFX-HUMGAPDH/M33197	311.1 P		
367.6 P	913.0 P		
530.59	2.93		
AFFX-HUMGAPDH/M33197	273.3 P		
329.4 P	848.2 P		
483.64	3.10		
AFFX-HUMGAPDH/M33197	350.1 P		
444.5 P	1079.8 P		
624.81	3.08		
AFFX-HUMGAPDH/M33197	149.9 P		
265.3 P	722.4 P		
379.22	4.82		
AFFX-HUMGAPDH/M33197	241.3 P		
334.4 P	864.5 P		
480.07	3.58		
AFFX-HUMGAPDH/M33197	8.8 P		
10.1 P	30.7 P		
16.52	3.51		
AFFX-HUMGAPDH/M33197	2.3 A		
3.0 P	12.7 P		6.02
5.52			
AFFX-HUMGAPDH/M33197	140.8 P		
182.5 P	465.1 P		
262.78	3.30		
AFFX-HUMGAPDH/M33197	7.0 P		
9.6 P	35.4 P		
17.34	5.03		
...			

表 15 データ処理スクリプト 7

```
$ awk 'Filename' sum.txt >
samplename.txt
```

表 16 データ処理スクリプト 7 の実行結果 (一部)

Filename:	#17-10.CHP
Filename:	#17-11.CHP
Filename:	#17-12.CHP
Filename:	#17-13.CHP
Filename:	#17-15.CHP
Filename:	#17-16.CHP
Filename:	#17-17.CHP
Filename:	#17-19.CHP
Filename:	#17-2.CHP
Filename:	#17-21.CHP
...	

上記スクリプト 4-7<表 9, 11, 13, 15>の実行結果のファイルである、present.txt, HSA.txt, GAPDH.txt, samplename.txt の改行コードを引き続き処理を行なう OS に適した物へ変換 (我々は CR へ変換) して、汎用表計算ソフトや統計ソフトにより開くことができた。

3) 発現解析-クラスプレディクション

QC パス後の 125 サンプルのデータを GeneSpring (Agilent Technology) ソフトウェアへ読み込ませ、前年までに報告した方法と同様に normalization をい行なった。短く記述すると発現量 0.01 未満の値を 0.01 へ切り上げ、BetterHK リストに従い perChip normalization を行なった。

その後昨年までと同様に noise reduction を行なった。短く記述すると、Filter on Flags を行ない、125 サンプル中 75 以上のサンプルにおいて P または、M のフラグを有するトランスクリプトのみをその後の解析に使用した。この処理により 45,063 transcript が 9,351 transcript までノイズを低減する事ができた。

その後、GeneSpring 埋め込みの Volcano Plot 法等を応用し、腎障害の有無と関連して発現状況が変わると考えられる transcript を検出した。

この腎障害の有無と関連して発現状況が変わると考えられる transcript のリストを数個から 100 以下程度までの範囲で変動させて、クラスプレディクションの結果を評価した。クラスプレディクションを行なう場合は、2005 年度に実験を行なったサンプル群 (2005 サンプル) の腎障害有り無しの情報に基づき、CrossVaridation 法によりクラスプレディクションに用いるパラメーターを探索した。2005 サンプルについて正確性が極大となるパラメーターを用いて、腎障害有無の情報をコンピュータへ与えることなしに 2006 サンプル (2006 年度に実験を行なったサンプル) の予測を行なわせた。その結果の腎障害の有無予測の正確性を評価した。

4) 毒性メカニズムの解析

As2O3 毒性メカニズムの解析

毒性メカニズムにアプローチする上で、様々な毒性の機序による、様々な薬物を一度に解析する方法は適していないと思われたため、今回は多くのデータの中で、arsenic trioxide 曝露後の発現実験により得られたデータについて着目した。発現データは、通常の normalization および noise reduction を行なわれた後、k-means クラスタリングによって、5 つのクラスターへ遺伝子が分類された。この中より遺伝子発現が経時的に上昇する傾向にある HMOX1 遺伝子を対象として、その発現を定量 PCR を行い確認した。蛋白質レベルでの発現誘導の有無をウエスタンブロットにより確認した。

また、HEK293 細胞において HMOX1 遺伝子を CMV プロモーター下に発現させたコンストラクトを作成しこれをリポゾーム法を応用して細胞へトランスフォームした。arsenic trioxide による殺細胞活性を、発現誘導していない細胞株と比較した。

arsenic trioxide 曝露により酸化ストレスと関連した HMOX1 遺伝子が誘導されたため、arsenic trioxide の細胞毒性と酸化ストレスの関連を検討した。1 mcM MCLA を arsenic trioxide 曝露有無で、培養上清へ添加し、その蛍光を IVIS (Xenogen 社) で測定した。

Allograft inflammatory factor-1(AIF-1)のクローニングと機能解析

AIF-1(accession No. U49392)の配列を参考にプライマーを設計し、遺伝子コード領域(147 アミノ酸)を PCR 法にて増幅した。増幅断片を pIRES2-EGFP(Clontech)にクローニングした。このプラスミドを pIRES-AIF1 とし、AIF-1 安定発現細胞株を作る目的で、ヒト胚腎由来で T 抗原を持つ 293T 細胞へ Lipofectamin2000 (Invitrogen)を使用して遺伝子導入した。導入 24 時間後にネオマイシン入りの培地に交換し、プラスミド導入細胞を選抜した。293T 細胞に遺伝子導入しネオマイシンで 3 代選抜した細胞を 293T/AIF1 とした。選抜中に細胞を回収し、AIF-1 がコントロール細胞より発現している事をウエスタンブロッティングで確認した。コントロールとして、pIRES2-EGFP コントロールベクターを導入し、293T/GFP とした。細胞増殖性は、トリパンブルー染色により生細胞を 24 時間毎にカウントする事により評価した。また、細胞活性は WST-1 試薬 (TaKaRa) によってミトコンドリアの活性を測定する事により評価した。また日本人ヒト血球 B 細胞由来の培養細胞株である HEV0034 へもリポフェクチン法で遺伝子導入を行い、WST-1 試薬により細胞活性の評価を行った。尚、HEV0034 への遺伝子導入は Lipofectamin-2000 では導入効率が非常に悪かったので、検討の結果、LipofectaminLTX (Invitrogen)に PLUSreagent (Invitrogen)を組み合わせた方法により導入を行った。

5) 発現データと外部データの比較

我々の発現データの信頼性を、外部のリソースと比較しその類似性を評価する事とした。本プロジェクトで解析対象とした日本人由来腎組織の遺伝子発現データで、高発現と評価したものと、CGAP, Cancer genome Anatomy Project の EST および SAGE リソースに存在しているデータ (EST として捕らえられる遺伝子発現産物は、トランスクリプトが存在している以上発現していると評価できる) の類似性を示す事とした。

CGAP リソースは以下の URL より入手した<表 17>。

表 17 CGAP リソース入手元

```
http://gcap.nci.nih.gov/Tissues/
http://cgap.nci.nih.gov/SAGE/
```

tissue->library finder->Homo sapiens, CGAP LIBRARY KIDNEY により表示される物のうち normal と表示された NCI-CGAP-Kid11 を解析対象とする事とした

CGAP の EST リストをダウンロードし改行コードを(LF) UNIX 用にしたファイル (dl_unix.txt)は 42943 行からなるが、この一部を示す<表 18>

表 18 dl_unix.txt (一部)

```
[west-95-141:Jun06/journal_reply/CGAP]
y_oshima% more download.txt
>gnl|UG|Hs#S11051156 NISC_gg01a05.x1
NCI_CGAP Kid11 Homo sapiens cDNA clone
IMAGE:32531
36 3', mRNA sequence /clone=IMAGE:3253136
/clone_end=3' /gb=CB047156 /gi=27785443
/ug=H
s.624011
/len=555^MAATGCTTTTTTTTTTTTTTTTTTGGACATGGA
GTCCTCACTCTGTTGCCAGGCCTTCTGTC^MTGTAAT
TTTAAATTTATCCCTTTGGGTCGACATTGATAATTTG
TGACTTTTCTTC^MTTGTATCAGTCTTAACAAAATGTGA
CTTTA
ATAGTCTCTCCAAGAACCACTTTTGT^MTTTCCTGAA
GTTTTATATCTTTGTTTCATTGATTACTTGTGTTGCTCTCT
TGTAT
ACTC^MTCTCGGCGTGCCTTACATTTTAAATTAATGCCAGA
CATTTGTGTTGACAATTTTATAGGT^MTATCTGAATGATCA
TTTGC
TGTTAGGATGGACGGCTTACCCACTCCTGTGTTAGGCAGA
^MTAGAATGCAGTTAGATCACCTTAATCACAAAACTTTGA
ATACT
GCAGTTTGCTATGCAG^MACTTTCCTTAGCTTCTTATCTGT
GTAGCTTCACAATTCAGTAGATATCGCTGGAGGGAAA^MTA
GTTAA
GCCTGTGGTTCAGTCTCTGTTTCTTTTTTATGACCAGAAT
CCTGGCCCTTCA^MAGTCTCTTAATATTT^M>gnl|UG|Hs
#S110
51157 NISC_gg01a06.x1 NCI_CGAP Kid11 Homo
sapiens cDNA clone IMAGE:3253138 3', mRNA
seq
uence /clone=IMAGE:3253138 /clone_end=3'
/gb=CB047157
```

配列情報を除く目的で以下のスクリプトを実行した<表 19>。

表 19 スクリプト

```
% awk '/>/' dl_unix.txt > g_num.txt
```

この結果得られるデータの形式を以下に示す<表 20>。

表 20 g_num.txt (一部)

```
>gnl|UG|Hs#S11051156 NISC_gg01a05.x1
NCI_CGAP_Kid11 Homo sapiens cDNA
clone IMAGE:3253136 3', mRNA
sequence /clone=IMAGE:3253136
/clone_end=3' /gb=CB047156
/gi=27785443 /ug=Hs.624011 /len=555
```

g_num.txt も行数としては依然 42943 行あるが、ファイルサイズは 25.8Mb から 8.3M まで減少し、汎用の表計算ソフト等でハンドリングできるようになった。g_num.txt を汎用表計算ソフトで読み込む場合、フィールドセパレータを'/'とする事で、この後の計算に必要な ug=Hs.*のフィールドを取り出す事が可能となる。この ug=Hs.フィールドを濃厚にしたファイルを ug=.xls とした。ug=Hx.*フィールドを有さないトランスクリプト情報が除かれたため、このファイルは 39345 行となった<表 21>。

表 21 ug=.xls (一部)

```
ug=Hs.100043
ug=Hs.100043
ug=Hs.100261
ug=Hs.100299
ug=Hs.10056
ug=Hs.100686
ug=Hs.100743
ug=Hs.100743
ug=Hs.100743
ug=Hs.100743
ug=Hs.100874
ug=Hs.100878
ug=Hs.100914
ug=Hs.101014
ug=Hs.101014
ug=Hs.101014
ug=Hs.101014
ug=Hs.1011
ug=Hs.1011
ug=Hs.1011
ug=Hs.1011
ug=Hs.1011
```

表 21 では並べ替えを行なった後の状態で示した。同じ遺伝子が複数回データベースに登録されたための重複がある事がわずかなサンプルからでも見て取る事ができる。このファイルから'ug='を取り除いて、さらに重複を取り除いて Kid11.txt ファイルへ出力した。Kid11.txt ファイルは 24519 行からなる。この Kid11.txt ファイルを CGAP ホームページの検索画面へ入力し、"get clone"を行なった。その結果のファイルを kid11getclone.txt として保存し

た。このファイルは 32057 行からなる。この中で GeneSymbol が存在する物のみをリストしたところ、24836 行となった。この中から重複する GeneSymbol を除くと、9227 行となった。このリストを kid11symbol.txt---(*)として保存した。

一方、GeneSpring 出力として発現有り(flags are positive/marginal)と発現無し(flags are negatie)のファイルを作成した<表 22>。

表 22 発現有りおよび無しリスト

```
発現有り:expression_pos.txt (10775 行)
発現無し:expression_neg.txt (5437 行)
```

このファイルから GeneSymbol のみを取り出し、重複を取り除いた<表 23>。

表 23 発現有りおよび無しリスト

```
発現有り:expression_pos_symbol.txt (8006 行)---(P)
発現無し:expression_neg_symbol.txt (4232 行)---(N)
```

(P)と(N)の両方のリストに含まれる gene が 1583 存在した。gene_in_both.txt---(B) (*)のりすとから(B)の gene を除いた (ダブリを除いた 8710 行)。kid2.txt---(**) GeneSpring 出力のファイルからもダブリを除いた。(P)のリストから(B)の gene を除いた EPS3.txt---(PP) 6365 行 (N)のリストから (B)の gene を除いた ENS3.txt---(NN) 2756 行 これらのファイルを照合する事で以下の結果を得た。

表 24 2X2 分割表

```
(PP) 中に存在し、(**) 中に存在するもの---3086
(PP) 中に存在し、(**) 中がないもの-----3279
(NN) 中に存在し、(**) 中に存在するもの---1120
(NN) 中に存在し、(**) 中がないもの-----1636
```

この情報を基に市販の汎用統計ソフトを用いて、Fisher's Exact test により検定を行なった。

6) 臨床試験

トキシコゲノミクス研究において、臨床の現場で何らかの疾患治療のために薬物を投与される患者を対象とし、末梢血リンパ球の遺伝子発現を解析

するというアプローチを目指して、基礎研究を行ってきた。本セクションでは臨床の現場で遺伝子発現研究を行う上での実施上の問題点を明らかにする事を目的として、比較的侵襲の少ない方法で、少人数の患者を対象として、実際に検体を得て解析する事とした。具体的な方法や対象は別添書類として添付する、遺伝子解析研究申請書の写しに記載する。この研究課題は遺伝子解析研究倫理審査を経て承認された。手短かに記述すると以下の通りとなる。

臨床試料等の採取の手順

通常の診療では、特別な合併症等の問題のない症例では第三大臼歯の抜去は3回（以上）の受診が行われ、2回目の受診の午後が手術日に予定される。

1回目受診：

術前検査（採血）：手術数日前、日常診療と同様に行う、同意の得られた患者については術前検査の採血に加え、研究用の2ml（1回目）の末梢血の採血を行う。

2回目受診：手術当日

手術手技は日常診療と同様に行う。被験者は昼食後（術前）に1回塩酸セフカペンピボキシル100mg（1回目）を服用する。被験者は手術終了後の夕食後に塩酸セフカペンピボキシル100mg（2回目）を服用する。

3回目受診：手術翌日

術後の消毒・術創の確認を日常診療と同様に行う。被験者は手術翌日の朝食後に、塩酸セフカペンピボキシル100mg（3回目）を服用する。研究用の2ml（2回目）の末梢血の採血を行う。

発現解析

得られた末梢血サンプルよりRNAが抽出する。Affymetrix GeneChip システムを用いて網羅的解析を行うとともに、定量PCR法を利用して発現解析を行う。

本研究計画書はヒトゲノム・遺伝子解析研究に関する当施設の取り扱い規定（別紙）に従い審査され許可された。（申請書・許可通知書を添付する）

C. 結果

1) rRNA [28S/18S]比

以下に電気泳動ゲル様表現した泳動情報を rRNA Ratio [28s/18s] 低値サンプル#17-01, #17-02 及び適値サンプル#17-11, #17-12 を例として示す<図 1>

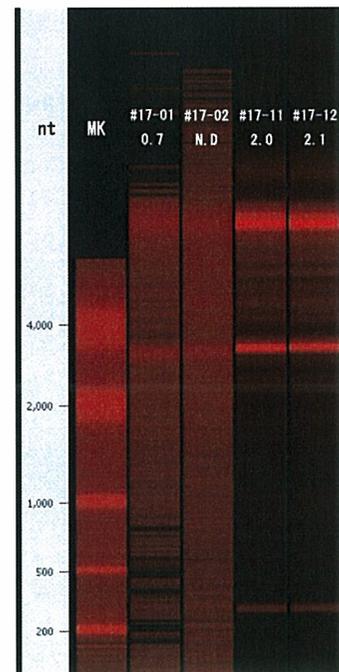


図 1 ゲル様表現した RNA 電気泳動情報
28S/18S 値が適切なサンプル(#17-11, #17-12)は、rRNA の2つの主要なバンドが明確に認識できるが、#17-01, #17-02 などの 28S/18S 値が低値のサンプルはこれらのバンドが不明確となり、一方、これらの分解産物を含むと思われるスメア状のバンドが背景に見られる。

これらのように明確な差が認められるサンプルは目視での判断で十分であるが、中間的に見えるサンプルの扱いを判断するため、より定量的な判断基準をヒストグラムをもとに考慮し、RNA28s/18s 比集計結果をヒストグラムとした<図 2>。

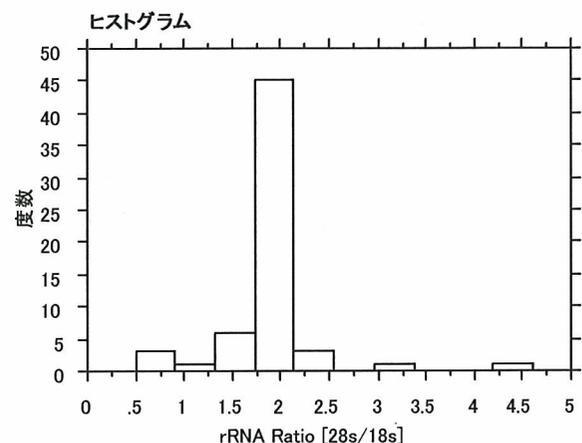


図 2 rRNA Ratio [28s/18s]集計結果

1.3-2.6 の範囲に主要なサンプルがピークを形成する。我々の経験では rRNA Ratio [28s/18s]値が1.0以下のサンプルは状態が不良であり、1.6以上は良好である。図では2005年に行った曝露遺伝子発現結果の集計結果が示されている。本研究では1.3以上を後の解析に使用する事とした。高値の方にはずれ値を示すサンプルが見られる。エレクトロフ

エログラムで確認するとこれらは計算機がピークやベースラインを誤認識していた。正確なベースラインやピークで補正したところ 28s/18s は適値であった。

2) Percent Present, 3'/5'比

以下に HG-U133 チップで計測した 142 サンプルのパラメータを示す。この中で、はずれ値を示した物を黄色の背景で示す<表 25>。

表 25 142 サンプルの QC パラメータ

Filename	GAPDH	HSA	%Present
#18-40.CHP	2.67	20.24	41.60%
#17-1.CHP	2.98	30.55	38.20%
#17-10.CHP	2.77	12.79	44.40%
#17-11.CHP	2.93	35.01	37.00%
#17-12.CHP	3.1	33.28	37.10%
#17-13.CHP	3.08	33.45	38.70%
#17-15.CHP	4.82	20.82	33.60%
#17-16.CHP	3.58	15.84	35.10%
#17-17.CHP	3.51	654.21	9.20%
#17-19.CHP	5.52	467.79	3.10%
#17-2.CHP	3.3	25.16	29.70%
#17-21.CHP	5.03	2521.32	5.30%
#17-22.CHP	6.04	389.15	5.50%
#17-23.CHP	4.08	467.84	6.60%
#17-24.CHP	4.63	274.87	2.40%
#17-26.CHP	4.69	1335.2	3.70%
#17-27.CHP	4.62	306.39	5.00%
#17-28.CHP	5.08	1264.45	5.60%
#17-29.CHP	4.77	455.49	6.10%
#17-3.CHP	2.32	16.98	46.50%
#17-30.CHP	3.96	228.8	5.40%

Filename	GAPDH	HSA	%Present
#17-41.CHP	3.16	22.58	47.70%
#17-42.CHP	2.7	13.37	49.60%
#17-43.CHP	2.37	16.18	41.40%
#17-44.CHP	2.47	13.4	47.30%
#17-45.CHP	2.3	12.68	47.90%
#17-47.CHP	2.67	15.91	39.50%
#17-48.CHP	2.51	22.58	39.50%
#17-49.CHP	2.94	23.01	36.50%
#17-5.CHP	2.8	16.03	45.40%
#17-51.CHP	2.91	21.64	53.40%
#17-53.CHP	2.18	14.04	48.40%
#17-54.CHP	2.1	13.48	48.40%
#17-55.CHP	2.28	13.72	44.70%
#17-56.CHP	2.1	14.37	53.20%
#17-57.CHP	2.27	16.63	43.60%
#17-58.CHP	2.44	20.49	54.20%
#17-59.CHP	2.26	17.95	43.30%
#17-6.CHP	3.55	21.58	22.60%
#17-60.CHP	2.15	17.62	40.00%
#17-61.CHP	2.29	19	42.20%
#17-62.CHP	2.73	24.33	50.10%
#17-64.CHP	2.37	20.98	44.50%
#17-65.CHP	2.65	20.92	43.80%
#17-68.CHP	2.85	26.56	42.30%
#17-69.CHP	2.08	19.32	45.20%
#17-70.CHP	2.81	21.82	39.90%
#17-71.CHP	3.18	19.16	42.70%
#17-72.CHP	2.81	18.08	44.50%

Filename	GAPDH	HSA	%Present
#17-73.CHP	2.82	33.42	42.40%
#17-74.CHP	3.07	35.48	41.70%
#17-75.CHP	3.17	21.23	51.00%
#17-77.CHP	3.05	20.82	44.90%
#17-78.CHP	2.94	16.56	43.00%
#17-8.CHP	3.23	24.56	30.10%
#17-80.CHP	2.94	20.79	51.50%
#17-81.CHP	2.64	15.68	43.40%
#17-82.CHP	3.45	21.06	45.90%
#17-83.CHP	4.67	552.98	28.40%
#17-84.CHP	3.32	26.64	42.00%
#17-85.CHP	4.77	39.63	43.20%
#17-9.CHP	2.36	13.75	40.50%
#18-1.CHP	3.63	22.38	38.50%
#18-10.CHP	2.55	18.72	35.00%
#18-11.CHP	3.63	31.26	37.30%
#18-12-4.CHP	2.39	40.45	38.30%
#18-12.CHP	3.23	35.06	22.60%
#18-13-4.CHP	3.03	12.4	48.80%
#18-14.CHP	3.41	10.7	47.30%
#18-15.CHP	4.06	16.2	40.00%
#18-2.CHP	1.55	6.53	46.20%
#18-3.CHP	1.93	15.16	40.90%
#18-4.CHP	1.96	16.4	38.30%
#18-41.CHP	2.94	31.79	52.60%
#18-42.CHP	3.03	25	53.40%
#18-43.CHP	3.64	28.07	52.80%

Filename	GAPDH	HSA	%Present
#18-44.CHP	3.09	24.07	52.40%
#18-45.CHP	2.67	27.62	54.10%
#18-46.CHP	3.51	34.88	52.10%
#18-48.1.CHP	3.92	35.72	50.30%
#18-5.CHP	1.95	11.11	41.80%
#18-50.CHP	6.62	53.93	44.50%
#18-51.1.CHP	4.99	45.38	49.40%
#18-52.CHP	5.96	46.59	48.40%
#18-53.CHP	4.66	31.28	50.00%
#18-55.CHP	6.04	37.69	47.10%
#18-57.CHP	5.5	38.74	49.30%
#18-58.CHP	6.39	45.2	45.60%
#18-59.CHP	5.4	35.99	49.50%
#18-6.CHP	2.41	13.99	43.70%
#18-60.CHP	4.86	34.96	46.70%
#18-61.CHP	3.38	26.43	46.70%
#18-62.CHP	4.14	28.45	39.70%
#18-63.CHP	5.32	39.07	36.60%
#18-64.CHP	4.08	23.47	45.00%
#18-65.CHP	3.5	24.86	43.40%
#18-66.CHP	3.45	23.3	46.20%
#18-67.CHP	4	22.11	45.20%
#18-68.CHP	3.15	25.49	42.40%
#18-69.CHP	3.68	37.23	40.50%
#18-7.CHP	2.29	14.61	38.90%
#18-70.CHP	2.66	26.06	44.70%
#18-71.CHP	4.02	31.48	42.00%

Filename	GAPDH	HSA	%Present
#18-73.CHP	3.64	23.76	44.00%
#18-74.CHP	2.83	32.17	43.40%
#18-75.CHP	3.86	27.51	43.40%
#18-77.CHP	16.68	252.54	16.50%
#18-78.CHP	4.09	35.26	53.40%
#18-79.CHP	6.05	33.97	52.10%
#18-8.CHP	2.3	17.22	36.80%
#18-80.CHP	3.55	32.84	53.60%
#18-82.CHP	4.5	31.31	52.80%
#18-83.CHP	4.23	27.73	54.20%
#18-85.CHP	3.84	43.2	51.60%
#18-86.CHP	6.36	37.7	52.40%
#18-87.CHP	3.4	29.53	54.20%
#18-88.CHP	3.54	25.26	53.10%
#18-89.CHP	4	30.67	51.60%
#18-9.CHP	2.14	14.05	40.40%
#17-31.CHP	3.05	20.98	42.80%
#17-32.CHP	2.6	25.74	39.20%
#17-33.CHP	2.94	19.44	38.20%
#17-34.CHP	2.49	16.99	41.70%
#17-37.CHP	3.04	20.07	45.40%
#17-38.CHP	2.9	18.09	41.30%
#17-39.CHP	3.1	25.38	40.70%
#17-40.CHP	2.49	22.64	40.30%
#18-13.CHP	2.8	16.53	43.40%
#18-16.CHP	2.4	23.77	46.40%
#18-17.CHP	2.85	23.18	41.70%
#18-18.CHP	13.46	0.44	0.30%

Filename	GAPDH	HSA	%Present
#18-23.CHP	2.41	31.95	34.10%
#18-26.CHP	2.94	21.05	42.30%
#18-27.CHP	2.64	21.08	37.50%
#18-28.CHP	2.71	21.4	40.50%
#18-29.CHP	2.81	35.33	33.60%
#18-30.CHP	2.56	24.8	30.80%
#18-32.CHP	2.34	15.11	41.80%
#18-34.CHP	2.32	22.88	44.10%
#18-37.CHP	3.29	17.78	39.50%
#18-38.CHP	2.93	18.15	43.30%
#18-39.CHP	2.54	21.91	39.90%

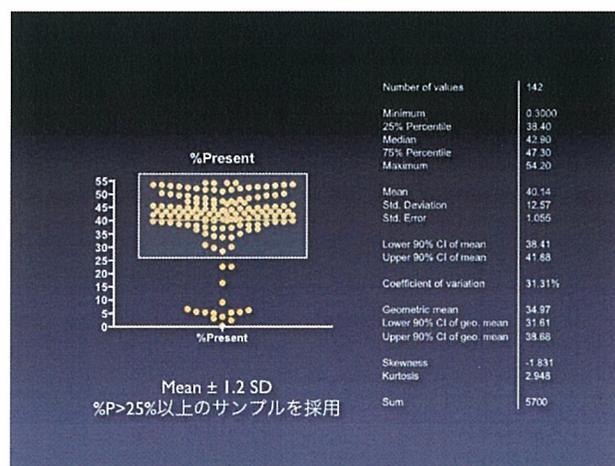


図3 %Present の分布図

%Present の分布図<図3>より 25%未満は不適切な状態のサンプル（または、ラベリング等のハンドリングが不適切であった）と考え、その後の解析から除く事とした。この条件で除かれたサンプルは #18-18, #17-24, #17-19, #17-26, #17-27, #17-21, #17-30, #17-22, #17-28, #17-29, #17-23, #17-17, #18-77, #17-6, #18-12 の 15 サンプルであった。

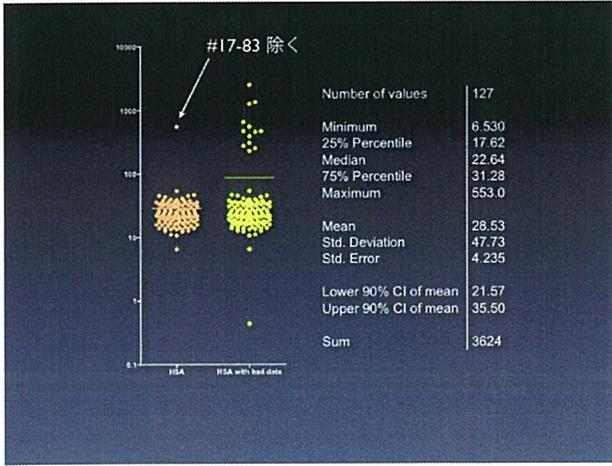


図 4 Actin 3'/5'比の分布図

左は%Present で除かれたサンプルを除いた状態で表現。右は%Present で除かれたサンプルも含む。ここで表現された情報からはずれ値を示した#17-83を除いたく図 4)。



この評価ではサンプルを除かず

図 5 GAPDH 3'/5'比の分布図

は%Present で除かれたサンプルを除いた状態で表現。右は%Present で除かれたサンプルも含む。ここで表現された情報からはずれ値を示した#17-83を除いたく図 5)。

結果以下の 125 サンプルを以降の解析に使用することとしたく表 26)。

表 26 解析に適したサンプルの ID

#17-1	#17-5	#17-80	#18-37	#18-63
#17-10	#17-51	#17-81	#18-38	#18-64
#17-11	#17-53	#17-82	#18-39	#18-65
#17-12	#17-54	#17-84	#18-4	#18-66
#17-13	#17-55	#17-85	#18-40	#18-67
#17-15	#17-56	#17-9	#18-41	#18-68
#17-16	#17-57	#18-1	#18-42	#18-69
#17-2	#17-58	#18-10	#18-43	#18-7
#17-3	#17-59	#18-11	#18-44	#18-70

#17-31	#17-60	#18-12-4	#18-45	#18-71
#17-32	#17-61	#18-13-4	#18-46	#18-73
#17-33	#17-62	#18-14	#18-48.1	#18-74
#17-34	#17-64	#18-15	#18-5	#18-75
#17-37	#17-65	#18-16	#18-50	#18-78
#17-38	#17-68	#18-17	#18-51.1	#18-79
#17-39	#17-69	#18-2	#18-52	#18-8
#17-40	#17-70	#18-23	#18-53	#18-80
#17-41	#17-71	#18-26	#18-55	#18-82
#17-42	#17-72	#18-27	#18-57	#18-83
#17-43	#17-73	#18-28	#18-58	#18-85
#17-44	#17-74	#18-29	#18-59	#18-86
#17-45	#17-75	#18-3	#18-6	#18-87
#17-47	#17-77	#18-30	#18-60	#18-88
#17-48	#17-78	#18-32	#18-61	#18-89
#17-49	#17-8	#18-34	#18-62	#18-9

3) 発現解析-クラスプレディクション

Filter on Volcano Plot でパラメータを以下の通りとした。Function, Parametric test, assumed variance equal, no compensation. Fold difference > x2, p-value cut off < 0.05. このフィルタリングにより 9,351 遺伝子が 4 遺伝子に絞られた。発現量の変化が大きく有意と判断された 4 遺伝子を含む 9,351 遺伝子を図に示すく図 6)。

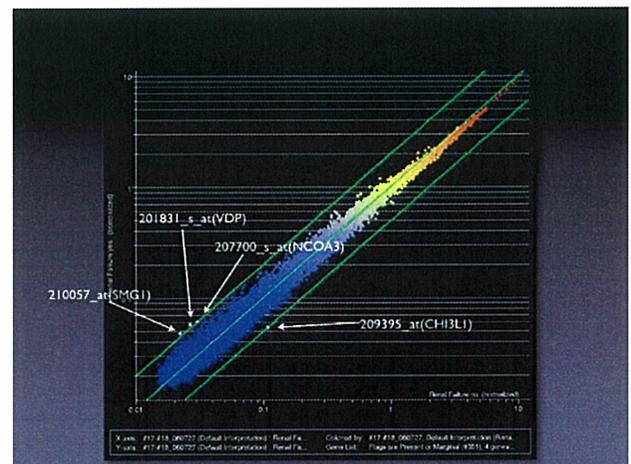


図 6 遺伝子発現量

全 125 サンプルで腎障害有りとなしの間での発現量の相関図であるが、多くの遺伝子発現量は腎障害の有無にかかわらず同様の発現量を示した。横軸が腎障害なし薬物曝露後の遺伝子発現量の平均で、縦軸が腎障害有り薬物曝露後の遺伝子発現量の平均である。フィルタリングにより発現量の変化が大きく有意と判断された 4 遺伝子を含む 9,351 遺伝子を図に示す。4 遺伝子については明示した。

フィルタリングにより絞られた 4 遺伝子は、Gene Symbol (affymetrix probe set ID) で VDP (201831_s_at), NCOA3 (207700_s_at), SMGI (210057_at), CHI3LI (209395_at) であった。腎障害有り (YES) となし (NO) で発現量がどのように異なるのかを次に示した (図 7)。

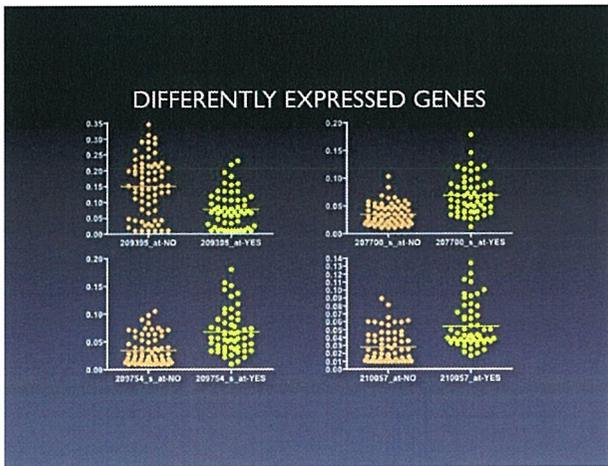


図 7 遺伝子発現量

フィルタリングにより絞られた 4 遺伝子 (Gene Symbol (affymetrix probe set ID) で VDP (201831_s_at), NCOA3 (207700_s_at), SMGI (210057_at), CHI3LI (209395_at)) の発現量の分布を腎障害有り (YES-黄色) となし (NO-オレンジ) で別に示した。

平均値は異なるものの、分布は交差しており、単独の遺伝子発現を見る事により腎障害の有無を予測する事は困難と思われた。これらの 4 遺伝子を用いて、様々なパラメータを与えた場合のクロスバリデーション (125 全サンプルを用いた) により、クラスプレディクションの正確性をラフに評価した (図 8)。

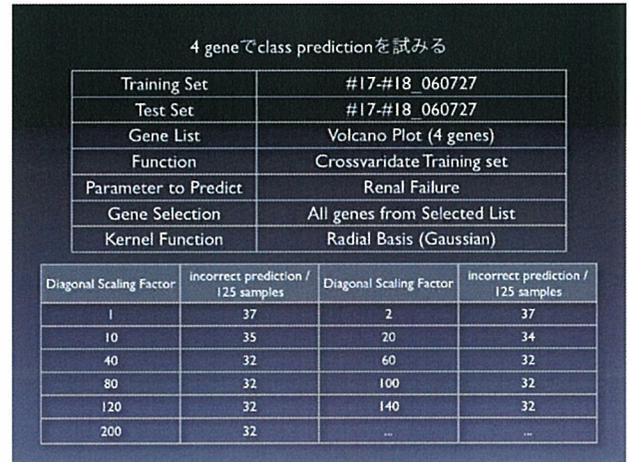


図 8 クラスプレディクション

4 遺伝子を用いてクロスバリデーション (125 全サンプルを用いた) によるクラスプレディクションの正確性の評価を示す。Kernel Function は Radial Basis (Gaussian) が、GeneSpring 埋め込みカーネルの中では最も正確に予測することができるようであった。Diagonal Scaling Factor (soft margin) 値は 40 以上で、74%ほどの正確性が得られた

74%の正確性では充分とは思われなかったのので、プレディクションに用いる遺伝子を増やす目的で、Volcano Plot のパラメータの厳しさを低下させた。すなわち、Filter on Volcano Plot でパラメータを以下の通りとした。Function, Parametric test, assumed variance equal, no compensation. Fold difference > x1.8, p-value cut off < 0.10. このフィルタリングにより 9,351 遺伝子が 20 遺伝子に絞られた。それら 20 遺伝子を用いてクラスプレディクションを行なった (図 9)。

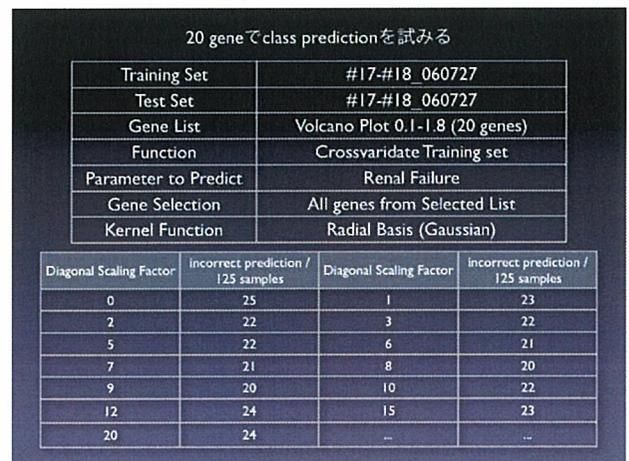


図 9 クラスプレディクション

20 遺伝子を用いてクロスバリデーション (125 全サンプルを用いた) によるクラスプレディクションの正確性の評価を示す。Kernel Function は Radial Basis (Gaussian) が、GeneSpring 埋め込みカーネルの中では最も正確に予測することができるようであった。Diagonal Scaling Factor (soft margin) 値が 8 で、84%ほどの正確性が得られた

さらに予測遺伝子を増やす目的で同じく、Volcano Plotのパラメータの厳しさをさらに低下させた。すなわち、Filter on Volcano Plotでパラメータを以下の通りとした。Function, Parametric test, assumed variance equal, no compensation. Fold difference > x1.6, p-value cut off < 0.05. このフィルタリングにより 9,351 遺伝子が 89 遺伝子に絞られた。それらのうち 50 遺伝子を用いてクラスプレディクションを行なった。89 遺伝子から 50 遺伝子を選択する方法は Golb Method を用いた<図 10>。

50 gene で class prediction を試みる

Training Set	#17-#18_060727
Test Set	#17-#18_060727
Gene List	Volcano Plot 005-1.6 (89 genes)
Function	Crossvaridate Training set
Parameter to Predict	Renal Failure
Gene Selection	50 selected by Golb Method
Kernel Function	Radial Basis (Gaussian)

Diagonal Scaling Factor	incorrect prediction / 125 samples	Diagonal Scaling Factor	incorrect prediction / 125 samples
0	23	1	21
2	23	3	23
4	24	5	25

図 10 クラスプレディクション

50 遺伝子を用いてクロスバリデーション (125 全サンプルを用いた) によるクラスプレディクションの正確性の評価を示す。Kernel Function は Radial Basis (Gaussian) が、GeneSpring 埋め込みカーネルの中では最も正確に予測することができるようであった。Diagonal Scaling Factor (soft margin) 値が 1 のときに 83% ほどの正確性が得られた

次に上記の Function, Parametric test, assumed variance equal, no compensation. Fold difference > x1.6, p-value cut off < 0.05. このフィルタリングにより 9,351 遺伝子が 89 遺伝子に絞られた。それら 89 遺伝子すべてを用いてクラスプレディクションを行なった<図 11>。

89 gene で class prediction を試みる

Training Set	#17-#18_060727
Test Set	#17-#18_060727
Gene List	Volcano Plot 005-1.6 (89 genes)
Function	Crossvaridate Training set
Parameter to Predict	Renal Failure
Gene Selection	All genes from Selected List
Kernel Function	Radial Basis (Gaussian)

Diagonal Scaling Factor	incorrect prediction / 125 samples	Diagonal Scaling Factor	incorrect prediction / 125 samples
0	22	1	19
2	21	3	21
4	21	5	22
10	24	--	--

図 11 クラスプレディクション

89 遺伝子を用いてクロスバリデーション (125 全サンプルを用いた) によるクラスプレディクションの正確性の評価を示す。Kernel Function は Radial Basis (Gaussian) が、GeneSpring 埋め込みカーネルの中では最も正確に予測することができるようであった。Diagonal Scaling Factor (soft margin) 値が 1 のときに 85% ほどの正確性が得られた<図 12>。

予測の正確性

	True Value Yes	True Value No
Prediction Yes	48	11
Prediction No	8	58

Sensitivity = 85.7%
Specificity = 84.1%
Accuracy = 84.8%

definition of sensitivity, specificity and accuracy

	True value Yes	True value No
Prediction Yes	a	b
Prediction No	c	d

sensitivity = a/(a+c) = 1-(pseudo.negativity)
specificity = d/(b+d) = 1-(pseudo.positivity)
accuracy = (a+d)/(a+b+c+d)

図 12 クラスプレディクションの結果

上記 89 遺伝子を用いてクロスバリデーション (125 全サンプルを用いた) によるクラスプレディクションの正確性・感度・特異度の評価を示す。

これらと同様のパラメータを用いて、未知のサンプルの予測の正確性を評価する目的で、2005 年度のサンプルでパラメータを微調整した。Volcano Plot によるフィルターリングは、Function, Parametric test, assumed variance equal, no compensation. Fold difference > x1.6, p-value cut off < 0.05 の条件で行なった。32 遺伝子が選択された。この 32 遺伝子を次の表に示す<表 27>。その上で 2006 年度のサンプルの予測を行なった。

表 27 選択された 32 遺伝子

probe set ID	gene symbol	description
218717_s_at	LEPREL1	Leprecan-like 1
221731_x_at	CSPG2	Chondroitin sulfate proteoglycan 2 (versican)
215073_s_at	NR2F2	Nuclear receptor subfamily 2, group F, member 2
204533_at	CXCL10	Chemokine (C-X-C motif) ligand 10
214036_at		CDNA: FLJ22256 fis, clone HRC02860
219777_at	hIAN2	Human immune associated nucleotide 2
215118_s_at	MGC27165	Hypothetical protein MGC27165
210105_s_at	FYN	FYN oncogene related to SRC, FGR, YES
220177_s_at	TMPRSS3	Transmembrane protease, serine 3
203430_at	HEBP2	Heme binding protein 2
219938_s_at	PSTPIP2	Proline-serine-threonine phosphatase interacting protein 2
202436_s_at	CYP1B1	Cytochrome P450, family 1, subfamily B, polypeptide 1
201131_s_at	CDH1	Cadherin 1, type 1, E-cadherin (epithelial)
215051_x_at	AIF1	Allograft inflammatory factor 1
209191_at	MGC4083	Tubulin beta MGC4083
209396_s_at	CHI3L1	Chitinase 3-like 1 (cartilage glycoprotein-39)
217771_at	GOLPH2	Golgi phosphoprotein 2
218764_at	PRKCH	Homo sapiens protein kinase C, eta (PRKCH), mRNA.
212698_s_at	SEPT10	Septin 10
205933_at	SETBP1	SET binding protein 1

probe set ID	gene symbol	description
211367_s_at	CASP1	Caspase 1, apoptosis-related cysteine protease (interleukin 1, beta, convertase)
205590_at	RASGRP1	RAS guanyl releasing protein 1 (calcium and DAG-regulated)
221584_s_at	KCNMA1	Potassium large conductance calcium-activated channel, subfamily M, alpha member 1
219410_at	FLJ10134	Hypothetical protein FLJ10134
207111_at	EMR1	Egf-like module containing, mucin-like, hormone receptor-like 1
209118_s_at	TUBA3	Tubulin, alpha 3
216557_x_at	A1VH3	Human rearranged immunoglobulin heavy chain (A1VH3) gene, partial cds.
203038_at	PTPRK	Protein tyrosine phosphatase, receptor type, K
205848_at	GAS2	Growth arrest-specific 2
210448_s_at	P2RX5	Purinergic receptor P2X, ligand-gated ion channel, 5
202552_s_at	CRIM1	Cysteine-rich motor neuron 1