厚生労働科学研究費補助金（化学物質リスク研究事業）
研究分担報告書

ES 細胞試験における化学物質影響の種間・細胞間比較に関する研究

研究分担者　大迫誠一郎　東京大学　准教授

研究要旨

化学物質の安全性評価で重要な問題であるヒトへの生体影響を予測するシステムを開発するため、ヒト胚性幹細胞試験（EST）を利用、影響評価のためのデータ適応法の標準化に関する開発を行う。モデル化合物としてメチル水銀と TCDD をヒトならびにマウス ES 細胞の神経系分化系に曝露し、その影響を比較検討した。その結果、細胞死と分化後の細胞形態表現型の影響において、両種間でメチル水銀の濃度に対する感受性が異なることがわかった。

共同研究者

何小明　東京大学　疾患生命工学センター

A. 研究目的

ヒト胚性幹細胞（ES 細胞）の再生医療技術への応用研究が進展し、随時大量に準備可能な ES 細胞から、各種細胞への分化培養系が確立されつつある。このような分化培養系は受精卵から胚、胎児を経て成熟個体に至るまでの過程を再現しており、ヒトの発生影響試験の理想的モデルと言える。本研究では化学物質の安全性評価で重要な問題であるヒトへの生体影響を予測するシステムを開発するため、ヒト胚性幹細胞試験（EST）を利用する。マウス ES 細胞実験との比較検討とともに、取得する影響指標（遺伝子発現情報および形態情報）から数理工学理論に基づくバイオインフォマティクスを駆使し、ヒトへの影響レベルの予測を試みるが、そのために使用する確率推論アルゴリズムに適用するための実験系確立ならびにシステム標準化を実施する。

B. 研究方法

マウス ES 細胞で確立された神経細胞分化系を、ヒト ES 細胞にも適用し、化学物質 2 種（メチル水銀・TCDD）を曝露、両 ES 細胞分化系から、細胞形態情報（IN Cell Analyzer 1000 解析（MAP2・GFAP ラベル）で神経突起長、分岐点数、交差点数を測定）、および遺伝子発現情報（神経細胞への分化誘導前後の遺伝子変動プロファイルをマイクロアレイと qPCR で測定）を取得した。これら情報を基に「表現型構成要素間ネットワーク」形成を試みた。

（倫理面への配慮）

東京大学ライフサイエンス委員会倫理審査専門委員会で 2009 年 12 月機関承認を得、それにより、2010 年 1 月文部科学省より使用許可を得た。

C. 研究結果

メチル水銀は 10 nM からマウス胚様体（EB）の生存率を低下させる作用を示したが、生存した神経細胞のデンドライト伸長には影響がなかった。一方、ヒト EB への生存率には影響がないものの、分化した神経細胞のデンドライト伸長には抑制的効果を示した。また、TCDD はヒトならびにマウス両分化系での 100 nM の高濃度においても分化レベルの影響は見られなかった。ヒトおよびマウス間で十分に比較可能なオーソログ遺伝子 20 種を定量 PCR で測定したデータセットと上記形態情報（IN Cell Analyzer 1000 解析（MAP2 染色））による確率推論モデルにおいても、マウスではメチル水銀に起因するネットワーク構造が見られないのに対して、ヒトではメチル水銀を最上位（親 Node）にもつネットワークを描くことがわかった（図 1）。

D. 考察

メチル水銀によりマウスではEB生存率低下を、ヒトではEB生存率に影響はないものの分化神経細胞のデンドライト伸長抑制効果が観察できたことは、この物質のヒト神経細胞への特有の効果だと考えられる。マウスではこれまで、実験動物による胎生期メチル水銀曝露で神経系影響モデルを作ることが困難とされてきたが、これは神経細胞機能異常に先んじて細胞死が起きていた可能性がある。またヒトでは細胞死よりも神経系機能異常が生じるため胎生期影響が顕著に出るのではないかと考察できる。

これらのデータはヒトの感受性を評価するためにヒトESとマウスES細胞を利用した本システムの有効性を証明したものと言える。また、TCDDは神経分化影響が重篤と予想されたが、ヒトならびにマウス両分化系での100 nMの高濃度においても形態レベルでの影響は見られず、ダイオキシン類のヒトへの神経発達影響は少ないものと考察できる。

E. 結論

上記の結果より、本研究課題で開発した神経細胞分化系は、同一化合物であっても、発生という観点からヒトにおける特徴的影響を観察できるポテンシャルをもつことが実証された。また、評価法として導入するベイズ推定モデルも遺伝子発現変動とその結果である細胞表現型に予測性を持つことが提示できた。これら研究成果はヒトの感受性を評価するためにヒトESとマウスES細胞を利用した本システムの有効性を如実に示している。

F. 健康危険情報

なし。

G. 研究発表

1. 論文発表

1. Sone H, Okura M, Zaha H, Fujibuchi W, Taniguchi T, Akanuma H, Nagano R, Ohsako S, and Yonemoto J. pCEC: a toxicogenomics database with a toxicoinformatics system for risk evaluation and toxicity prediction of environmental chemicals. *J Toxcol Sci,* (in press).
2. Alam MS, Ohsako S, Matsuwaki T, Zhu XB, Tsunekawa N, Kanai Y, Sone H, Tohyama C, and Kurohmaru M. Induction of spermatogenic cell apoptosis in prepubertal rat testes irrespective of testicular steroidogenesis: A possible estrogenic effect of di(n-butyl) phthalate. *Reproduction*, (in press).
3. Ohsako S, Fukuzawa N, Ishimura R, Kawakami T, Wu Q, Nagano R, Zaha H, Sone H, Yonemoto J, and Tohyama C. Comparative contribution of the aryl hydrocarbon receptor gene to perinatal stage development and dioxin-induced toxicity between the urogenital complex and testis in the mouse. *Biol Reprod*, (in press).
4. Ishihara K, Ohsako S, Tasaka K, Harayama H, Miyake M, Warita K, Tanida T, Mitsuhashi T, Nanmori T, Tabuchi Y, Yokoyama T, Kitagawa H, and Hoshi N. When does the sex ratio of offspring of the paternal 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) exposure decrease: In the spermatozoa stage or at fertilization? *Reprod Toxicol,* (in press).
5. **大迫誠一郎. プログラムされる"病"の新たな仮説－環境化学物質による代謝系遺伝子の次世代エピゲノム変化. 科学**, 79, 984-988, (2009).
6. Ishimura R, Kawakami T, Ohsako S, and Tohyama C. Dioxin-induced toxicity on vascular remodeling of the placenta. *Biochemical Pharmacol,* 77, 660-669, (2009).
7. Kawakami T, Ito T, Ohsako S, Shiizakia K, Murakami Y, Hirowatarid K, Sato M, and Tohyama C. Possible Involvement of arylhydrocarbon receptor variants in TCDD-induced thymic atrophy and XRE-dependent transcriptional activity in Wistar Hannover GALAS rats. *J Toxicol Sci*, 34, 209-220,

(2009).

8. Sone H, Imanishi S, Akanuma H, Nagano R, Fukuda T, Ohsako S. Gene expression signatures of environmental chemicals in cancer and in developmental disorders. In "The roles of free radicals in biology and medicine", MEDIMOND SRL Publ., 45-52, (2009).

2. 学会発表

1. Ohsako S. Perinatal exposure to dioxins alters the epigenomic program that influences the susceptibility to carcinogens exposed in adulthood. Horiba Conference/CDBIM Symposium, 21st Century Advances in the Molecular Toxicology of Environmental Chemicals and Pathogenesis of Disease, Tokyo, JAPAN, (2009) Oct 27th.
2. 大迫誠一郎. ダイオキシンの胎生期曝露による肝臓エピゲノム変化. 第 82 回日本生化学会、シンポジウム「化学物質による遺伝子修飾と毒性発現」. 神戸 (2009) 10 月
3. Ohsako S, Tohyama C. Reduction of methylation frequency in the matured mouse hepatic CYP1A1 promoter region by perinatal dioxin exposure. The 5th International Congress of Asian Society of Toxicology, Taipei, Taiwan, (2009) Sept 11th.
4. 大迫誠一郎. 環境汚染物質の胎生期曝露影響は生後の肝臓においてエピジェネティックメモリーとして残る. 第 3 回日本エピジェネティクス研究会、東京 (2009) 7 月
5. Miyazaki W, Nagano R, Sone H, Tohyama C, Ohsako S. Effects of dioxin and OH-PCB on neural differentiation from mouse ES cells. Horiba/CDBIM Symposium, Tokyo, JAPAN, (2009)
6. Nagano R, Akanuma H, Koikegami S, Imanishi S, Miyazaki W, Okura M, Zaha H, Ohsako S, Sone H. Development of an image profiling system to evaluate for the effects of Thalidomide and permethrin in neural differentiation from mouse embryonic stem (ES) cells. Horiba/CDBIM Symposium, Tokyo, JAPAN, (2009)
7. Miyazaki W, Nagano R, Sone H, Tohyama C, Ohsako S. Effects of dioxin and OH-PCB on neural differentiation from mouse ES cells. DIOXIN2009, Beijin, China (2009)
8. Nagano R, Akanuma H, Koikegami S, Imanishi S, Miyazaki W, Okura M, Zaha H, Ohsako S, Sone H. Development of an image profiling system to evaluate for the effects of chemicals in neural differentiation from mES cells. The 32nd Ann Meeting of Japanese Society of Neuroscience, Nagoya, JAPAN, (2009)

H. 知的財産権の出願・登録状況

1. 特許取得

曽根秀子、大迫誠一郎、永野麗子、今西聡、赤沼宏美、宮崎航.「胎生プログラミングに対する影響を評価するための方法」. 特願 2009-81497（識別番号 100078662）(2009)

大迫誠一郎、栗田尚佳.「CpG メチル化頻度の変動をゲノムワイドに比較解析するための DNA 試料作成方法」. 米国仮出願 （出願番号：61/309971）(2010)

2. 実用新案登録

なし

3. その他

なし

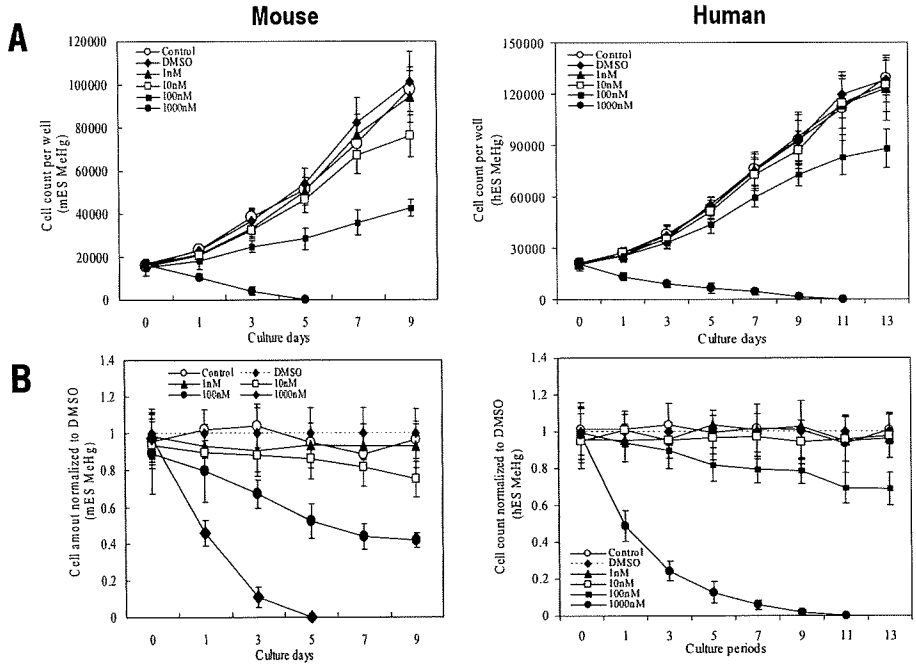

図1. マウスES細胞およびヒトES細胞の神経系細胞培養系に及ぼすMeHgの影響。MTTアッセイによって用量依存的な細胞活性の測定を行った。MeHg濃度:1, 10, 100, 1000 nM。A, 培養時間によるウェル内細胞数の推移。B, 培養時間による細胞活性の推移を細胞数当たりに換算して計算した。各時間後とのDMSO対照群の活性を1とした。

Human; Baysian network model analysis
Integrated type: p>0.5
(14 gene sets, 3 morphology sets, 1 chemical (MeHg))

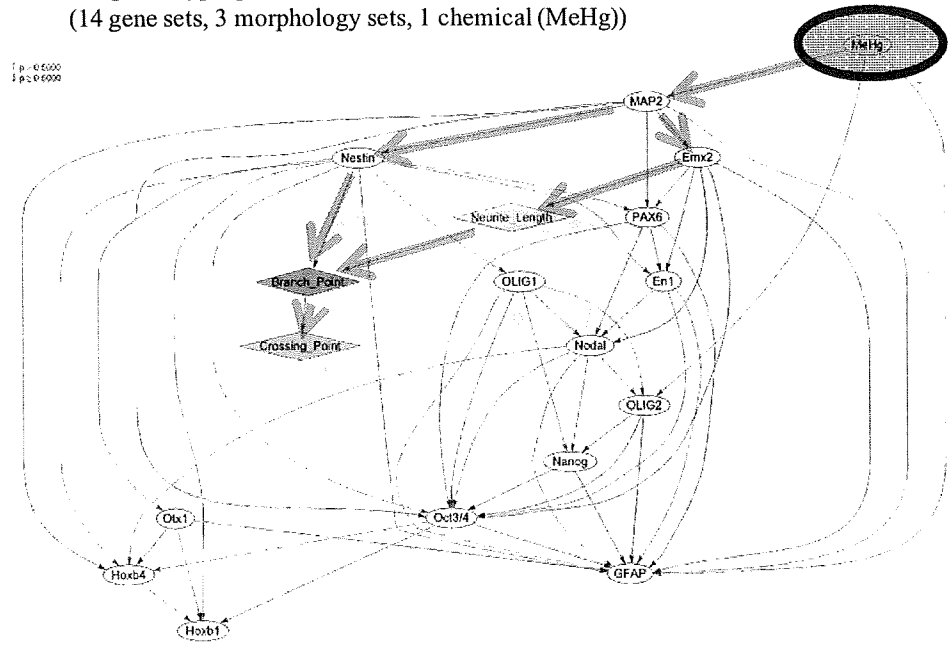

図2. マウスES細胞およびヒトES細胞の神経系細胞培養系に及ぼすMeHgの影響の確率推論アルゴリズム（ベイジアンネットワークモデル）に基づく統合化ネットワーク解析結果。マウス細胞培養系のDay 23における14種の遺伝子発現データと3種の形態情報（Neurite length, Branching points, Crossing points）ならびにHeHgとの関係をコンピュータ解析し、自動的に図式化させた。ベイズモデルにおけるベータ値が正の関係のものを赤い矢印、負の関係を青い矢印で示した。相互作用確率でp>0.5のものを抽出した。

研究成果の刊行一覧表

別紙4

雑誌

発表者氏名	論文タイトル名	発表誌名	巻名	ページ	出版年
Seiichiroh Ohsako, Noriho Fukuzawa, Ryuta Ishimura, Takashige Kawakami , Qing Wu, Reiko Nagano, Hiroko Zaha, Hideko Sone, Jyunzo Yonemoto, and Chiharu Tohyama.	Comparative contribution of the aryl hydrocarbon receptor gene to perinatal stage development and dioxin-induced toxicity between the urogenital complex and testis in the mouse.	*Biol Reprod*			in press
Kana Ishihara, Seiichiroh Ohsako, Ken Tasaka, Hiroshi Harayama, Masashi Miyake, Katsuhiko Warita, Takashi Tanida, Tomoko Mitsuhashi, Takashi Nanmori, Yoshiaki Tabuchi, Toshifumi Yokoyama, Hiroshi Kitagawa and Nobuhiko Hoshi	When does the sex ratio of offspring of the paternal 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) exposure decrease: In the spermatozoa stage or at fertilization?	*Reprod Toxicol*			in press
Hideko Sone H, Hiromi Akanuma H, Tomokazu Fukuda.	Oxygenomics and environmental stressor.	*Redox Reports*			in press
Jean-François Pessiot, Hirokazu Chiba, Hiroto Hyakkoku, Takeaki Taniguchi, Wataru Fujibuchi PeakRegressor identifies composite sequence motifs responsible for STAT1 binding sites and their potential rSNPs, *PLoS ONE* (under revision).	PeakRegressor identifies composite sequence motifs responsible for STAT1 binding sites and their potential rSNPs	*PLoS ONE*			under revision
Yuki Tochigi, Natsuko Sato, Takehiko Sahara, Chun Wu, Shinya Saito, Tsutomo Irie, Wataru Fujibuchi, Takako, Goda, Ryoichi Yamaji, Masahiro Ogawa, Yoshihiro Ohmiya, Satoru Ohgiya	A sensitive and convenient yeast reporter assay for high-throughput analysis by using a secretory luciferase from Cypridina noctiluca. Analytical Chemistry	*Analytical Chemistry*			accepted
Hideko Sone, Masahiro Okura, Hiroko Zaha,Wataru Fujibuchi, Takeaki Taniguchi, Hiromi Akanuma, Reiko Nagano, Seiichiro Ohsako, Junzo Yonemoto	Profiles of Chemical Effects on Cells (pCEC): a toxicogenomics database with a toxicoinformatics system for risk evaluation and toxicity prediction of environmental chemicals.	*J Toxicol Sci.*	35	115-23	2010
Hideko Sone, Tomokazu Fukuda, Hiroyoshi Toyoshiba, Takeharu Yamanaka, Perham Fred, Christopher J Portier	Importance of CDK7 for G1 re-entry into the mammalian cell cycle and identification of new downstream networks using a computational method.	*The Open Cell Signaling Journal*	2	1-12	2010
Mohammad Shan Alam, Seiichiroh Ohsako, Takashi Matsuwaki, Xiao Bo Zhu, Naoki Tsunekawa, Yoshiakira Kanai, Hideko Sone, Chiharu Tohyama and Masamichi Kurohmaru.	Induction of spermatogenic cell apoptosis in prepubertal rat testes irrespective of testicular steroidogenesis: A possible estrogenic effect of di (n-butyl) phthalate.	*Reproduction*	139	427-437	2010

発表者氏名	論文タイトル名	発表誌名	巻名	ページ	出版年
幡野晶子、Harry Amri Moesa、千葉啓和、谷口丈晃、永家聖、山根木康嗣、藤渕航	細胞分化解析を目指した網羅的ヒト細胞データベース「CELLPEDIA」	*情報処理学会研究報告*	2010-BIO-20	No.12	2010
金蕙鈴、加藤毅、茂櫛薫、田中博、藤渕航	正則化正準相関解析を用いた抗がん剤の影響による共通パスウェイ解析	*情報処理学会研究報告*	2010-BIO-20	No.9	2010
Jean-François Pessiot, Hirokazu Chiba, Hiroto Hyakkoku, Takeaki Taniguchi, Wataru Fujibuchi	PeakRegressor identifies composite sequence motifs responsible for STAT1 binding sites and their potential rSNPs	*Proceedings of Critical Assessment of Massive Data Analysis 2009*		4-11	2009
Wataru Fujibuchi, Hirokazu Chiba, Hideo Akiyama, and Hitoshi Shiku	Designing Pyro-Primer Sequences Using a Simulated Annealing Algorithm, to Critically Target mRNAs in Quantitative Cell Analysis	*Proceedings of the 6'th International Forum on Post-genome Technologies*		253-7	2009
Hideaki Karasawa, Koh Miura, Wataru Fujibuchi, Kazuyuki Ishida, Naoyuki Kaneko, Makoto Kinouchi, Mitsunori Okabe, Toshinori Ando, Yukio Murata, Hiroyuki Sasaki, Kazuhiro Takami, Akihiro Yamamura, Chikashi Shibata and Iwao Sasaki	Down-regulation of cIAP2 enhances 5-FU sensitivity through the apoptotic pathway in human colon cancer cells.	*Cancer Sci.*	100	903-913	2009
百石弘澄、杉原稔、諏訪牧子、加藤毅、山名早人、藤渕航	2-way prediction 法によるGPCRリガンドの結合予測	*情報処理学会研究報告*	2009-BIO-18	No.2	2009
Wataru Fujibuchi , Hyeryung Kim, Yoshifumi Okada, Takeaki Taniguchi, Hideko Sone	High-performance gene expression module analysis tool and its application to chemical toxicity data.	*Methods Mol Biol*	577	55-65	2009
Ryuta Ishimura, Takashige Kawakami, Seiichiroh Ohsako, and Chiharu Tohyama.	Dioxin-induced toxicity on vascular remodeling of the placenta.	*Biochemical Pharmacol*	77	662-669	2009
Kana Ishihara, Seiichiroh Ohsako, Ken Tasaka, Hiroshi Harayama, Masashi Miyake, Katsuhiko Warita, Takashi Tanida, Tomoko Mitsuhashi, Takashi Nanmori, Yoshiaki Tabuchi, Toshifumi Yokoyama, Hiroshi Kitagawa and Nobuhiko Hoshi	Possible Involvement of arylhydrocarbon receptor variants in TCDD-induced thymic atrophy and XRE-dependent transcriptional activity in Wistar Hannover GALAS rats.	*J Toxicol Sci*	34	209-220	2009

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
藤渕航	シミュレーテッドアニーリングによる多重プライマー配列デザイン法	神原秀記、松永是、上田充美	シングルセル解析の最前線	シーエムシー出版	日本	2010	265-273
加藤毅、藤渕航	Kernel Classification Methods for Cancer Microarray Data	Frank Emmert-Streib and Matthias Dehmer	Medical Biostatistics for Complex Diseases	Wiley-Blackwell	ドイツ	2010	279-303
Wataru Fujibuchi, Hyeryung Kim, Yoshifumi Okada, Takeaki Taniguchi, Hideko Sone	High-performance gene expression module analysis tool and its application to chemical toxicity data	Hisashi Koga	Methods in Mol. Biol. 577: Reverse Chemical Genetics	Humana Press	米国	2009	55-65
Hideko Sone, Satoshi Imanishi, Reiko Nagano, Hiromi Akanuma, Tomokazu Fukuda, Seiishiroh Ohsako	Gene expression signatures of environmental chemicals in cancer and in developmental disorders. In: Biophys.Soc.China (BSC)ed.	Baolu Zhao, Michael Davies, Enrique Cadeans	The Roles of Free radicals in Biology and Medicine	Medimond S.r.l.	中国	2009	45-52
大迫誠一郎	プログラムされる"病"の新たな仮説－環境化学物質による代謝系遺伝子の次世代エピゲノム変化		科学	岩波書店	日本	2009	984-989

研究成果の刊行物・別刷り

The Journal of Toxicological Sciences (J. Toxicol. Sci.)
Vol.35, No.1, 115-123, 2010

Toxicogenomics/proteomics Report

Profiles of Chemical Effects on Cells (pCEC): a toxicogenomics database with a toxicoinformatics system for risk evaluation and toxicity prediction of environmental chemicals

Hideko Sone[1], Masahiro Okura[1], Hiroko Zaha[1], Wataru Fujibuchi[2], Takeaki Taniguchi[3], Hiromi Akanuma[1], Reiko Nagano[1], Seiichiro Ohsako[4] and Junzo Yonemoto[1]

[1]Research Center for Environmental Risk, National Institute for Environmental Studies, 16-2 Onogawa, Tsukuba, Ibaraki 305-8506, Japan
[2]Advanced Industrial Science and Technology (AIST), Computational Biology Research Center, 2-42 Aomi, Koto-ku, Tokyo 135-0064, Japan
[3]Mitsubishi Research Institute, Inc., Practice Areas and Industry Sectors, 3-6 Otemachi 2-chome Chiyoda-ku, Tokyo 100-8141, Japan
[4]Graduate School and Faculty of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

(Received September 3, 2009; Accepted October 21, 2009)

ABSTRACT — Profiles of Chemical Effects on Cells (pCEC) is a toxicogenomics database with a system of classifying chemicals that have effects on human health. This database stores and handles gene expression profiling information and categories of toxicity data. Chemicals are classified according to the specific tissues and cells they affect, the gene expression changes they induce, their toxicity and biological functions in this database system. The pCEC system also analyzes relationships between chemicals and the genes they affect in specific tissues and cells. The reason why we developed pCEC is to support decision-making within the context of environmental regulation. Especially, exposure to environmental chemicals during fetal and newborn development may result in a predisposition to various disorders such as cancer, learning disabilities and allergies later in life. The identification and prediction of hazardous chemicals using limited information are important issues in human health risk management. Therefore, various toxicity information including lethal dose 50 (LD50), toxicity pathways and pathological data were loaded into pCEC. pCEC is also a facility for query, analysis and prediction of unknown toxicochemical reaction pathways and biomarkers which are based on toxicoinformatical data mining approaches. This database is available online at http://project.nies.go.jp/eCA/cgi-bin/index.cgi. The current version of the database has information on the hepatotoxicity, reproductive toxicity and embryotoxicity of chemicals.

Key words: Toxicogenomics, Toxicoinformatics, Risk assessment, Toxicity prediction, Chemical profiling

INTRODUCTION

Risk assessment for human health is now standing in front of a new door of toxicity testing in the 21st century. Many scientists at universities and regulatory agents have proposed the need for a paradigm-shift from the old-fashioned-style toxicity test using many animals and high doses of chemicals (Andersen and Krewski, 2009; Bushnell *et al.*, 2007; Fentem *et al.*, 2004). Exposure to environmental chemicals during fetal and newborn development may result in a predisposition to various disorders such as cancer, learning disabilities and allergies later in life. The identification and prediction of these chemicals using limited information are important issues in human

Correspondence: Hideko Sone (E-mail: hsone@nies.go.jp)

health risk management (Krewski *et al.*, 2009; Woodruff *et al.*, 2008). The National Research Council (NRC) of the National Academies released a report entitled, "Toxicity testing in the 21st Century" in 2007 (NRC, 2007). The central components of their vision are toxicity pathway and targeted testing, and the main components of their vision are collection and computational modeling of physical and chemical properties, environmental concentrations and stability, routes of human exposure, the potential for bioaccumulation, metabolites and molecular interactions (Kavlock *et al.*, 2008). The United States Environmental Protection Agency (U.S. EPA) has already developed and publicly opened access to Aggregated Computational Toxicology Resource (ACToR), which is a database holding essentially all publicly available information on the identity, structure, physical-chemical properties, *in vitro* assay results, and *in vivo* toxicology data of chemicals (Judson *et al.*, 2008, 2009). In Europe, the European Union established a new regulation concerning the Registration, Evaluation, Authorization and Restriction of Chemicals (REACH) (Ahlers *et al.*, 2008), in which all relevant industrial chemicals must now be assessed by industries themselves and the industries are responsible for risk assessment. REACH sets certain minimum data requirements in order to achieve a high level of protection for human health and the environment. All available data from the different steps have to be integrated to come to an overall conclusion on the toxicity of the chemical (Mattingly, 2009; Maxwell *et al.*, 2008). In this background, SuperToxic (Schmidt *et al.*, 2009) and ToxRTool (Schneider *et al.*, 2009) from Germany and MMSINC (Masciocchi *et al.*, 2009) from Italy were released. SuperToxic is a database that compiles a wide range of compounds with large amounts of bioassay data, and ToxRTool is a tool that assesses the reliability of toxicological data. MMSINC is a chemical structure database, where data on chemicals are appropriately stored and annotated. Thus, databases that collected chemical properties and toxicological data are being used to forecast the unknown toxicity of unstudied chemicals. In order to predict the unknown toxicity of chemicals, categorization according to toxicity is an important step after completing data collection, as is making various templates of toxic type or toxicity pathways (Wullenweber *et al.*, 2008). The reliance on toxicity pathway perturbations for human health risk assessment will require sufficient understanding of such pathways to explain phenotypic outcomes in animals. The successful prediction of chemical toxicity will require the development of three areas: a comprehensive suite of toxicity pathway assays, analytical tools for the data, and regulatory and political infrastructure enabling their use in health risk assessment (Mattingly, 2009, Maxwell *et al.*, 2008). To solve these problems, genomic analysis such as microarray data analysis is one strong approach (Wang, 2008; Waters and Jackson, 2008).

Toxicogenomics has been widely used for elucidating the molecular and cellular actions of chemicals and other environmental stressors on biological systems (Guyton *et al.*, 2009; Waters and Fostel, 2004; Waters *et al.*, 2003). Classification of known or new toxicants based on signatures of gene expression will be a basis for predicting toxicity before any potential functional damage (Aardema and MacGregor, 2002; Benigni *et al.*, 2007; Lambert *et al.*, 2009). Therefore, we developed a database accompanied by software for classification of known or new toxic chemicals based on signatures of gene expression with other toxicology data. The system was named the Profiles of Chemical Effects on Cells (pCEC) system. pCEC shows classifications of chemicals that act on particular cells from the viewpoint of gene expression signatures. pCEC also stores and handles gene expression profiling information and categorizations of bioassay data to elucidate toxicity. The compiled data in pCEC have been organized into a variety of chemical groups that are classified according to the type of molecular pathway or type of toxicity (Fig. 1). In comparison with other databases (Supplementary Table 1), intuitive visualization of gene expression information using clustering techniques such as a self-organizing map and minimum-spanning tree in pCEC is very unique and allows us to make toxicity predictions and find biomarkers more easily. Thus, this is a unique database containing information on the health effects of chemicals combined with gene alteration profiles in specific tissues and cell types. This is directly useful for risk evaluation and assessment of the effect of chemicals on human health. All of the data and search functions of pCEC are accessible through a user-friendly web interface that we describe later in this paper.

MATERIALS AND METHODS

pCEC was established on the basis of data from publicly available resources at the websites, Gene Expression Omnibus (GEO) database (Barrett *et al.*, 2005, 2007) in National Center for Biotechnology Information (NCBI), The Distributed Structure-Searchable Toxicity (DSSTox) (Williams-Devane *et al.*, 2009b; Williams-DeVane *et al.*, 2009a), the Chemical Effects in Biological Systems (CEBS) database in the National Institute of Environmental Health Sciences (NIEHS) (Fostel, 2008; Waters *et al.*, 2008), NCI60 project in the National Cancer Institute (Shoemaker, 2006) and the Toxicology data network

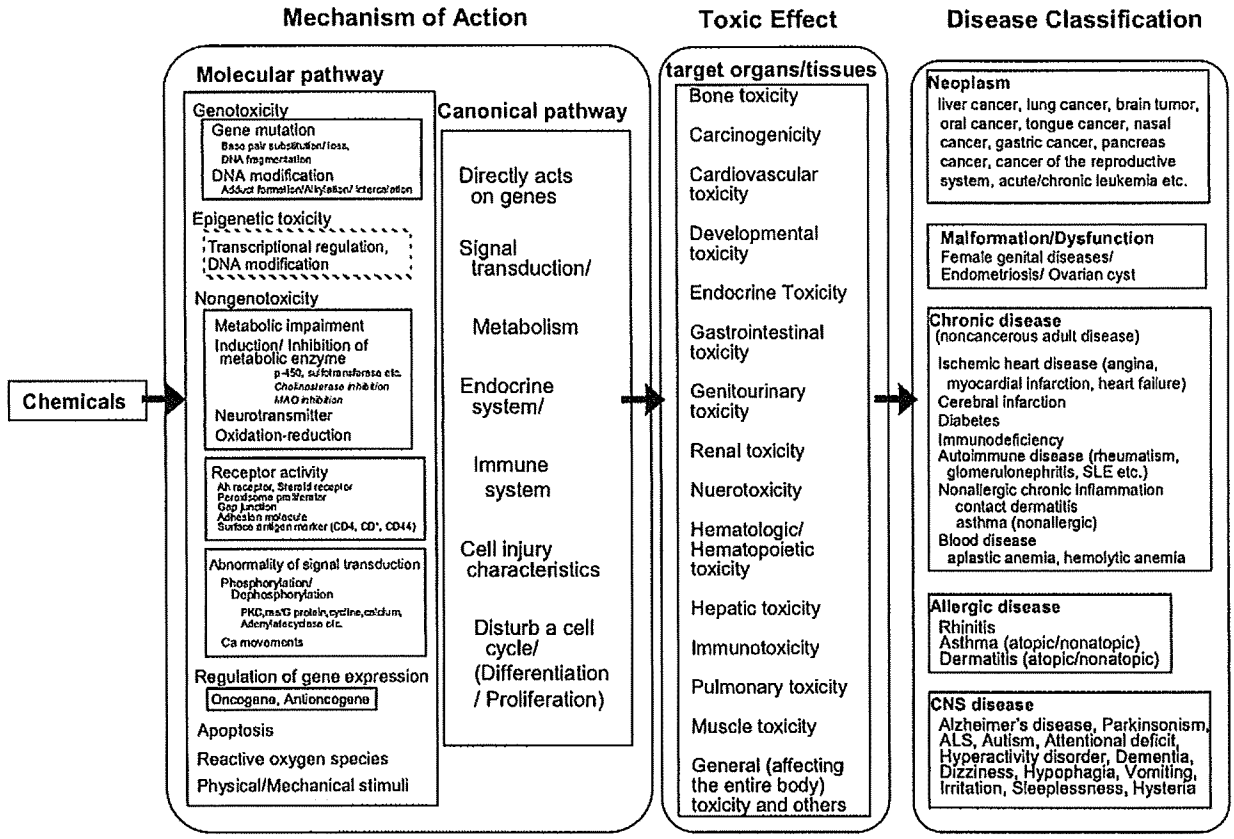

Fig. 1. Categorization of chemicals by their toxicity and molecular mechanism of action. The flow of pictures represents the molecular pathway, toxicity types and disease classification of a chemical when exposure to the chemical affects cells, tissues and organs. This flow influences the structure of pCEC. Neurotoxicity is subclassified as CNS (central nervous system including brain and spinal cord), PNS (peripheral nervous system including motor nerves and sensory nerves) and ANS (autonomic nervous system including sympathetic nerve, parasympathetic nerve).

(TOXNET) database (Tomasulo, 2005). By exploring comprehensively public datasets among these websites, we obtained three categories datasets. In other words, we could not find a battery of datasets about the other target organs such as kidneys and hearts. Three categories were datasets for the liver toxicity, the neuronal toxicity and the reproductive toxicity. To analyze expression values from large-scale gene expression data, we first normalize and convert them into a unified file format. Gene expression data normalized by Z score (*) transformation can be used directly in the calculation of significant changes in gene expression between different samples and experimental conditions. After obtaining formatted data, gene expression patterns for each chemical were used to create a heat map on a matrix using a self-organizing maps (SOM). The expression patterns extracted by the SOM were used to represent graphically Z-score values of the downregulated and upregulated genes. The heat-maps as an expression pattern of chemicals are then classified using a minimum spanning tree (MST) algorithm. For example, in "2007_rat_liver (102 chemicals, 2,488 genes)", data on the effects of chemicals on gene expression in the rat liver were downloaded from the website http://cebs.niehs.nih.gov/. There were a total of 133 chemical groups containing 964 microarray experiments. In the present study, 298 arrays among 9,215 probes in the dataset for both chemical exposure and control data were collected. After deleting low-abundance genes that showed no expression in all 298 experiments, a total of 7,614 probes were selected. Every pair of gene expression values was transformed into log-fold-change abundance by subtracting control values from the chemical exposure after taking log 2 and

subtracting the median value in each array. The log-fold-change values are normalized to Z-score by (x - mean) / SD. The gene expression data in the other projects were also normalized and converted into a unified file format.

The Z-score for gene expression values in cells can be approximated by the logarithm normal distribution with Zipf's law.

$Z = (x- \mu)/ \delta$

Where x is the gene expression value, μ is the mean expression value for all genes, and δ is the standard deviation of all genes.

Primary resources

The original microarray data sets used by this database were as follows, 2007_rat_liver: CEBS Accession 004-00002-0010-000-7; 2008_mouse_neuro: GSE367, GSE587, GSE1076, GSE1077, GSE1588, GSE1800, GSE3253, GSE3412, GSE5763; 2008_mouse_repro: GSE280, GSE438, GSE499, GSE3348, GSE4650; 2008_mouse_embryostem: GSE18503.

Supported web browsers

The following web browsers are fully supported, i.e., all of the features of pCEC including JavaScript and CSS styles should work properly: Internet Explorer 7, Firefox 3, Safari 3.

RESULTS AND DISCUSSION

pCEC (http://project.nies.go.jp/eCA/cgi-bin/index.cgi) has been developed as a database with a system of classifying chemicals that affect cells and induce gene expression changes, according to their toxicity and biological functions (Figs. 1 and 2). This database stores and handles gene expression profiling information and categorizes toxicity data. The system can separate chemicals into a variety of groups by the type of influence. Gene expression profiling was achieved by the SOM technique (Luo *et al.*, 2004; Törönen *et al.*, 1999), and the toxicity data are shown using MST algorithms (Xu *et al.*, 2001, 2002). The projects categorize chemicals according to the types of toxicity. The component in each option of pCEC reflects the framework of categorization of chemicals as shown in Fig. 1. All projects are tagged at the holder header by year, animal, and cell type, such as "2007 rat liver". The present latest version includes the following projects: 2007_rat_liver (102 chemicals, 2,488 genes); 2008_mouse_neuro (7 chemicals, 974 genes); 2008_mouse_repro (4 chemicals, 661 genes); 2008_mouse_embryostem (12 chemicals, 17,042 genes). The primary data can be linked to "the primary resource" in the top page of each project holder. pCEC has 4 tool boxes which are described below (Figs. 3, 4 and 5).

1) Chemicals. This section contains physiological and toxicity information on chemicals. The system categorizes data on chemicals that induce toxicity in animals, according to their mechanism of action, toxicity and structure (Fig. 3A).

2) Chemical Expression Neighbor. The user can compare the gene expression signature, which indicates gene expression changes affected by chemicals between different samples and experimental conditions (Fig. 3B). The gene expression signature of a chemical presents graphically genes whose expression is upregulated or downregulated with a SOM. The entire SOM signature for chemicals in each project is classified by the Minimum spading tree technique.

3) Correspondence Analysis (CA). CA is one of the multivariate analyses, a statistical visualization method of picturing the associations between the levels of a two-way contingency table. In this analysis, some measure of correspondence between the rows and columns of each datum value are mapped. The CA viewer of pCEC can display the results by plotting them in two-dimensional space (Fig. 3C).

4) Chemical Selector. The Chemical Selector allows the user to search for information on chemicals including physical and toxicity information according to information in the toxicity pathway. The Chemical Selector in pCEC provides three levels of categorization according to toxicity pathways, target organ toxicity and disease as shown in Fig. 1. For example, level 1 and level 2 represent classification of toxicity mechanisms and pathways according to the concept of categorization of toxicity at the molecular, cellular and tissue levels. In level 3, chemicals are listed according to the categories that the user had selected in levels 1 and 2 (Fig. 4).

5) Specific search options. If users selected one chemical in the site of "Chemical", there are many options to find the similar expression profiles and particular gene expression patterns (Fig. 5). As an example, Fig. 5 shows screenshots of toxicogenomic information for the chemical, antimycin A3 (Fig. 5A). Fig. 5B displayed networks for gene expression signatures of each chemical by using the shortest path problem algorism, which is the problem of finding a path between two nodes such that the sum of the weights of its constituent edges is minimized. Up and down regulated genes and distribution of expression values were listed in Fig. 5C. pCEC also directly link to SAMURAI, which is another program to find a module, and which is a minimal unit of common responsive genes affected by exposure to chemicals (Okada *et al.*, 2007).

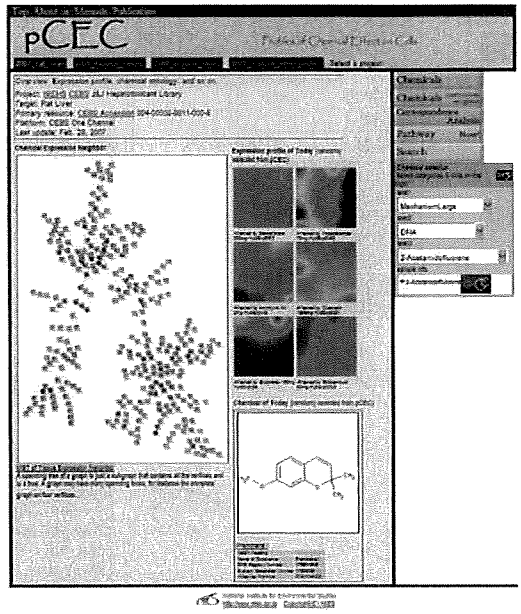

Fig. 2. Top page of Browser's window. Heat maps of the Gene expression patterns of chemicals are gathered by similarity of the map patterns.

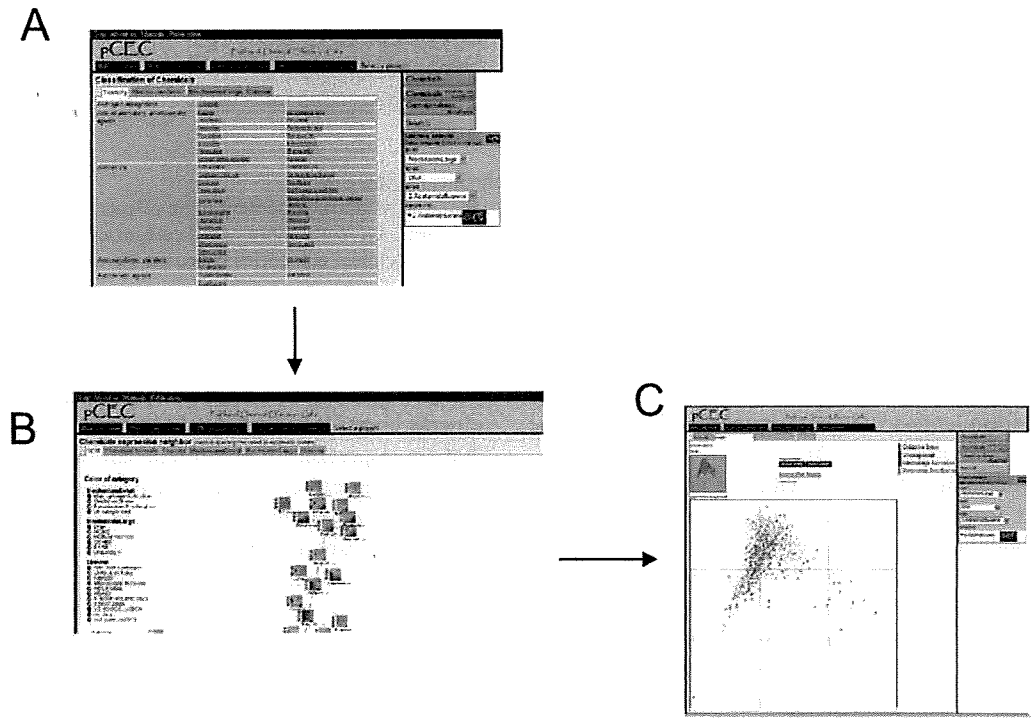

Fig. 3. "Chemicals" (A), "Chemicals Expression neighbor" (B), and "CA" (C). After clicking each button at the top-right corner, the respective page appears.

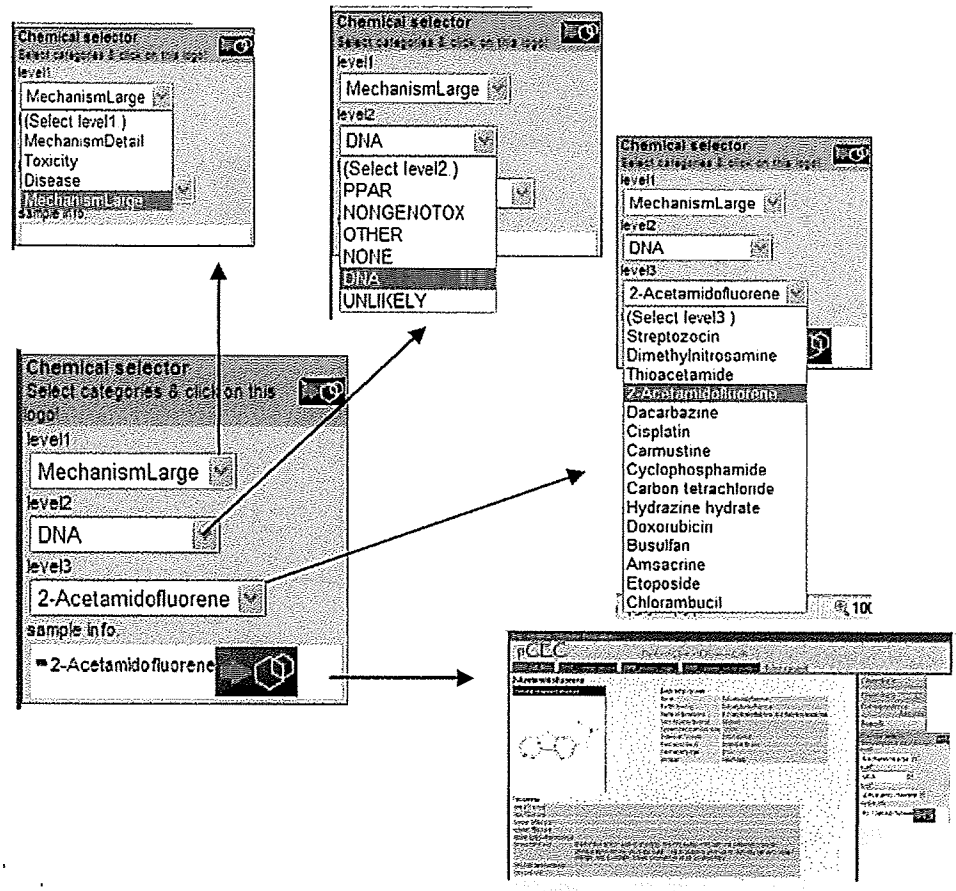

Fig. 4. Chemical selector this query option is categorized by the chemical at three levels. Level 1 indicates categories of chemical effects, which are the process during apparent toxic effects and disease outcomes from the first target on the cellular event as shown in Fig. 1. Level 2 indicates more detail categories than those of level 1 for mechanism of action and toxicity pathways. Level 3 indicates chemical categories based on the mother chemical structure.

The users can find specific molecular marker of chemical exposure from large-scale gene expression data using the SAMURAI program (Fig. 5D).

In a future plan, the pCEC database would be created to support decision-making within the context of environmental regulation, especially human health. Our research goals are to: (1) classify the different types of toxicity and the mechanism of action of environmental chemicals using analysis of gene expression induced by exposure to chemicals, as well as toxicological data; (2) develop analysis systems that categorize multiple profiling based on multidimensional information on the effects of chemicals on rodents and human health, including analysis of the relationship between toxicity data and related diseases; (3) create a high-quality categorized index for diverse chemical-based databases and lists of chemicals that are important for environmental regulation; (4) construct Health Effects Alert System (HEALS) that includes other systems to evaluate chemical effects using a variety of databases and algorithms. Further development of structure-activity and structure-toxicity databases as well as toxicity molecular endpoint computerized libraries is required.

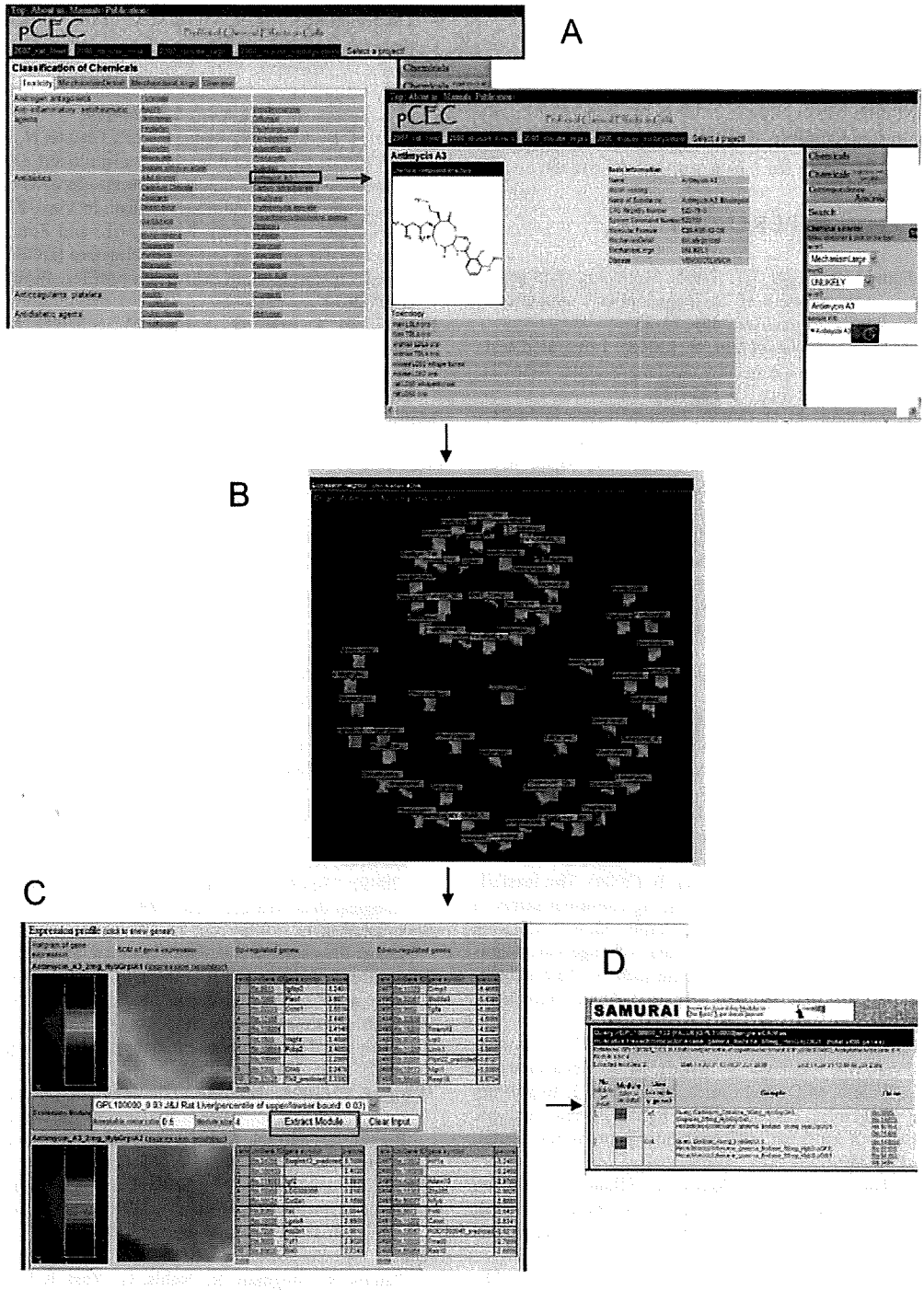

Fig. 5. Chemical selector. This query option is categorized by the chemical at three levels. The user chooses options in Levels 1 and 2 and chooses the desired chemical in Level 3.
Level 1 indicates the categories of the different types of chemical effects, which are parts of the process during apparent toxic effects and disease outcomes from the first target on the cellular event as shown in Fig. 1. Level 2 indicates more detailed categories of the mechanism of action and toxicity pathways than those of level 1. Level 3 indicates chemicals that are related to the structure of the mother chemical.

ACKNOWLEDGMENTS

We thank Prof. Richard E. Peterson (the University of Wisconsin-Madison) for his comments on this work. This work was supported by an Environmental Risk Program of the National Institute for Environmental Studies.

REFERENCES

Aardema, M.J. and MacGregor, J.T. (2002): Toxicology and genetic toxicology in the new era of "toxicogenomics": impact of "-omics" technologies. Mutat. Res., **499**, 13-25.

Ahlers, J., Stock, F. and Werschkun, B. (2008): Integrated testing and intelligent assessment-new challenges under REACH. Environ. Sci. Pollut. Res. Int., **15**, 565-572.

Andersen, M.E. and Krewski, D. (2009): Toxicity testing in the 21st century: bringing the vision to life. Toxicol. Sci., **107**, 324-330.

Barrett, T., Suzek, T.O., Troup, D.B., Wilhite, S.E., Ngau, W.C., Ledoux, P., Rudnev, D., Lash, A.E., Fujibuchi, W. and Edgar, R. (2005): NCBI GEO: mining millions of expression profiles--database and tools. Nucleic. Acids. Res., **33**, D562-566.

Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., Rudnev, D., Evangelista, C., Kim, I.F., Soboleva, A., Tomashevsky, M. and Edgar, R. (2007): NCBI GEO: mining tens of millions of expression profiles--database and tools update. Nucleic. Acids. Res., **35**, D760-765.

Benigni, R., Netzeva, T.I., Benfenati, E., Bossa, C., Franke, R., Helma, C., Hulzebos, E., Marchant, C., Richard, A., Woo, Y.T. and Yang, C. (2007): The expanding role of predictive toxicology: an update on the (Q)SAR models for mutagens and carcinogens. J. Environ. Sci. Health C Environ. Carcinog. Ecotoxicol. Rev., **25**, 53-97.

Bushnell, P.J., Boyes, W.K., Shafer, T.J., Bale, A.S. and Benignus, V.A. (2007): Approaches to extrapolating animal toxicity data on organic solvents to public health. Neurotoxicology, **28**, 221-226.

Fentem, J., Chamberlain, M. and Sangster, B. (2004): The feasibility of replacing animal testing for assessing consumer safety: a suggested future direction. Altern. Lab. Anim., **32**, 617-623.

Fostel, J.M. (2008): Towards standards for data exchange and integration and their impact on a public database such as CEBS (Chemical Effects in Biological Systems). Toxicol. Appl. Pharmacol., **233**, 54-62.

Guyton, K.Z., Kyle, A.D., Aubrecht, J., Cogliano, V.J., Eastmond, D.A., Jackson, M., Keshava, N., Sandy, M.S., Sonawane, B., Zhang, L., Waters, M.D. and Smith, M.T. (2009): Improving prediction of chemical carcinogenicity by considering multiple mechanisms and applying toxicogenomic approaches. Mutat. Res., **681**, 230-240.

Judson, R., Richard, A., Dix, D., Houck, K., Elloumi, F., Martin, M., Cathey, T., Transue, T.R., Spencer, R. and Wolf, M. (2008): ACToR--Aggregated Computational Toxicology Resource. Toxicol. Appl. Pharmacol., **233**, 7-13.

Judson, R., Richard, A., Dix, D.J., Houck, K., Martin, M., Kavlock, R., Dellarco, V., Henry, T., Holderman, T., Sayre, P., Tan, S., Carpenter, T. and Smith, E. (2009): The toxicity data landscape for environmental chemicals. Environ. Health Perspect., **117**, 685-695.

Kavlock, R.J., Ankley, G., Blancato, J., Breen, M., Conolly, R., Dix, D., Houck, K., Hubal, E., Judson, R., Rabinowitz, J., Richard, A., Setzer, R.W., Shah, I., Villeneuve, D. and Weber, E. (2008): Computational toxicology--a state of the science mini review. Toxicol. Sci., **103**, 14-27.

Krewski, D., Andersen, M.E., Mantus, E. and Zeise, L. (2009): Toxicity testing in the 21st century: implications for human health risk assessment. Risk Anal., **29**, 474-479.

Lambert, C.B., Spire, C., Renaud, M.P., Claude, N. and Guillouzo, A. (2009): Reproducible chemical-induced changes in gene expression profiles in human hepatoma HepaRG cells under various experimental conditions. Toxicol. In Vitro, **23**, 466-475.

Luo, F., Khan, L., Bastani, F., Yen, I.L. and Zhou, J. (2004): A dynamically growing self-organizing tree (DGSOT) for hierarchical clustering gene expression profiles. Bioinformatics, **20**, 2605-2617.

Masciocchi, J., Frau, G., Fanton, M., Sturlese, M., Floris, M., Pireddu, L., Palla, P., Cedrati, F., Rodriguez-Tomé, P. and Moro, S. (2009): MMsINC: a large-scale chemoinformatics database. Nucleic. Acids. Res., **37**, D284-290.

Mattingly, C.J. (2009): Chemical databases for environmental health and clinical research. Toxicol. Lett., **186**, 62-65.

Maxwell, G., Aleksic, M., Aptula, A., Carmichael, P., Fentem, J., Gilmour, N., Mackay, C., Pease, C., Pendlington, R., Reynolds, F., Scott, D., Warner, G. and Westmoreland, C. (2008): Assuring consumer safety without animal testing: a feasibility case study for skin sensitisation. Altern. Lab. Anim., **36**, 557-568.

National Research Council (2007): Toxicity Testing in the 21st Century: A Vision and a Strategy. National Academies Press, Washington, DC.

Okada, Y., Okubo, K., Fujibuchi, W. and Horton, P. (2007): Exhaustive Search Method of Gene Expression Modules and Its Application to Human Tissue Data. IAENG International Journal of Computer Science, **34**, 119-126.

Schmidt, U., Struck, S., Gruening, B., Hossbach, J., Jaeger, I.S., Parol, R., Lindequist, U., Teuscher, E. and Preissner, R. (2009): SuperToxic: a comprehensive database of toxic compounds. Nucleic. Acids. Res., **37**, D295-299.

Schneider, K., Schwarz, M., Burkholder, I., Kopp-Schneider, A., Edler, L., Kinsner-Ovaskainen, A., Hartung, T. and Hoffmann, S. (2009): "ToxRTool", a new tool to assess the reliability of toxicological data. Toxicol. Lett., **189**, 138-144.

Shoemaker, R.H. (2006): The NCI60 human tumour cell line anticancer drug screen. Nat. Rev. Cancer, **6**, 813-823.

Tomasulo, P. (2005): ITER- International Toxicity Estimates for Risk, new TOXNET database. Med. Ref. Serv. Q, **24**, 55-66.

Törönen, P., Kolehmainen, M., Wong, G. and Castrén, E. (1999): Analysis of gene expression data using self-organizing maps. FEBS Lett., **451**, 142-146.

Wang, R.Y. (2008): Medical toxicology and public health-update on research and activities at the Centers for Disease Control and Prevention and the Agency for Toxic Substances and Disease Registry. J. Med. Toxicol., **4**, 143-144.

Waters, M. and Jackson, M. (2008): Databases applicable to quantitative hazard/risk assessment--towards a predictive systems toxicology. Toxicol. Appl. Pharmacol., **233**, 34-44.

Waters, M., Stasiewicz, S., Merrick, B.A., Tomer, K., Bushel, P., Paules, R., Stegman, N., Nehls, G., Yost, K.J., Johnson, C.H., Gustafson, S.F., Xirasagar, S., Xiao, N., Huang, C.C., Boyer, P., Chan, D.D., Pan, Q., Gong, H., Taylor, J., Choi, D., Rashid, A., Ahmed, A., Howle, R., Selkirk, J., Tennant, R. and Fostel, J. (2008): CEBS--Chemical Effects in Biological Systems: a public data repository integrating study design and toxicity data with microarray and proteomics data. Nucleic. Acids. Res., **36**, D892-900.

Waters, M.D. and Fostel, J.M. (2004): Toxicogenomics and systems toxicology: aims and prospects. Nat. Rev. Genet., **5**, 936-948.

Waters, M.D., Olden, K. and Tennant, R.W. (2003): Toxicogenomic approach for assessing toxicant-related disease. Mutat. Res., **544**, 415-424.

Williams-DeVane, C.R., Wolf, M.A. and Richard, A.M. (2009a): DSSTox chemical-index files for exposure-related experiments in ArrayExpress and Gene Expression Omnibus: enabling toxico-chemogenomics data linkages. Bioinformatics, **25**, 692-694.

Williams-Devane, C.R., Wolf, M.A. and Richard, A.M. (2009b): Toward a public toxicogenomics capability for supporting predictive toxicology: survey of current resources and chemical indexing of experiments in GEO and ArrayExpress. Toxicol. Sci., **109**, 358-371.

Woodruff, T.J., Zeise, L., Axelrad, D.A., Guyton, K.Z., Janssen, S., Miller, M., Miller, G.G., Schwartz, J.M., Alexeeff, G., Anderson, H., Birnbaum, L., Bois, F., Cogliano, V.J., Crofton, K., Euling, S.Y., Foster, P.M., Germolec, D.R., Gray, E., Hattis, D.B., Kyle, A.D., Luebke, R.W., Luster, M.I., Portier, C., Rice, D.C., Solomon, G., Vandenberg, J. and Zoeller, R.T. (2008): Meeting report: moving upstream-evaluating adverse upstream end points for improved risk assessment and decision-making. Environ. Health Perspect., **116**, 1568-1575.

Wullenweber, A., Kroner, O., Kohrman, M., Maier, A., Dourson, M., Rak, A., Wexler, P. and Tomljanovic, C. (2008): Resources for global risk assessment: the International Toxicity Estimates for Risk (ITER) and Risk Information Exchange (RiskIE) databases. Toxicol. Appl. Pharmacol., **233**, 45-53.

Xu, Y., Olman, V. and Xu, D. (2001): Minimum spanning trees for gene expression data clustering. Genome Inform., **12**, 24-33.

Xu, Y., Olman, V. and Xu, D. (2002): Clustering gene expression data using a graph-theoretic approach: an application of minimum spanning trees. Bioinformatics, **18**, 536-545.

Open Access

Importance of CDK7 for G1 Re-Entry into the Mammalian Cell Cycle and Identification of New Downstream Networks Using a Computational Method

Hideko Sone[*,1,2,§], Tomokazu Fukuda[3,§], Hiroyoshi Toyoshiba[1,§], Takeharu Yamanaka[1], Fred Parham[1] and Christopher J. Portier[1]

[1]*Laboratory of Computational Biology and Risk Analysis, National Institute of Environmental Health Sciences, 111 T.W. Alexander Drive, Research Triangle Park, NC 27709, USA*

[2]*Health Effects Team, National Institute for Environmental Studies, 16-2 Onogawa, Tsukuba 305-8506, Japan*

[3]*Laboratory of Animal breeding and Genetics, Graduate school of Agricultural Science, Tohoku University, Tsutsumidori-amamiyamachi 1-1 Aoba-ku, Sendai 981-8555, Japan*

Abstract: Many of the key molecules in cell cycle progression (e.g. pRB, cyclin complexes) and their basic interactions are oncogene or tumor suppressor genes, which are well characterized in the clinical and experimental analysis. However, there are still unknown mechanisms for the cell cycle regulation, which is critical step for the progression of the cancer development. Especially it is not fully understood how the cells move to G1 phase from quiescent G0 phase in the mammalian cells. To find out the new gene networks associated with the two transition of the mammalian cell cycle (G0 to G1 and G1 to S phase), we analyzed the linkages between 39 representative oncogene or tumor suppressor genes, which related to the cell cycle regulation, with gene expression sets obtained from the publicly opened microarray data for mouse embryonic fibroblasts that synchronized by the serum starvation or hydroxyurea treatment. Analyses with a qualitative algorithm based on Bayesian networks that assume a log-linear relationship between genes have applied, and newly found networks were validated. Results highlighted the importance of two master genes, *Cdk7* and *Cdkna2* for the re-entry to G1 from G0, and suggested a new network connection from *Cdk7* to downstream molecules, including the *EGF* receptor and *N-myc*. Introduction of a recombinant *Cdk7* with retrovirus decreased endogenous EGFR and N-myc protein levels. The results supported the computational prediction of the *Cdk7* network. Taken together, these result showed the existence of new regulating pathway from Cdk7 to Egfr and N-myc, suggesting this analytical approach provides an assessment of regulatory networks in complex mammalian cells, and the process of the carcinogenesis.

Keywords: Gene network, cell cycle, Cdk7, mammalian, Bayesian theory.

INTRODUCTION

Cell division and tissue growth represent two of the most fundamental biological processes and play essential roles in development, aging, cancer [1, 2], and many other diverse events. Although gene transcripts have been comprehensively catalogued in yeast, much work remains to be done in higher organisms. Especially, for tumor progression, the gene networks underlying the regulation of the cell cycle are not well understood in cancer cells or the initiated precancerous cells. Several groups have utilized microarrays to perform serial analyses of gene expression during cellular replication in normal or cancer human and mouse cell lines [3-6]. These microarray data have been analyzed using clustering approaches such as hierarchical clustering and k-means to identify stage-specific or co-regulated genes through each phase of the cell cycle. However, these methodologies can only identify genes with expression levels that correlate over time, and the network dynamics of the cell cycle is not yet fully understood.

Integrated and networked functions in mammalian cells can be identified and quantified through the use of a computational model. Efforts to systematically define specific gene network structures to further understand the functions and dynamics of each gene and its protein products have lead to a new generation of *in silico* analysis tools that use diagrams to depict the logical relationships between genes [7-9]. To infer unknown gene networks from microarray gene expression data, the methods adopted need to incorporate the two different aspects of Bayesian models and associated validation tools. The application of these biostatistical methods has the potential to elucidate unknown mechanisms underlying the key regulatory systems of mammalian cells [10-12].

The regulatory mechanisms for the G0 quiescent stage of the cell cycle remain largely unknown. For the efficient progression from the G0 to G1 phase, the protein level of the p27/kip1 is known to have a important role in T cell from *in vitro* study and a knockout mouse study [13, 14]. In the normal cells, the protein level of p27 is high during G0 phase

*Address correspondence to this author at the Research Center for Environmental Risk, National Institute for Environmental Studies, 16-2 Onogawa, Tsukuba 305-8506, Japan; Tel: +81.298.50.2464; Fax: +81.298.50.2546; E-mail: hsone@nies.go.jp

[§]These authors equally contributed to this work.

1876-3901/10 2010 Bentham Open

but decreases rapidly on the entry to G1 [14, 15]. The degradation of p27 is controlled by an SCF complex, which involves SKP2 [16, 17]. Although these findings for G0-G1 regulation have had a significant impact, it is not clear whether these mechanisms can be applied to the all type of cells and tissues. For cancer therapeutics, the G0-G1 transition of the cell cycle has been a strong target to prevent tumor growth and progression [18-20].

In our current study, we employed the gene datasets from the publicly opened microarray data for the mouse fibroblasts, which synchronized with the serum starvation and hydroxyarea, which are the study of the transition from a quiescent state into the cell cycle in mouse embryonic fibroblast (MEF) cells reported by Ishida *et al.* [4]. In order to elucidate new gene networks related to the progression of the cell cycle, the gene expression datasets were analyzed using a series of approaches in which putative network structures are elucidated using Bayesian networks. These approaches involve a likelihood-based selection algorithm to qualitatively infer the identity of the network structure [21] and a quantitative algorithm involving a Markov chain Monte Carlo (MCMC) method [22, 23] is then used to quantify the structure. The identified interactions between genes that are based upon these predicted gene networks were then validated using a retrovirus expression system.

MATERIALS AND METHODS

Microarray Data Sets

Previously published mouse embryonic fibroblast (MEF) cell microarray datasets were used in our analyses [4]. Briefly, the cells were synchronized by either serum starvation or hydroxyurea treatment. We used the data sets obtained from the serum starved cells for the analysis of re-entry into G1 from G0 (0, 6, 12, 15, 18, 21, 24 hours after serum starvation), and those from the hydroxyurea exposed fibroblasts for the G1-S analysis (0, 3, 6, 9, 12, 15, 18 hours after the treatment). The detailed methods used to obtain these microarray data have been previously described [4].

Selection of the Subset Database

The original gene expression data, comprising about 6437 genes, were screened for genes that showed at least a 2.0-fold change (up- or down-regulation) using GenMAPP [24]. The distribution and frequency of the fold changes (relative to the time 0) at each time point were analyzed by MAPFinder 1.0 beta, an accessory tool of GenMAPP, to identify the optimal biological maps. From this collection of maps, we selected those related to cell cycle processes that had a "z" score greater than 1.95 (the z score represents the difference between the observed number of genes meeting the criteria and the expected number of genes meeting the criteria in each map based on gene ontology). As detailed in Table **1**, 10 maps were selected based on gene ontology (denoted MAPP) and the relationship to the cell cycle. A subset of 39 genes was chosen from among the MAPP maps selected (Table **2**). The abbreviated names of the genes that were analyzed in this report are presented according to the displays listed in GenMAPP.

Mathematical Models

We applied the expression-associated network modeling method previously developed by Yamanaka *et al.* [21] to the

Table 1. List of Maps with More than 1.95 Z Score Selected from Maps Analyzed by MAPFinder. Maps are the Database from Mouse Biological Processes that are Contain in GenMAPP

MAPP Name	A	B	C	D	E	R	z Score	Time Point
Mm_cell cycle	4	15	104	26.7	14.4	95	2.014	18h
	4	15	104	26.7	14.4	87	2.211	21h
Mm_cell cycle arrest	2	2	11	100	18.2	92	4.175	12h
	2	2	11	100	18.2	92	4.175	15h
	2	2	11	100	18.2	109	3.795	18h
	2	2	11	100	18.2	118	3.626	21h
	2	2	11	100	18.2	121	4.335	24h
Mm_cell cycle control	15	48	124	31.2	38.7	132	3.205	12h
	9	48	124	18.8	38.7	87	2.102	21h
Mm_cell growth and or maintenance	17	55	153	30.9	35.9	132	3.357	12h
Mm_cell growth	3	7	16	42.9	43.8	118	2.317	21h
Mm_cell proliferation	3	4	28	75	14.3	95	4.177	18h
Mm_G1 S transition of	1	1	5	100	20	101	2.771	6h
Mitotic cell cycle								
Mm_mitosis	2	6	23	33.3	26.1	87	1.95	21h
Mm_mitotic cell cycle	1	1	7	100	14.3	101	2.771	6h
Mm_M phase of mitotic cell cycle	2	6	23	33.3	26.1	87	1.95	21h

A, the number of genes meeting the criterion in this specific MAPP; B, the total number of genes measured in this specific MAPP; C, Number on MAPP; D, Percent Changed; E, Percent present; N, the total number of genes measured (= 894), R, the total number of distinct genes meeting the criterion. Criterions were set at > 2.0 or < 0.5 of the expression ratio. Each time point means a sampling time after serum starvation. Z Score= $(A-B*R/N)/\sqrt{(B(R/N)(1-R/N)(1-B-1/N-1))}$.