

トキシコジェノミクス・プロジェクトにより実施された毒性試験における  
病理組織学的検査所見についての検証

研究分担者 三森国敏

東京農工大学大学院 教授

研究要旨：73 化学物質のラットの毒性試験における肝臓と腎臓のヘマトキシリン・エオジン（HE）染色標本から診断された病理組織所見およびその組織写真 640 件について検証を実施し、診断された所見に対して用語の統一や整合性をとる作業を実施した。また、肝臓について脂肪染色を施し、HE 染色標本で観察された空胞変性との関連性を検索した。

#### A. 研究目的

トキシコジェノミクス・インフォマティクスプロジェクトにより実施された 73 化学物質のラットにおける毒性試験における肝臓と腎臓のヘマトキシリン・エオジン（HE）染色標本から診断された病理組織所見およびその組織写真 640 件についてレビューを行い、組織写真で示される変化とそれに対して取られた所見の妥当性についての確認を行った。また、HE 染色標本で、脂肪変性ないし空胞変性と診断されたものについて脂肪染色を施し、その本態が何であるかを解析した。

#### B. 研究方法

1) HE 染色標本についての検証：各化学物質により誘発された肝臓と腎臓における毒性病変の組織診断とその組織所見が撮影された組織写真 640 件をレビューし、診断名の適切性および写真との整合性を調査した。また、研究成果報告書に記載された所見について、同じ病変について同一用語を用いて診断しているかについても調査した。組織写真から病理診断が適切になされているか否か判定が難しいものについては、その HE 染色標本を取り寄せ、顕微鏡を用いて、その診断が適切であるか否かを確認した。さらに、HE 染色標本のみでは、診断困難なものについては、パラフィンブロックを取り寄せ、組織切片を作成して、種々の抗体を用いて免疫組織化学的解析を行った。

2) 脂肪染色による解析：HE 染色標本で、脂肪変性ないし空胞変性と診断された以下の化学物質の毒性試験については、ズダンブラック B とオイルレッド O 染色を実施してその本態が何であるかを検索した。

diltiazem、thioacetamide、methapyrilene、amiodarone、imipramine、amitriptyline、clomipramine、ethambutol、acetamidofluorene、lomustine、ketoconazole、chlormezanone、ethionine、naphthyl isothiocyanate

（倫理面への配慮）

特になし

#### C. 研究結果

1) HE 染色標本についての検証：組織写真 640 件をレビューした結果、以下のような不備が認められた。

- ① 毒性試験を実施した試験実施施設間で、用語の統一がされていない。そのため、同じ病変にもかかわらず、いろいろな用語が使用されており、これらの用語の統一を実施した。
- ② 肝臓については、肝細胞肥大を所見として取った上で、小胞体増生の特徴である「ground glass appearance（すりガラス状変化）」を所見として付記している試験実施施設とそうでない施設があり、所見を特定化する作業を実施した。さらに、Fatty（脂肪）やHydropic（水腫性）に細区分した試験実施施設と、統一的に「Degeneration, vacuolar（空胞変性）」を使用する施設があり、「Degeneration, vacuolar」に用語統一する作業を行った。「Change, eosinophilic」と「Degeneration, granular, eosinophilic（好酸性顆粒状変性）」の区別があいまいであり、鑑別基準をどこに置いたのかが不明な試験実施施設があった。

- ③ 腎臓については、病変が発現している部位がどこであるのか(どの部位の尿細管なのか?) 明記されていないものが多かった。例えば、「renal tubule」のみ記載されているが、写真を見ると、「Proximal tubule, convoluted」であることがわかるものが多々あり、このような所見のつけ方では、病変発現部位を特定することができず、毒性の特徴を考える上で不十分な記載を行っている試験実施施設があった。このような病変発現部位を特定していない写真については、病変部位を特定した。また、尿細管の

「regeneration(再生)」を「degeneration(変性)」や「necrosis(壊死)」と誤診しているものが目立った。その他、「hyaline droplet(硝子滴)」と「eosinophilic body(好酸性小体)」の区別があまりであり、鑑別基準をどこに置いたのかが不明な試験実施施設があり、それらの用語統一も行った。また、尿細管の「swelling(肥大:肉眼所見用語)」の表記が多く、「hypertrophy(肥大:組織所見用語)」に用語統一した。

- ④ 添付されている組織写真には、組織所見に記載されている所見以外に、さらに他の所見がみられているものがあり、整合性に欠けるものがあった。また、組織写真の質が悪く、認められた所見がうまく撮影されていないものがみかけられた。さらに、組織写真のみでは、記載されている病変の特徴をつかむことができないものがあった。そのような写真については、HE切片を取り寄せて、病変を再度確認し、必要に応じて組織写真の取り直しを実施した。

2) 脂肪染色による解析: 脂肪染色の結果、HE 染色で空胞変性と診断された diltiazem、thioacetamide、methapyrilene、amiodarone、imipramine、amitriptyline、clomipramine は、ズダンブラック B とオイルレッド O の両方に陽性を示した。HE 染色で空胞変性と診断された ethambutol、acetamidofluorene は、ズダンブラック B に陽性、オイルレッド O に陰性を示した。HE 染色で空胞変性と診断された lomustine は、ズダンブラック B とオイルレッド O の両方に陰性であった。

HE 染色標本で空胞変性を認めなかった ketoconazole は、ズダンブラック B とオイルレッド O の両方に陽性を示した。HE 染色標本で空胞変性を認めなかった chlormezanone、ethionine、

naphthyl isothiocyanate は、ズダンブラック B に陽性、オイルレッド O に陰性であった。

#### D. 考察

以上のように、毒性試験を実施した試験実施施設間で、用語の統一がなされていないことが今回のレビューで明確となり、誘発された毒性病変の質が同一にも関わらず、異なった用語が使用されていたことについては、非常に驚くべき事象であった。日本毒性病理学会が「毒性病理専門家」という認定制度を設置しているが、それらの専門家が共通した病理診断用語を使うべきであることが再確認された。

化学物質により誘発された病変の内、数種類の所見が混在する場合があるが、試験実施施設により、それらの所見を全て記載する実施施設と、それらの所見を総括して一つの診断名にする施設が見受けられ、これらについても整合性をもたせるべきであると思われた。特に、肝細胞肥大と「ground glass appearance」は、小胞体が増生する場合に頻繁に見られる変化であるが、肝細胞肥大は、いろいろな細胞内小器官が増加した場合に見られる変化であるので、肝細胞肥大のみ所見として取るべきではなく、「ground glass appearance」を付記すべきである。

今回のレビューで驚いた点は、HE 染色標本での病変の見逃しが多々認められたことである。貼付されている組織写真をみた場合、所見として記載されている病変に加えて、他の所見が認められており、明らかに鏡検担当者の所見の見逃しがあったことが否定できなかった。このような不備については、試験実施施設の鏡検担当者の資質が問われるものであり、今後、このような不備が起らないような防止策が必要である。

以上のようなレビュー結果から、トキシコジェノミクスデータベース内に格納されている病理所見を横断検索して活用する場合は、病理学を理解した専門家により、検索したい病変の同義語も含めて検索してもらう必要がある。また、化学物質により誘発された病変を試験実施施設の鏡検担当者が見逃している場合があるので、トキシコジェノミクスデータベース内に格納されている病理所見のみから横断検索すると、重要な誘発病変の記載が欠落していることも考えられるので、十分な注意が必要である。このようなことから、今回のレビュー結果は、トキシコジェノミクス・プロジェクトのデータ集として発

刊される「(仮称) 毒性試験データ集」に、全ての検索化学物質において誘発された病変の組織写真と一緒に、それらの誘発病変の病理用語を統一した正誤表的なリストを添付するとのことであるので、そのリストを参照されたい。

脂肪染色結果では、HE 染色で空胞変性と診断されたものの中には中性脂肪やリン脂質であるものが含まれていることが明らかとなった。しかし、HE 染色からそれらを鑑別することは必ずしも容易ではないことが確認された。さらに、HE 染色では空胞変性と診断されなかった化学物質において、ズダン黒に陽性のものが認められたことから、HE 染色のみでは、脂質の存在を確認することは困難であることが明確となり、遺伝子解析で脂質代謝に関連する遺伝子の変動を考慮した上で、脂質についての特殊染色を実施すべきであることが強く示唆された。

#### E. 結論

73 化学物質のラットの毒性試験での肝臓と腎臓の HE 染色標本から診断された病理組織所見とその組織写真 640 件についてレビューを実施した結果、用語の統一が不十分であること、誘発された病変に対して適切な診断用語が使われていないこと、誘発された病変を見逃していることなどが発見され、大幅な用語の統一や所見の整合性を計る作業を実施した。また、肝臓について脂肪染色を施し、HE 染色標本で観察された空胞変性との関連性を検索した。その結果、空胞の内容が中性脂肪やリン脂質である場合と単なる空胞であるものが区別され、HE 染色のみの検索の限界が再確認された。

#### F. 健康危険情報

なし

#### G. 研究発表

##### 1. 論文発表

###### 1) 書籍

なし

###### 2) 雑誌

なし

##### 2. 学会発表

なし

#### H. 知的財産所有権の出願・登録状況

##### 1. 特許取得

なし

##### 2. 実用新案登録

なし

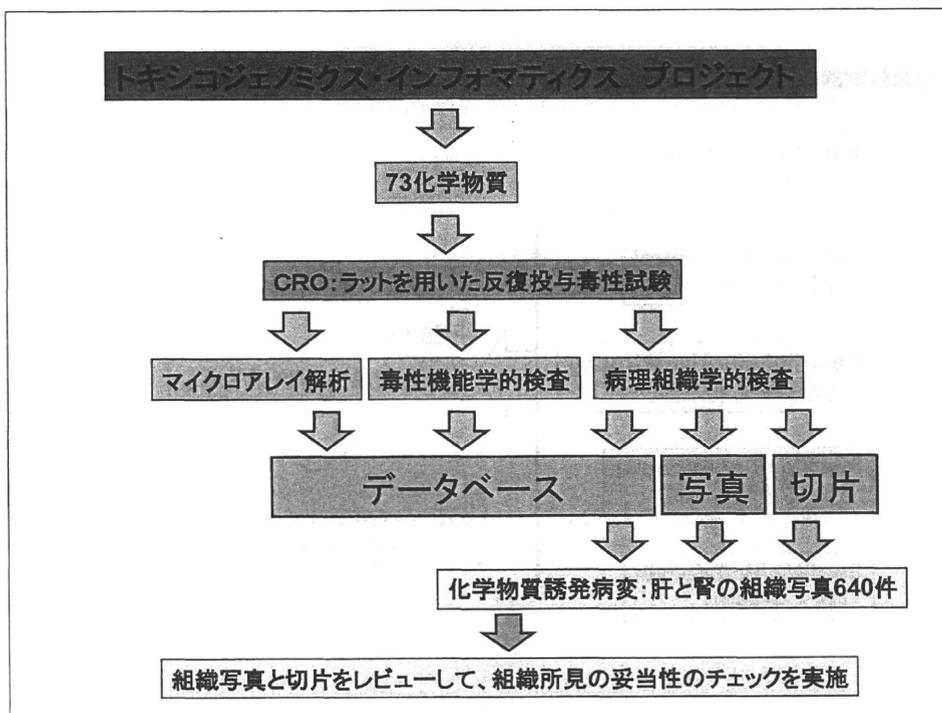
##### 3. その他

なし

# トキシコジェノミクスデータベースを活用した 医薬品安全性評価における 病理所見活用上の問題点

三森国敏

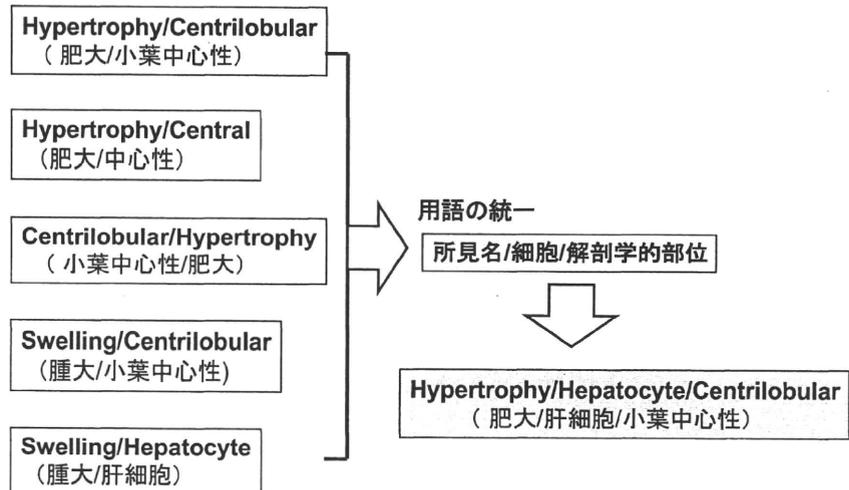
東京農工大学大学院農学研究院  
動物生命科学部門

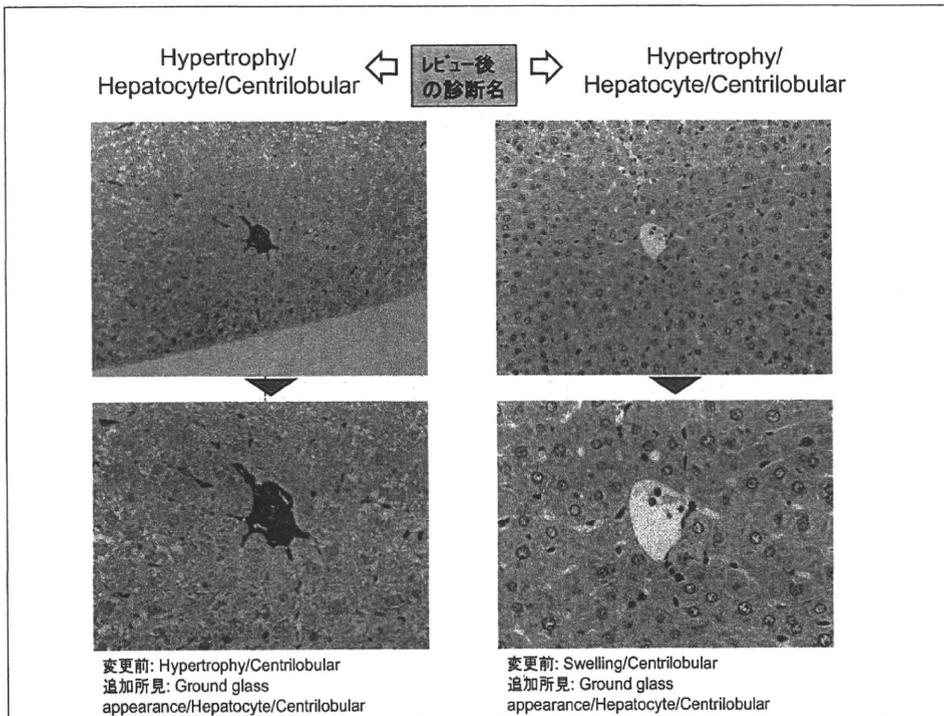


**組織写真640件をレビューした結果、以下のような不具合が認められた。**

- 1) 毒性試験を実施した試験実施施設間で、用語の統一がされていなかった。
- 2) 肝臓:
  - ① 肝細胞肥大を所見として取った上で、「ground glass appearance(すりガラス状変化)」を所見として付記している試験実施施設とそうでない施設があった。
  - ② 空胞変性の病変に対して種々の診断用語が使用されていた。
  - ③ 「Change, eosinophilic」と「Degeneration, granular, eosinophilic(好酸性顆粒状変性)」の区別があいまいであり、鑑別基準をどこに置いたのかが不明な施設があった。
- 3) 腎臓:
  - ① 病変が発現している部位がどこであるのか明記されていないものが多かった。
  - ② 尿細管の「Regeneration(再生)」を「Degeneration(変性)」や「Necrosis(壊死)」と誤診しているものが目立った。
  - ③ 「Hyaline droplet(硝子滴)」と「Eosinophilic body(好酸性小体)」の区別があいまいであり、鑑別基準をどこに置いたのかが不明な試験実施施設があった。
  - ④ 尿細管の「Swelling(肥大:肉眼所見用語)」の表記が多く、「Hypertrophy(肥大:組織所見用語)」に用語統一した。
- 4) 添付されている組織写真には、組織所見に記載されている所見以外に、さらに他の所見がみられているものがあり、整合性に欠けるものがあった。
- 5) 組織写真の質が悪く、認められた所見がうまく撮影されていないものがあった。
- 6) 組織写真のみでは、記載されている病変の特徴をつかむことができないものがあった。

**1) 毒性試験を実施した試験実施施設間で、用語の統一がされていなかった。**





肝: ① 肝細胞肥大を所見として記録した上で、「ground glass appearance(すりガラス状変化)」を所見として付記している試験実施施設とそうでない施設があった。

**CRO-A: Hypertrophy/Centrilobular; Ground glass appearance/Centrilobular**

**CRO-B: Hypertrophy/Centrilobular**  
(Ground glass appearanceが存在していても所見として記録していない。)

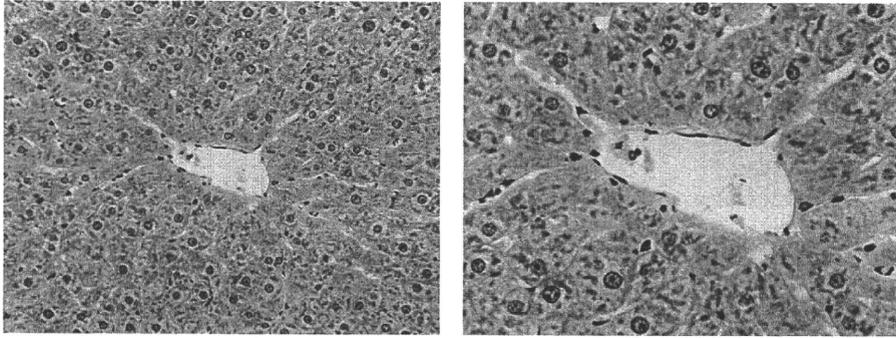
**CRO-C: Ground glass appearance/Centrilobular**  
(Hypertrophyが存在していても所見として記録していない。)

用語の統一



**Hypertrophy/Hepatocyte/Centrilobular;  
Ground glass appearance/Hepatocyte/Centrilobular**

Hypertrophy/Hepatocyte/Centrilobular;  
Ground glass appearance/Hepatocyte/Centrilobular



変更前: Hypertrophy/Centrilobular  
追加所見: なし

肝: ②空胞変性の病変に対して種々の診断用語が使用されていた。

CRO-A: Vacuolization, cytoplasmic

CRO-B: Deposit, lipid

CRO-C: Vacuolar change of hepatocyte

CRO-D: Vacuolization, cytoplasmic/Hepatocyte

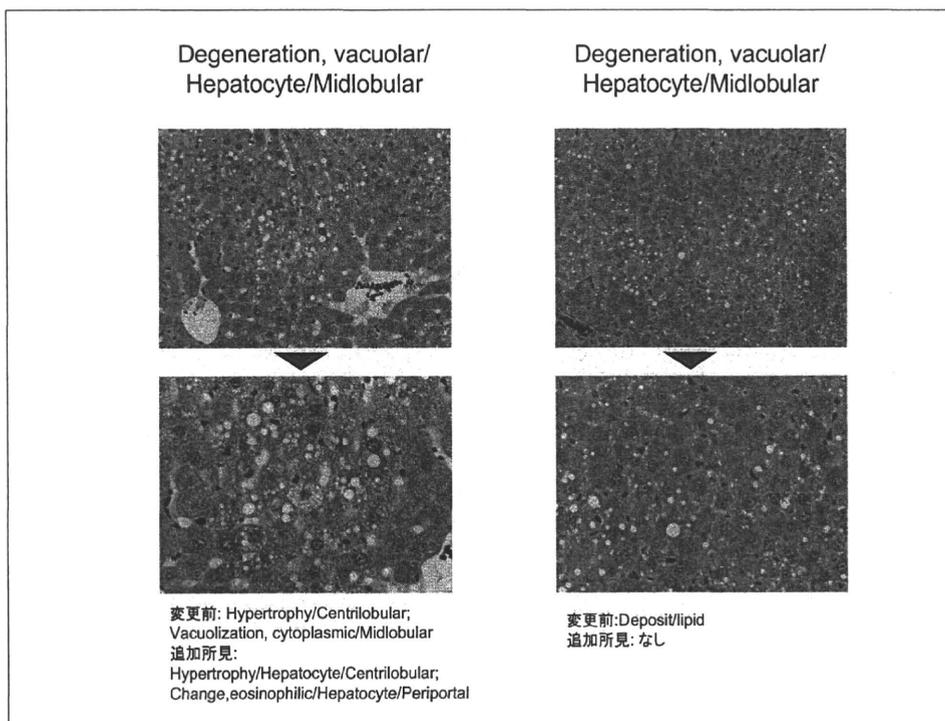
用語の統一

Degeneration, vacuolar/Hepatocyte

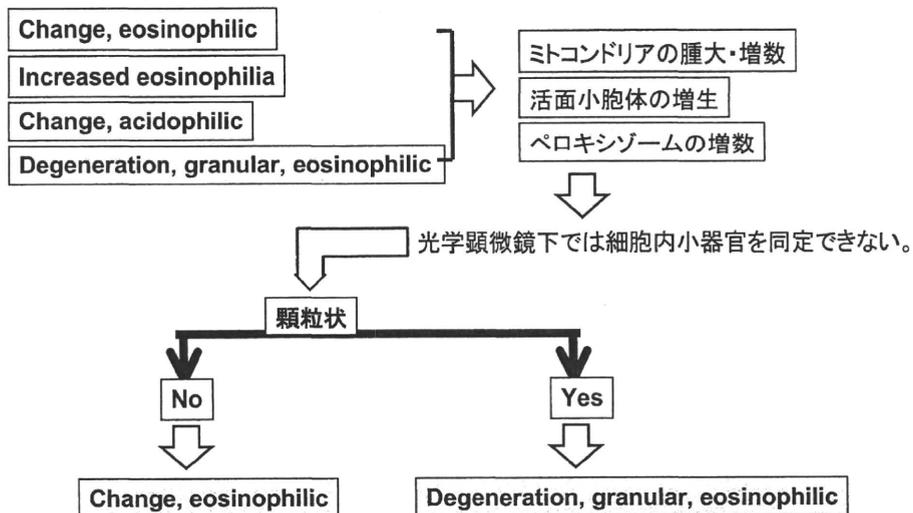
CRO: Fatty degeneration/Hepatocyte

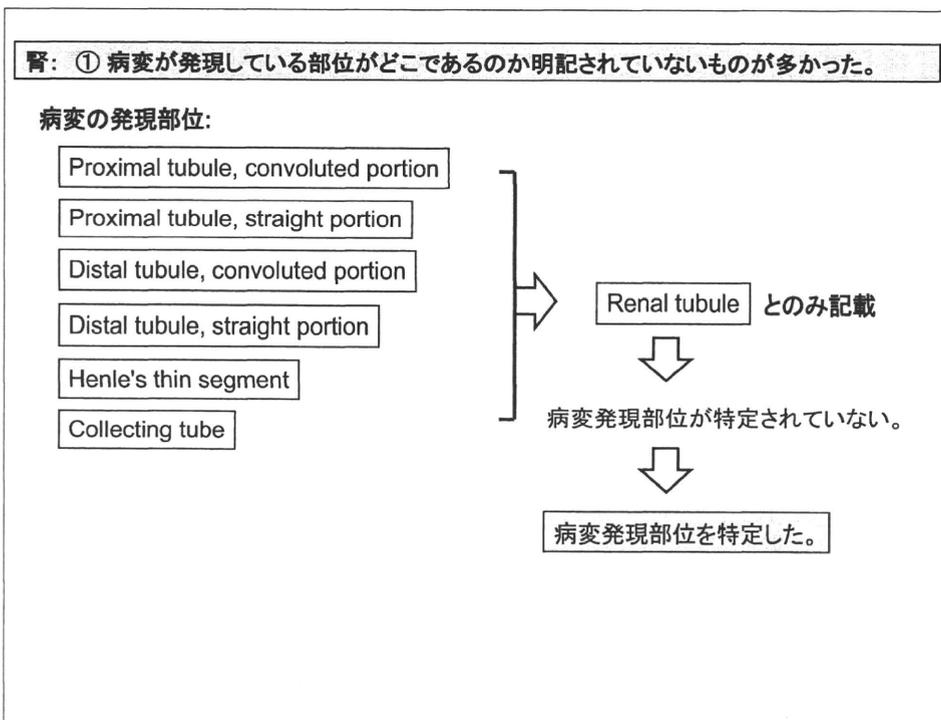
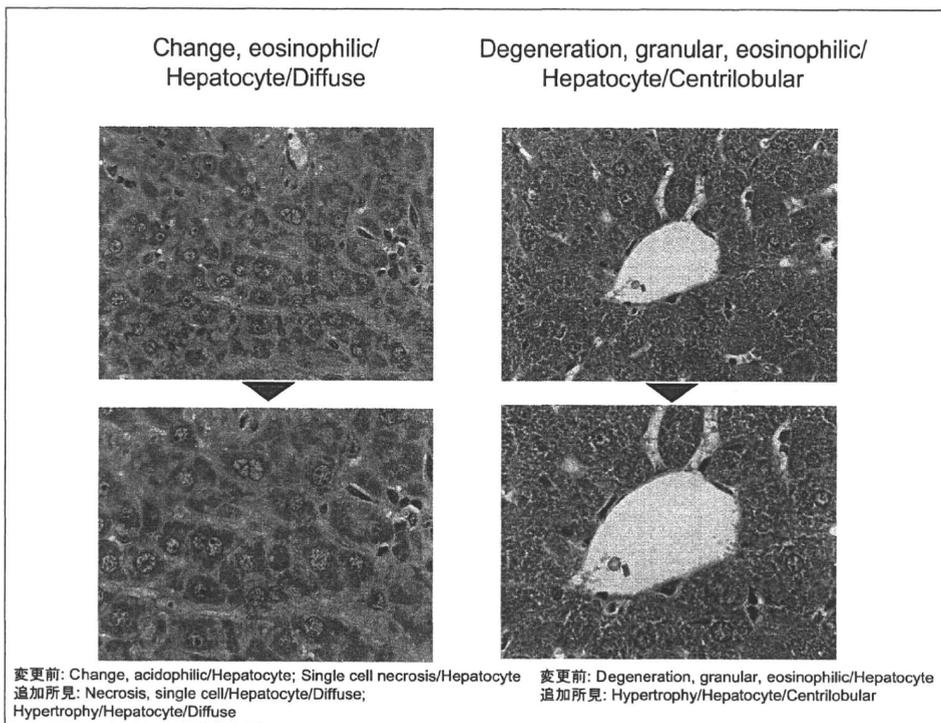
脂肪染色を実施していないにもかかわらず、「脂肪変性」という用語を使用している施設があった。

⇒ 脂肪染色を実施 ⇒ 脂肪の存在を確認

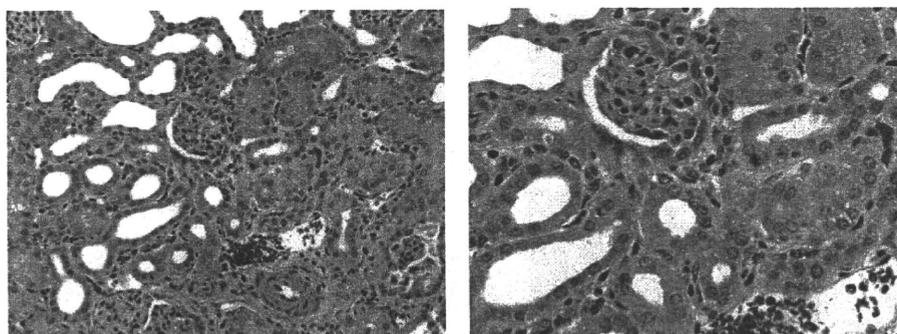


肝: ③「Change, eosinophilic」と「Degeneration, granular, eosinophilic(好酸性顆粒状変性)」の区別があいまいであり、鑑別基準をどこに置いたのが不明な施設があった。





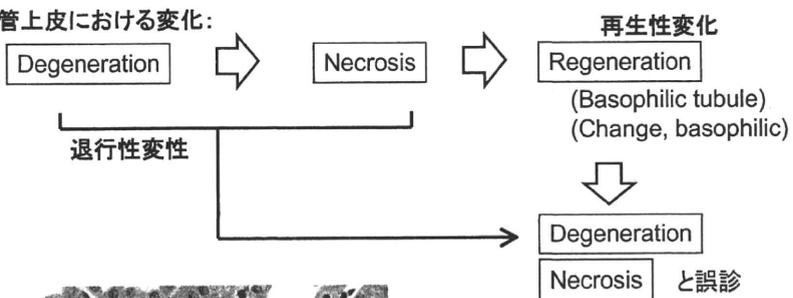
Regeneration/Proximal tubule, convoluted



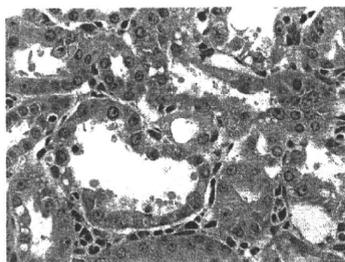
変更前: Regeneration/Renal tubule  
追加所見: なし

腎: ② 尿細管の「Regeneration(再生)」を「Degeneration(変性)」や「Necrosis(壊死)」と誤診しているものが目立った。

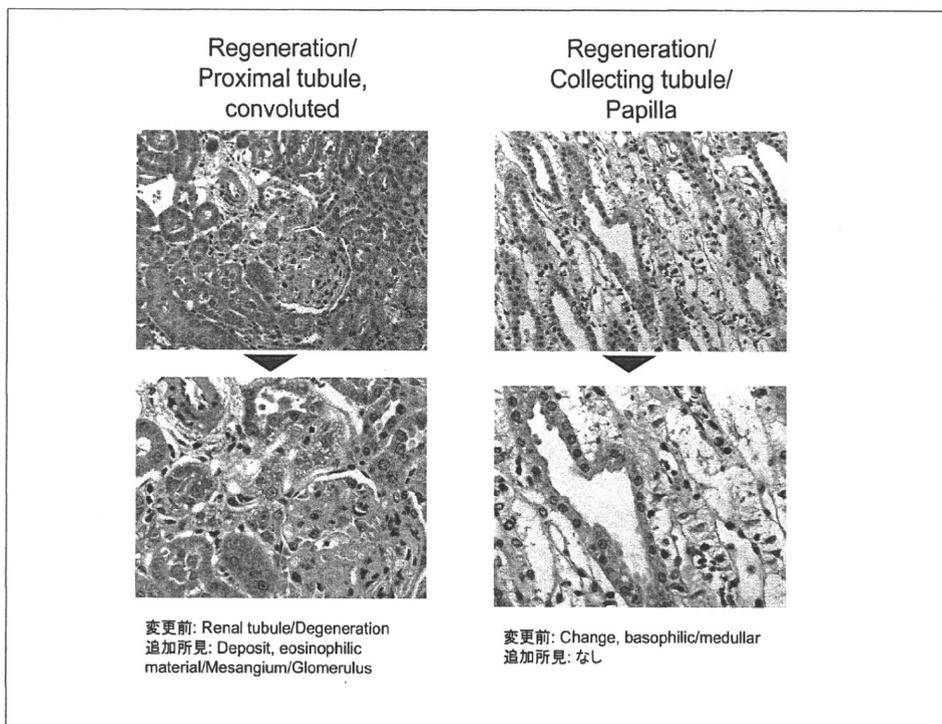
尿細管上皮における変化:



変更前:  
Necrosis/  
Renal  
tubule



変更後:  
Regeneration/Proximal  
tubule, convoluted;  
Necrosis/Proximal tubule,  
convoluted



**腎: ③ 「Hyaline droplet(硝子滴)」と「Eosinophilic body(好酸性小体)」の区別があいまいであり、鑑別基準をどこに置いたのが不明な試験実施施設があった。**

**Hyaline droplet** 細胞質におけるエオジン好性の顆粒であり、PASに陽性、α2uグロブリン陽性のもの。

**Eosinophilic body** エオジン好性のサイズが硝子滴よりも大きい細胞質内の小体で、α2uグロブリン陽性に染まる場合とそうでないものの2種類がある。

-----

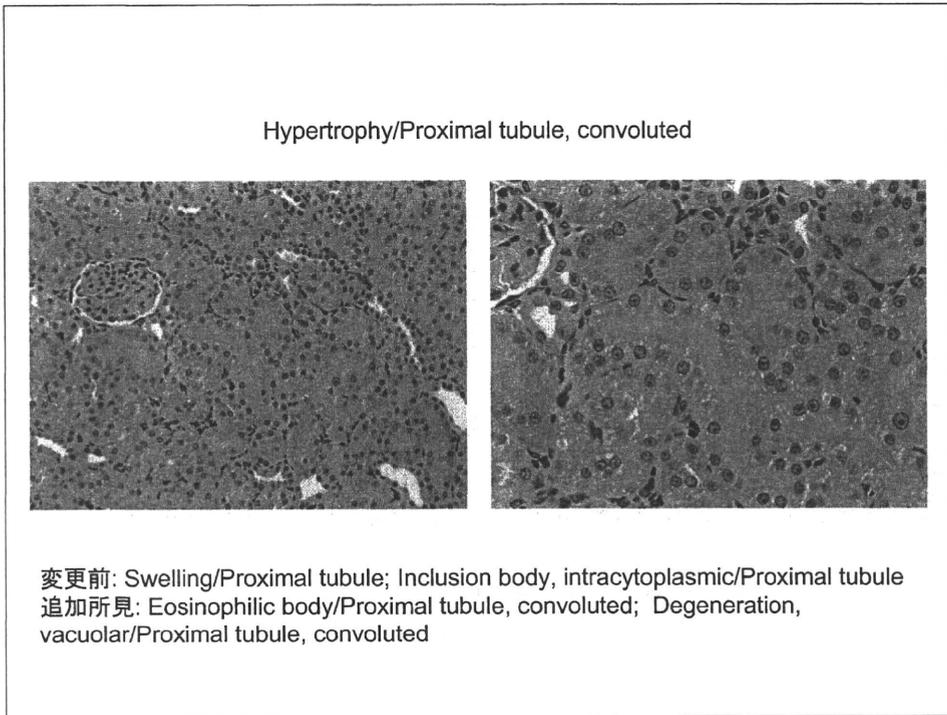
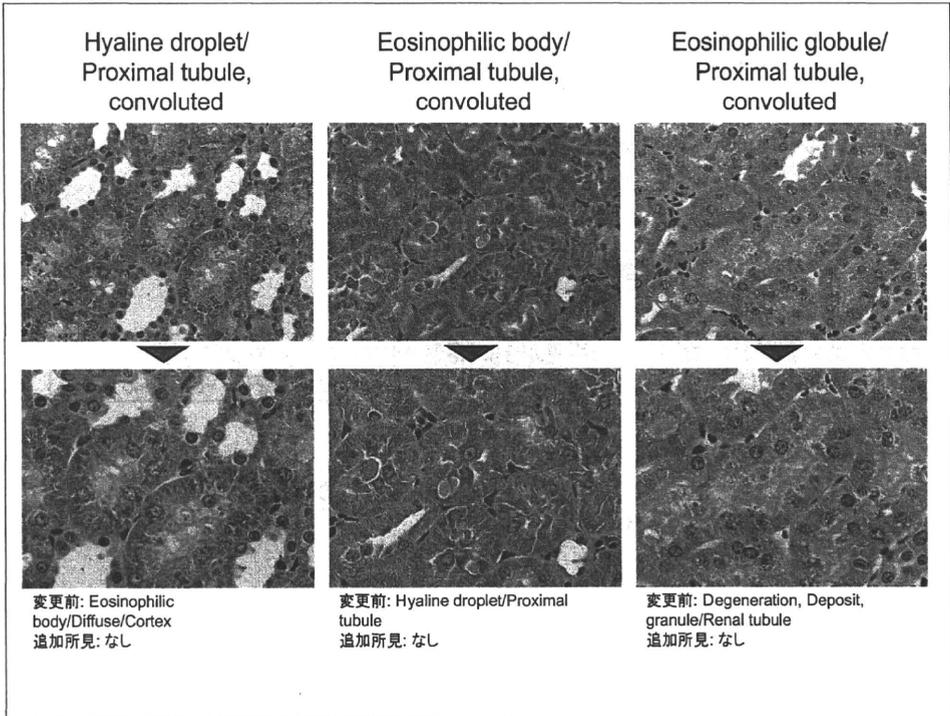
**Granular deposition** 上記の二つとは異なる、エオジン好性の細胞質内微細小球であり、「hyaline droplet」とは異なるもの。

↳ **Eosinophilic globule** に用語統一

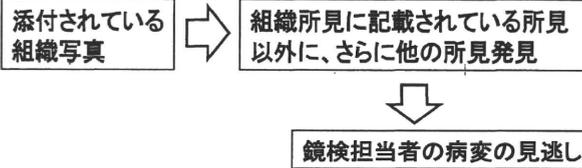
**腎: ④ 尿細管の「Swelling」の表記が多く、「Hypertrophy」に用語統一した。**

尿細管上皮の肥大

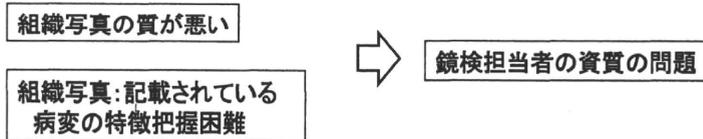
**Swelling** (肥大: 肉眼所見用語)      ↳      **Hypertrophy** (肥大: 組織所見用語)



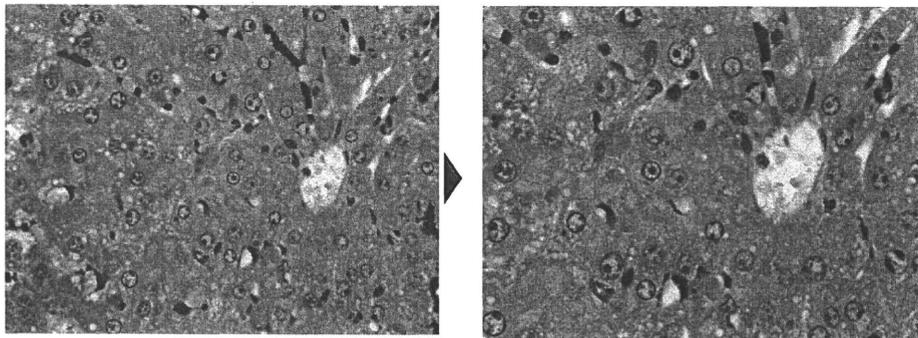
4) 添付されている組織写真には、組織所見に記載されている所見以外に、さらに他の所見がみられているものがあり、整合性に欠けるものがあった。



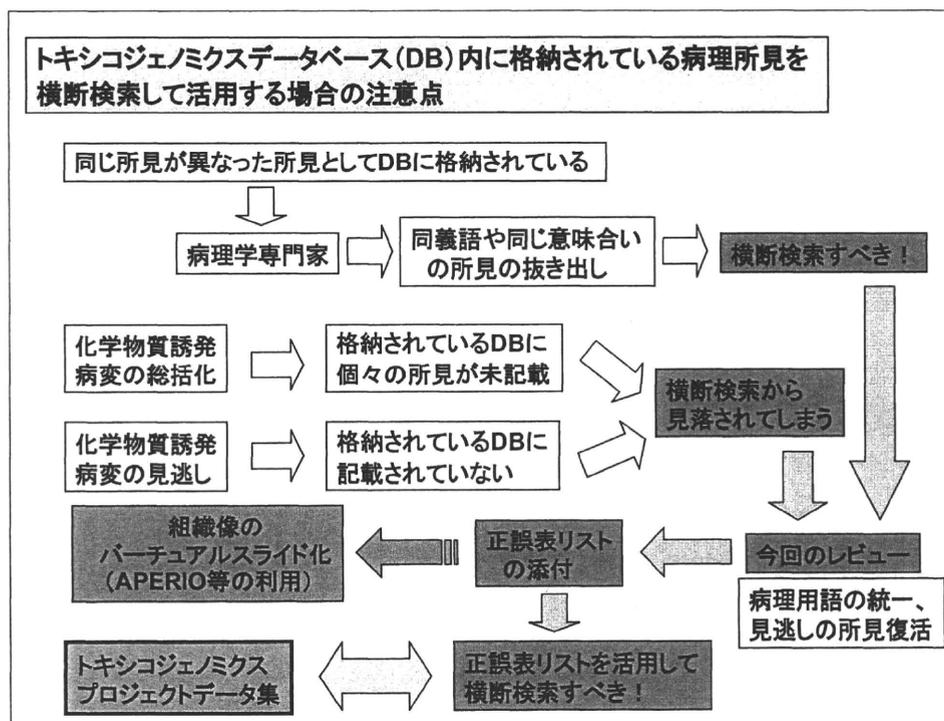
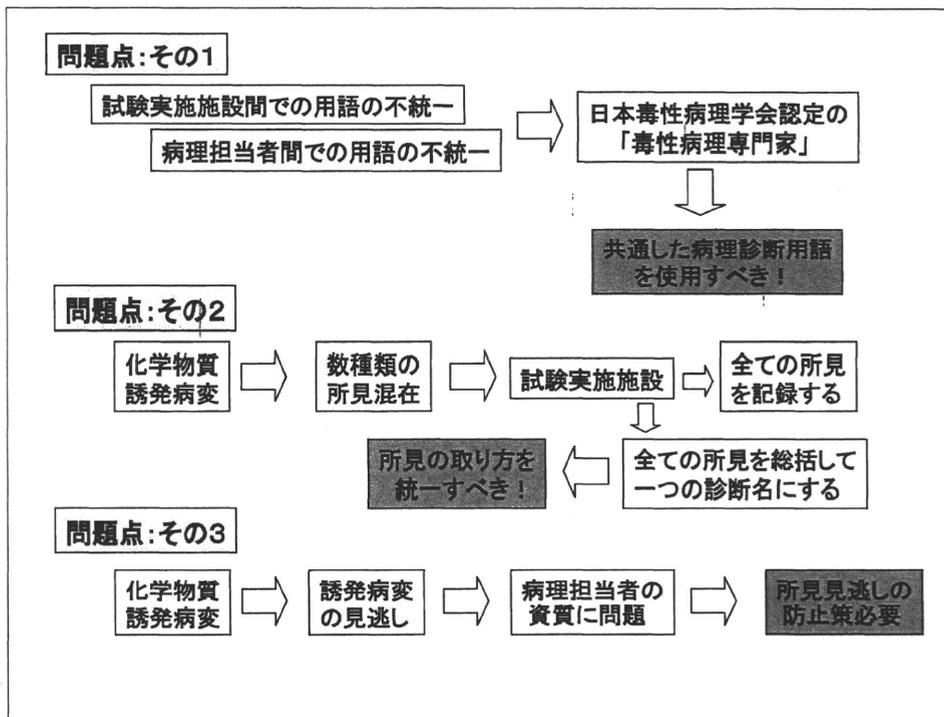
5) 組織写真の質が悪く、認められた所見がうまく撮影されていないものがあった。  
6) 組織写真のみでは、記載されている病変の特徴をつかむことができないものがあった。



#### Anisokaryosis/ Hepatocyte/Centrilobular



変更前: Vacuolization, cytoplasmic/Centrilobular  
追加所見: Degeneration, vacuolar/Hepatocyte/Centrilobular



## 別添5

## 研究成果の刊行に関する一覧表

## 書籍

著者氏名	論文タイトル名	書籍全体の編集者名	書籍名	出版社名	出版地	出版年	ページ
Yasushi Okuno, Yohsuke Minowa, Hiroshi Yamada, Yasuo Ohno and Tetsuro Urushidani.	In Silico Toxicology Prediction Using Toxicogenomics Data.	Daniel A. Casciano, Saura C. Sahu	Handbook of Systems Toxicology	John Wiley & Sons	USA	2011	591-598
漆谷徹郎	毒性評価 トキシコゲノミクス	金子周一、堀池靖浩	バイオチップ実用化ハンドブック	NTS	東京	2010	268-274

## 雑誌

発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
Sumida K, Igarashi Y, Toritsuka N, Matsushita T, Abe-Tomizawa K, Aoki M, Urushidani T, Yamada H, Ohno Y.	Effects of DMSO on gene expression in human and rat hepatocytes.	Hum. Exp. Toxicol.		Feb 21. [Epub ahead of print]	2011
Weihua Gao, Yumiko Mizukawa, Noriyuki Nakatsu, Yosuke Minowa, Hiroshi Yamada, Yasuo Ohno and Tetsuro Urushidani.	Mechanism-based biomarker gene sets for glutathione depletion-related hepatotoxicity in rats.	J. Toxicol. Appl. Pharmacol	247(3)	211-221	2010
Uehara T, Ono A, Maruyama T, Kato I, Yamada H, Ohno Y, Urushidani T.	The Japanese toxicogenomics project: Application of toxicogenomics.	Mol. Nutr. Food Res.	54(2)	218-227	2010

漆谷 徹郎	トランスレーショナル リサーチ④ トキシコ ゲノミクスプロジェク トと安全性試験	日本薬理学会 雑誌	136 (1)	46-49	2010
Suzuki A, Igarashi K, Aisaki KI, Kanno J, Saga Y.	NANOS2 interacts with the CCR4-NOT deadenylation complex and leads to suppression of specific RNAs.	Proc Natl Acad Sci U S A.	107(8)	3594-3599	2010
Arase S, Ishii K, Igarashi K, Aisaki K, Yoshio Y, Matsushima A, Shimohigashi Y, Arima K, Kanno J, Sugimura Y.	Endocrine Disrupter Bisphenol A Increases In Situ Estrogen Production in the Mouse Urogenital Sinus.	Biol Reprod.	84 (4)	734-742	2011
Yoshida T, Sekine T, Aisaki K, Mikami T, Kanno J, Okayasu I.	CITED2 is activated in ulcerative colitis and induces p53-dependent apoptosis in response to butyric acid.	J Gastroente rol.	46(3)	339-349	2011

# *In Silico* Toxicology Prediction Using Toxicogenomics Data

Yasushi Okuno<sup>1</sup>, Yohsuke Minowa<sup>2</sup>, Hiroshi Yamada<sup>2</sup>, Yasuo Ohno<sup>2</sup> and Tetsuro Urushidani<sup>2,3</sup>

<sup>1</sup>Department of Systems Bioscience for Drug Discovery, Graduate School of Pharmaceutical Sciences, Kyoto University, Kyoto, Japan, <sup>2</sup>Toxicogenomics-Informatics Project, National Institute of Biomedical Innovation, Ibaraki, Osaka, Japan and <sup>3</sup>Department of Pathophysiology, Faculty of Pharmaceutical Sciences, Doshisha Women's College of Liberal Arts, Kodo, Kyoto, Japan

## 1 INTRODUCTION

Toxicogenomics holds the promise of unprecedented advances in two broad, overlapping fields: mechanistic or investigative toxicology, and predictive toxicology. Mechanistic or investigative toxicology is the area of toxicology focused on biological responses to drug toxicity, and provides important perspectives on risk assessments of various compounds. In contrast, predictive toxicology focuses on identifying the potential toxicity of compounds (Cunningham and Lehman-McKeeman, 2005; Boverhof and Zacharewski, 2006). In previous studies, toxicogenomics has contributed to in-depth investigations of molecular mechanisms and the modes of toxin/chemical/environmental stressor action that was difficult to be achieved by conventional toxicological approaches. At the same time, expanding toxicogenomic data has promoted a valuable platform for the establishment of biomarkers to predict a compound's toxicity.

The progress of toxicogenomics has been supported by DNA microarray technology, a powerful tool for directly monitoring patterns of cellular perturbations through the identification and quantification of global shifts in gene expression

resulting from pathological alterations within cells and tissues. Microarrays provide a large amount of transcriptional expression data for thousands of individual genes under various experimental conditions. Bioinformatics technologies can determine which genes are meaningful, facilitating the analysis of huge pools of toxicogenomics data in mechanistic and predictive toxicology. Meaningful genes are referred to as "signature genes" with characteristic gene expression profiles for exposure or toxicological response to specific classes of toxic compounds. This chapter is devoted to computational approaches for the data mining of biomarker genes from toxicogenomics data, leading to toxicity prediction.

## 2 MICROARRAY DATA FOR TOXICOGENOMICS

Microarray techniques have been applied not only to toxicogenomics but also to various clinical purposes, such as disease classification and patient diagnostics. These applications have revealed the challenges of achieving reproducibility and stability of outcomes with microarray technologies caused by cross-platform chips and differing experimental

conditions. Therefore, data sets for constructing reliable prediction models must be obtained with the same platform and experimental conditions. For this purpose, many public databases for toxicologically relevant microarray data have been developed, including the Comparative Toxicogenomics Database (CTD) (Mattingly *et al.*, 2006), Environment, Drugs and Gene Expression database (EDGE) (Hayes *et al.*, 2005), and Chemical Effects in Biological System (CEBS) knowledgebase (Waters *et al.*, 2003), among others.

In 2002, we began the toxicogenomics project (TGP), a public-private collaborative project of the National Institute of Health Sciences, the National Institute of Biomedical Innovation, and 15 pharmaceutical companies in Japan (Urushidani and Nagao, 2005). With an emphasis on the uniformity of data quality, TGP has generated a large-scale toxicology database of transcriptomes intended to predict the toxicity of new chemical entities in the early stages of drug development. Drug reactions such as efficacy and toxicity are associated with the dosage and time course after treatment, and so precisely monitoring drug reactions requires multiple dose- and time-dependent experiments for each drug. Thus far, about 150 chemicals, primarily medicinal compounds, have been selected for the database. Over 27 000 gene expression profiles have been compiled for multiple doses and times in rat livers and kidneys, as well as rat and human hepatocytes, through comprehensive analysis using the Affymetrix GeneChip® Affymetrix, Inc., Santa Clara, CA, USA. These gene expression profiles, conjugated with histopathological changes, blood biochemical examination results, and the other phenotypic profiles, are stored in our database with a web-based tool for statistical analysis, genomics-assisted toxicity evaluation system developed by the toxicogenomics project in Japan (TG-GATEs). Thirteen of the 150 chemicals were typical nephrotoxicants or drugs showing clinical side-effects (e.g., cisplatin, carboplatin, gentamicin, vancomycin, phenacetin, and buccetin), and 20 chemicals exhibited nephrotoxicity in addition to hepatotoxicity (e.g., phenylbutazone, ethionine, and indomethacin).

### 3 IDENTIFICATION OF BIOMARKER GENES WITH TOXICOGENOMICS DATA

The first step from toxicogenomics to mechanistic and predictive toxicology is the identification

of an individual gene or a cluster of genes detective or predictive of certain types of toxicity; these "signature genes" are employed as biomarkers. A biomarker is defined by the International Programme on Chemical Safety (IPCS) of the WHO as any substance or its product, structure, or process that can be measured in the body and that can influence or predict the incidence of disease outcome. The ideal biomarker provides a sensitive, informative, and reproducible indicator of potential adverse effects at times or doses preceding overt tissue damage, toxicity, or disease initiation. The discovery and validation of biomarkers is useful for application in high-throughput experimental systems to characterize target organ effects and to detect specific toxicity end-points in the early steps of a compound's development. The identification and utilization of biomarkers through toxicogenomics have several further applications: from current use in pre-clinical toxicology to risk characterization and risk assessment of chemicals; from early clinical stages of drug development to the later stages; and even into daily clinical use in diagnostics, disease classification, and therapeutic monitoring.

### 4 GENE SELECTION FROM MICROARRAY DATA

Identifying biomarker genes in huge sets of microarray data is referred to as the gene selection problem. In selecting genes from a microarray with good separation between toxic and non-toxic drug-treated samples, one seeks the significant genes that are affected by the adverse drug effects, or even those that caused the adverse reaction. This is a key step toward understanding mechanistic and predictive toxicology through the underlying biological process.

Gene selection is also relevant in the classification problems in machine learning, in which the class of toxic response (including non-toxic responses) of a sample (e.g., drug-treated organ or tissue) is determined by a classifier. A sample is represented as a feature vector  $\mathbf{x}$ . Each dimension in the feature vector  $\mathbf{x}$  holds the expression value of a particular gene, which is obtained from a DNA microarray experiment. The classifier is constructed by inputting  $N$  feature vectors (called training data) with known toxic response outcomes into the machine-learning algorithms. However, because of the low number of samples ( $N$ : the number of the feature vectors)

and the high number of observed genes [dim(x): dimension of each feature vector], using all genes to classify the samples into good and bad outcomes incurs a high risk of over-training (or over-fitting). Over-training, in this case, means including the noise in the data, which may increase the generalization error (the error rate of the resulting predictor on samples that were not used during the training phase). This may yield a result that achieves high accuracy levels for the training data, but does not generalize to new data. The underlying problem is that if the sample size is much smaller than the number of genes, then one can distinguish different types of toxicity based on the noise present in these measurements, rather than on distinct biological characteristics of their gene expression levels. One approach to overcome this serious problem is to reduce the number of genes by removing irrelevant and redundant features; a method known as "feature selection" in computer science.

## 5 FEATURE SELECTION ALGORITHMS

Feature selection methods can be broadly categorized into the filter model and the wrapper model. The filter model separates feature selection steps from the machine-learning algorithm, and relies on the general characteristics of the training data to select features (Figure 1a). The wrapper model uses the predictive accuracy of a pre-determined learning algorithm to determine the quality of a selected subset (Figure 1b).

### 5.1 Filter Methods

Traditional methods in gene selection fall within the filter model. Filter approaches remove irrelevant features based on a ranking of all the genes according to their individual relevance or discriminative power to the target class. Filter approaches generally follow three simple steps (Figure 1a); they: (a) rank all the genes in the microarray data using a filter method; (b) choose the top-ranked  $n - 1$  genes as the best feature gene subset, and (c) construct a classifier in learning algorithm using the selected  $n - 1$  genes. The filter-type feature selection (steps 1 and 2) is independent of the classifier-learning algorithm (step 3). One major problem with the filter model is the selection of a threshold for discarding irrelevant features. All the features are given a score

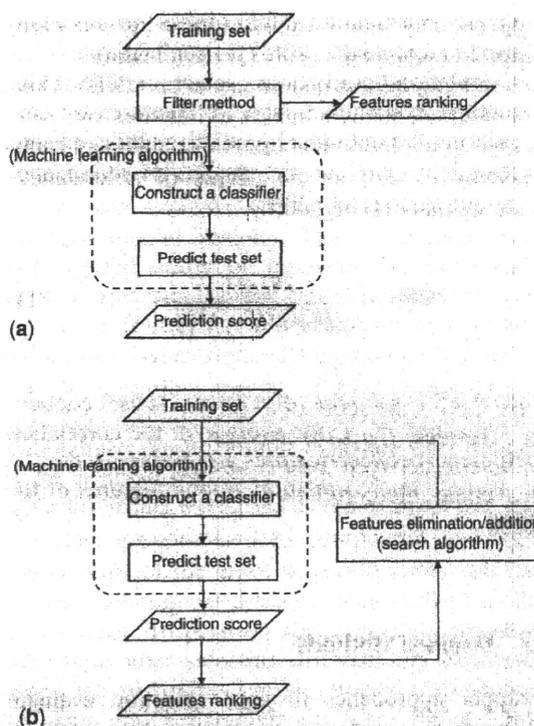


Figure 1. Calculation procedures of filter method (a) and wrapper method (b).

by the filter algorithm, but determining the optimal threshold " $n$ " for the data is difficult. We must choose the threshold " $n$ " using some appropriate criteria.

Filter methods are selected to produce the most relevant possible ranking. Many methods have been developed based on statistical tests. They are:

1. *t*-Statistic: This criterion measures the worth of a feature gene by computing the value of the *t*-statistic between toxic and non-toxic classes (Thomas *et al.*, 2001; Tsai, Chen and Chen, 2003).
2. ReliefF: This is a feature-weighting algorithm sensitive to feature interactions. The key idea of ReliefF is to rate features according to how well their values distinguish among instances of different classes and how well they cluster instances of the same class (Kononenko, 1994). To this end, ReliefF repeatedly chooses a single instance at random from the data, and then locates the nearest instances of the same class and the nearest instances pertaining to different

classes. The feature values of these instances are used to update the scores for each feature.

3. Correlation-based feature selection (CFS): This method evaluates a subset of features by considering the individual predictive ability of each feature as well as the degree of redundancy among them (Hall, 2000),

$$\text{CFS} = \frac{k\bar{\tau}_{cf}f}{\sqrt{k + k(k-1)\bar{\tau}_{ff}}} \quad (1)$$

where CFS is the score of a feature subset containing  $k$  features,  $\bar{\tau}_{cf}$  is the average of the correlation coefficients between features and classes, and  $\bar{\tau}_{ff}$  is the average intercorrelation among features of the subset.

## 5.2 Wrapper Methods

Wrapper approaches to gene selection evaluate the prediction performance of a learning machine trained for each given feature subset using a search algorithm to explore possible combinations of features. The strategy of wrapper approaches is composed of four steps (Figure 1b):

1. Choose a machine-learning algorithm to evaluate the prediction score of a feature subset. Many machine-learning algorithms have been developed, including the Naïve Bayes classifier, linear discriminant methods, support vector machines (SVM) (Vapnik, 1998), and artificial neural networks (Bishop, 1995). Cross-validation (CV) is often used for scoring prediction performance.
2. Choose a search algorithm.
3. Perform the search algorithm and note the best subset encountered.
4. Output the encountered subset with the best score.

Search algorithms compare all possible feature combinations, which are usually too numerous to be exhaustively explored. Of these search methods, greedy methods (forward selection or backward elimination) are the most popular. At each round of CV, forward selection adds the best feature; backward elimination deletes the worst feature.

## 5.3 Recursive Feature Elimination with Support Vector Machine (RFE-SVM)

This is a type of wrapper algorithm that couples recursive feature elimination with linear SVM (Guyon *et al.*, 2002). In a linear SVM, the decision function (classifier) is given as  $f(x) = w^T x + b$  or  $f(x) = \sum_{k=1}^n w_k x_k + b$ . For a given feature  $x_k$ , the absolute value of its weight  $w_k$  indicates the significance of its contribution to the margin of the linear SVM and to the construction of a linear classifier. Hence, the weight  $w_k$  is used as a feature-ranking coefficient in RFE-SVM. This algorithm first constructs a linear SVM classifier from the microarray data with  $n$  genes; then, the gene with the lowest  $w_k^2$  is removed and another classifier is trained on the remaining  $n - 1$  genes. This process is repeated until only one gene remains.

A gene ranking is produced from the order in which the genes were eliminated, and the last remaining gene is the most relevant gene. However, because of computational cost considerations, the algorithm is often implemented in such a way that several features are reduced at once. In such implementations, the method produces a feature subset ranking instead of a feature ranking. The optimal signature genes are the gene sets with the best prediction performance throughout the elimination process.

## 5.4 Filter Approaches Versus Wrapper Approaches

The filter methods are useful in practice because they are much faster than wrapper methods. Filters exclude irrelevant genes but cannot remove redundant genes because methods that select important genes based on individual gene information fail to account for mutual information among genes. The problem of redundancy among selected genes is twofold: the selected gene set may less comprehensively represent the target class than one of the same size without redundant genes would and in including all representative genes, redundant genes unnecessarily increase the size of the selected gene set, which will in turn affect the mining performance of the small sample.

In contrast, wrapper methods were developed to select discriminative genes while decreasing gene