

TG-GATEsデータの判別結果-3

			low			middle			high			hr
			2	8	24	2	8	24	2	8	24	
			00058	MCT	monocrotaline	NA	NA	NA	NA	0.134	0.123	
00059	VA	vitamin A	NA	NA	NA	NA	0.189	0.202	NA	0.357	0.224	
00060	TAC	tacrine	NA	NA	NA	NA	0.072	0.230	NA	0.057	0.228	
00061	MXS	moxisylyte	NA	NA	NA	NA	0.127	0.185	NA	0.080	0.313	
00062	IPA	iproniazid	NA	NA	NA	NA	0.098	0.300	NA	0.017	0.102	
00064	CMP	chloramphenicol	NA	NA	NA	NA	0.022	0.035	NA	0.009	0.022	
00065	NFZ	nitrofurazone	NA	NA	NA	NA	0.166	0.091	NA	0.091	0.167	
00069	IMI	imipramine	NA	NA	NA	NA	0.106	0.105	NA	0.172	0.170	
00070	AMT	amitriptyline	NA	NA	NA	NA	0.100	0.162	NA	0.045	0.081	
00072	IBU	ibuprofen	NA	NA	NA	NA	0.252	0.236	NA	0.858	0.979	
00074	QND	quinidine	NA	NA	NA	NA	0.077	0.256	NA	0.026	0.355	
00078	FUR	furosemide	NA	NA	NA	NA	0.090	0.334	NA	0.009	0.255	
00080	CPP	chlorpropamide	NA	NA	NA	NA	0.093	0.166	NA	0.139	0.485	
00081	NIC	nicotinic acid	NA	NA	NA	NA	0.061	0.237	NA	0.016	0.756	
00082	EME	erythromycin ethylsuccinate	NA	NA	NA	NA	0.026	0.226	NA	0.049	0.074	
00085	FAM	famotidine	NA	NA	NA	NA	0.062	0.146	NA	0.039	0.127	
00086	RAN	ranitidine	NA	NA	NA	NA	0.178	0.074	NA	0.021	0.063	
00090	CHL	chlorpheniramine	NA	NA	NA	NA	0.092	0.220	NA	0.074	0.332	
00091	NIF	nifedipine	NA	NA	NA	NA	0.110	0.149	NA	0.020	0.046	
00092	DIL	diltiazem	NA	NA	NA	NA	0.236	0.084	NA	0.017	0.019	
00093	TAN	tannic acid	NA	NA	NA	NA	0.088	0.191	NA	0.029	0.129	
00094	CAP	captopril	NA	NA	NA	NA	0.077	0.282	NA	0.026	0.440	
00095	ENA	enalapril	NA	NA	NA	NA	0.072	0.190	NA	0.117	0.943	
00096	TEO	theophylline	NA	NA	NA	NA	0.012	0.002	NA	0.001	0.001	
00097	CAF	caffeine	NA	NA	NA	NA	0.129	0.080	NA	0.019	0.245	
00098	PAP	papaverine	NA	NA	NA	NA	0.013	0.062	NA	0.008	0.001	
00099	PEN	penicillamine	NA	NA	NA	NA	0.169	0.423	NA	0.028	0.722	
00100	SUL	sulindac	NA	NA	NA	NA	0.060	0.606	NA	0.003	0.000	
00102	DIS	disopyramide	NA	NA	NA	NA	0.044	0.582	NA	0.008	0.157	
00103	MEX	mexiletine	NA	NA	NA	NA	0.117	0.236	NA	0.047	0.170	

NSAID

数字はprobability, ピンク: ポジティブ判定(probability>0.5)

TG-GATEsデータの判別結果-4

			low			middle			high			hr
			2	8	24	2	8	24	2	8	24	
			00104	TIO	tiopronin	NA	NA	NA	NA	0.093	0.136	
00108	ACZ	acetazolamide	NA	NA	NA	NA	0.154	0.262	NA	0.171	0.206	
00109	DSF	disulfiram	NA	NA	NA	NA	0.147	0.090	NA	0.024	0.202	
00110	PMZ	promethazine	NA	NA	NA	NA	0.061	0.061	NA	0.019	0.207	
00113	COL	colchicine	NA	NA	NA	NA	0.009	0.001	NA	0.007	0.000	
00114	TLB	tolbutamide	NA	NA	NA	NA	0.367	0.913	NA	0.448	0.953	
00115	SLP	sulpiride	NA	NA	NA	NA	0.213	0.775	NA	0.034	0.261	
00116	ACA	acarbose	NA	NA	NA	NA	0.095	0.112	NA	0.127	0.182	
00117	SST	simvastatin	NA	NA	NA	NA	0.206	0.291	NA	0.167	0.656	
00118	AJM	ajmaline	NA	NA	NA	NA	0.215	0.195	NA	0.064	0.205	
00119	DTL	dantrolene	NA	NA	NA	NA	0.122	0.112	NA	0.160	0.151	
00121	CPM	clomipramine	NA	NA	NA	NA	0.033	0.134	NA	0.039	0.086	
00123	TBF	terbinafine	NA	NA	NA	NA	0.099	0.085	NA	0.048	0.047	
00124	MLX	meloxicam	NA	NA	NA	NA	0.084	0.110	NA	0.126	0.336	
00125	LNX	lornoxepam	NA	NA	NA	NA	0.158	0.120	NA	0.205	0.241	
00126	CLM	chlormadinone	NA	NA	NA	NA	0.048	0.078	NA	0.034	0.006	
00127	DNZ	danazol	NA	NA	NA	NA	0.104	0.111	NA	0.003	0.000	
00130	BZD	benzodaron	NA	NA	NA	NA	0.672	0.885	NA	0.722	0.952	
00131	ETP	etoposide	NA	NA	NA	NA	0.110	0.196	NA	0.019	0.154	
00135	ETH	ethionamide	NA	NA	NA	NA	0.106	0.347	NA	0.025	0.119	
00136	NIM	nimesulide	NA	NA	NA	NA	0.317	0.743	NA	0.012	0.001	
00137	ETN	ethanol	NA	NA	NA	NA	0.030	0.072	NA	0.127	0.054	
00138	PCT	phenacetin	NA	NA	NA	NA	0.073	0.121	NA	0.044	0.119	
00139	BCT	bucetin	NA	NA	NA	NA	0.067	0.096	NA	0.194	0.161	
00140	NPA	phenylanthranilic acid	NA	NA	NA	NA	0.233	0.144	NA	0.694	0.966	
00144	AAF	acetamidofluorene	NA	NA	NA	NA	0.061	0.168	NA	0.014	0.154	
00145	DEN	nitrosodiethylamine	NA	NA	NA	NA	0.108	0.142	NA	0.020	0.157	
00146	TCP	ticlopidine	NA	NA	NA	NA	0.121	0.128	NA	0.069	0.221	

ラット
in vitro
でも陽性
 NSAID

数字はprobability, ピンク: ポジティブ判定(probability>0.5)

外部データの判別結果

化合物名	用量 (μM)	時点 (hr)	Probability	判定 結果	GEO ID	TGP用量 (μM)	TG-GATEsデータで 陽性判定
WY-14643	50	6	0.876	P	GSE17251	6, 30, 150	2hr-H, 8hr-L&M&H, 24hr-M&H
WY-14643	50	24	0.992	P	GSE17251		
Atorvastatin	10	24	0.247	N	GSE24187	simvastatin:	simvastatin:
Rosuvastatin	10	24	0.240	N	GSE24187	6, 30	24hr-H
Rifampicin	5	24	0.122	N	GSE24187	2.8, 14, 70	なし

NCBI GEOから各データセットのCELファイルをダウンロードし、TG-GATEsのデータと同様にデータ処理後、判別モデルにかけた。

ピンク:陽性化合物, 水色:陰性化合物, クリーム色:典型的PPAR α アゴニストではないが陽性が期待された化合物

スタチンでの陽性判定は出なかったが、用量不足の可能性も。WYとRIFは正しく判別された。



外部データにも適用可能。

小括

●ラット肝臓の*in vivo*, *in vitro*共通に利用可能なグレードIVのPPAR α アゴニストマーカーについて検証を行った。その過程で、PCAによる評価からSVMによる判別を用いた方法に変更し、*in vivo*, *in vitro*それぞれについて判別モデルを作成した。

●ヒト肝細胞についても発現変動遺伝子抽出後判別モデルを作成し、マーカーを選抜した。

●判別結果はラットについては*in vivo*, *in vitro*ともに非常に良好だった。ヒトはラットほどではないが、十分利用可能なレベルであった。いずれも外部データについても利用可能であった。

●マーカー遺伝子にはラットとヒトで共通のものはないが、いずれもPPARとの関連が既知の脂質関連遺伝子が多く含まれた。また遺伝子発現変動としては一部は共通している。

まとめ

本分担研究で検証を行った各種マーカーはプラットフォームやデータ処理法を合わせる必要がある場合が多いものの、ほとんどが外部データについても有効であった。

これらのマーカーを組み合わせると一度の実験で多くのフェノタイプについて推定でき、新規化合物の毒性プロファイリングなどに有用と考えられる。

また未だ毒性メカニズムが明らかではない化合物のメカニズム解析にも役立つと考えられる。

創薬基盤としての分子毒性学研究

研究分担者 菅野 純 国立医薬品食品衛生研究所・毒性部長

研究要旨

実験動物のデータをヒトへ外挿し毒性予測をする際、外来化学物質代謝機能の種差が問題となる。SXR(マウスでは PXR)は代謝酵素の誘導を制御し、代謝機能調節に関わる重要な受容体であるが、そのリガンド選択性に種差が大きいことが知られている。

本分担研究では、ヒト受容体 SXR のリガンド選択性を導入しつつ、全身諸臓器の毒性検討が可能且つ、その受容体の発現調節機構が生理的なマウスモデルを作出するために、遺伝子相同組換え技術を用い、hSXR のリガンド結合ドメイン(LBD)のみをマウスのLBDと入れ替えたノックインマウス(hSXRki mouse)を作製した。本マウスでは、マウスゲノム内に多数存在する cis-element への結合パターンに変化が生じないことに加え、転写開始点上流の配列が維持されることにより、マウス SXR (mSXR)本来の発現組織分布が hSXRki においても保たれることが見込まれる。

本年度は、本マウスの SXR リガンドに対する反応のヒト型化程度を評価し、本マウスが今後毒性研究において有用なツールとして活用しうるかどうか検証した。マウス SXR に対する選択性が高いリガンドとして Pregnenolone-16alpha-carbonitrile (PCN)を、ヒト SXR に対する選択性が高いリガンドとして Rifampicin (RIF)を用い、SXR が機能していることがよく知られている肝、小腸に加え、脳(海馬)、胸腺、肺、腎、精巣の合計 7 臓器について、Perclorome 法を適用したトランスクリプトーム解析を実施した。

その結果、肝、小腸において、CYP3A11 や CES6 といった既知の SXR 標的遺伝子の発現誘導パターンがヒト型化していることが確認され、本マウスが毒性研究における有用なツールであることが示された。一方、肝、小腸以外の臓器においては、SXR 活性化パターンを示す遺伝子はごく少ないことが分かったが、SXR の発現が肝、小腸に限られるわけでは無いことを踏まえると、他の臓器における機能が予想されたため、米国カリフォルニア大アーバイン校と共同研究を行い、白血球細胞における作用を見出した。

hSXRki を全身に生理的発現する本マウスは、ヒト SXR に対するリガンド特異性を反映するマウスとして、今後毒性研究における有用なツールとしての活用が期待される。

A. 研究の目的

経口摂取された化学物質は消化管から吸収され、門脈を経由し肝で代謝を受けた後、或いは乳び管を経由して大循環に入る。小腸や肝には齧歯類では PXR、ヒトでは Steroid X Receptor (SXR)とも呼ばれる核内

受容体(hSXR)が外来異物センサーとして存在し、化学物質を極性物質に代謝し体外への排出を促進する酵素 CYP3A 等の発現誘導を制御することが知られている。

CYP3A は複数の医薬品の代謝における相互干渉の問題において特に重要視されて

おり、治療効果減少や、毒性増強に関わるとされている。この様な場合を含め、mSXR と hSXR のリガンド選択性に大きな種差があることから、この受容体系の反応が毒性を左右する物質については、齧歯類を用いた試験結果からヒトでの毒性の予測は行えない可能性が指摘されている。選択性の高いリガンドの実例は RIF で、hSXR 系は活性化するが mSXR は活性化しない。このタイプの物質の毒性評価は通常のマウスを用いた試験からは困難であることになる。逆の事例に PCN 等の物質が知られている。

ヒト型の反応を示すマウスの作製は、既に試みられている。初期のものに、hSXR 遺伝子をアルブミンプロモータ下に肝に強制発現させるトランスジェニックマウスがある。この場合、発現部位が肝に限定されることと、mSXR も発現していることから、ヒト型の反応性の有無を得るに留まるものとなっている。

mSXR と hSXR を比較すると、DNA 結合ドメイン (DBD) のアミノ酸配列は 95% 保存されているのに対し、LBD は 77% にとどまっている。ヒトとラットのキメラタンパク質を用いた解析により、LBD 内に種差を規定する領域が存在することが明らかにされており、ヒトの Leu308 とラット/マウスの Phe305 が重要であることが示されている。hSXR の結晶構造を元にした homology modeling により、これらのアミノ酸は located within or neighboring the flexible loop that forms part of the pore to the ligand-binding cavity であることが示唆されている。実際に、このアミノ酸を置換することで Rifampicin に対する種特異性が逆転する。これらのことから、mSXR の LBD を hSXR のものに変えることにより、ヒト型の活性化を示すマウスが作製されることは想像に難しくなく、実際

に、幾つかのノックインマウスが報告されている。しかし、それらは、相同組換え体の設計の問題から、hSXR の発現臓器が限定されている等の人工的な要素が加わっており、マウスを用いての、全身を対象とした網羅性を重視する毒性評価に利用する為には十分な特性を有していないことが考えられた。

そこで、我々は LBD の部分のみをヒト型化し、それ以外の部分はマウス型のままとするノックインマウスを作製することとした。これにより、マウスゲノム内に多数存在する cis-element への結合パターンに変化が生じないこと、及び、転写開始点上流の配列が維持されることから、マウス本来の組織発現パターンとその誘導様態が保たれることを見込むものである。LBD のみをヒト型化したキメラタンパク質 (hSXR) をコードする cDNA を mSXR 遺伝子座の Exon3 に knock-in したマウスを作製し、そのマウスが実際に hSXR 特異的リガンドに対する反応を全身性に示したことを報告する。このマウスは hSXR を介した全身毒性の研究を行うための有用なツールとなると期待される。

B. 研究の方法

(1) hSXRki マウス

マウス LBD の代わりにヒト LBD を持つ hSXR chimeric protein を発現するノックインマウスを用いた。ES 相同組み換え体取得に利用したネオマイシン耐性遺伝子は CAG-Cre マウスとの交配により除去し、得られたマウスを C57BL/6 CrSlc に少なくとも 6 回バッククロスし、ホモ体を得て本解析に供した。

(2)リガンド反応性検討

Vehicleとしてcorn oil+0.1%DMSOを用い、PCNは20、70、200mg/kg、RIFは10、30、100mg/kgをhSXRkiマウス及び野生型マウスの雄に単回強制経口投与し、投与8時間後に脳(海馬)、胸腺、肺、肝、腎、小腸、精巣の7臓器を採取した。肝は小片を5mm径の皮膚生検用トレパンにて打ち抜き、RNA laterに浸漬した。小腸は粘膜上皮のみを搔爬採取しRNA laterに浸漬した。なお、臓器採取は十分に手順の効率化を図り複数人のチームにより実施することで時間を2分半~3分/匹程度に維持し、mRNA発現量の日内変動及び変性の影響を排除するよう努めた。

PCBの実験は、本マウスを米国カリフォルニア大学アーバイン校のBruce Blumbergに供与し、先方で実施した。

(3)Percellome手法を適用したTotal RNAの分離精製

組織は採取後すみやかにRNA later (Ambion社)に4°C一晩浸漬し、RNA抽出操作までは-80°Cにて保存した。抽出に当たっては、RNA laterを除いた後、RNeasyキット(キアゲン社)に添付されるRLT bufferを添加し、ジルコニアビーズを用いて破碎液を調製した。得られた破碎液の10µLを取り、DNA定量蛍光試薬Picogreenを用いてDNA含量を測定した。DNA含量に応じ、あらかじめ臓器毎に設定した割合でGSC (Graded dosed spike cocktail; Bacillus subtilis由来RNA 5種類を、濃度を振って混合した溶液)を添加し、TRIZOLにより水層を得、RNeasyキットを用いて全RNAを抽出した。100ngを電気泳動しRNAの純度及び分解の有無を検討した。

(4)GeneChip解析

全RNA 5 µgを取り、アフィメトリクス社のプロトコルに従い、T7プロモーターにより付されたオリゴdTプライマーを用いて逆転写しcDNAを合成した。得られたcDNAの第二鎖を合成し、二本鎖DNAとした。次にT7 RNAポリメラーゼ(ENZO社キット)を用い、ビオチン化UTP、CTPを共存させつつcRNAを合成した。cRNAはAffymetrix社キットにて精製後、300-500bpとなるよう断片化し、GeneChipターゲット液とした。GeneChipにはMouse Genome 430 2.0を用いた。ハイブリダイゼーションは45°Cにて18時間行い、バッファーによる洗浄後、phycoerythrin (PE)ラベルストレプトアビジンにて染色し、専用スキャナーにて発現データを得た。得られた発現データは、当方が開発した各種ソフトウェアを用いてコピー数に変換(Percellome法)し解析に供した。

(5)定量的RT-PCR(Percellome PCR)

Percellome法(BMC genomics 2006, 7, 64)に従い、以下の手順で実施した。組織破碎液中のDNA含量を、DNA定量蛍光試薬Picogreenを用いて測定し、それに応じてspike RNA cocktailを添加した後に精製したtotal RNAを、逆転写反応にてcDNAとし、Ces6に対するprimer

(FW:GGAGCCTGAGTTCAGGACAGAC、RV:ACCCTCACTGTTGGGGTTC)、

Cyp3a11に対するprimer

(FW:CAGCTTGGTGCTCCTCTACC、RV:TCAAACAACCCCATGTTTT)、及び、Bacillus subtilis由来RNA 5種類のspike RNAに対するprimerを用いて定量RT-PCRを行い、spike RNA量を基準にCES6のコピ

一数を算出した。

(6) *In situ* hybridization 解析

CYP3A11 検出用の DIG ラベル cRNA プロブを作製するために、FW:GATTGGTTTTGATGCCTGGT、RV:CAAGAGCTCACATTTTTCATCA を用い、肝 cDNA を鋳型に PCR 増幅し、配列を確認後、T7 RNA polymerase の promoter を含むリンカーを Block-iT T7-TOPO (Invitrogen, Carlsbad, CA) キットにより付加し、2nd PCR を行い、sense、antisense 用の鋳型を調製し、DIG RNA labeling kit (Roche Diagnostics, Germany) を用いて、DIG ラベルされた cRNA probe を作製した。パラフィン包埋切片に対する *in situ* hybridization は、HistoVT One (nacalai tesque, Japan) 中で 98°C、15 分間賦活化する処理を加え、定法に従って実施した。

(倫理面への配慮)

動物飼育、管理、動物実験に当たっては、国立医薬品食品衛生研究所・動物実験委員会の制定になる国立医薬品食品衛生研究所・動物実験の適正な実施に関する規程 (平成 19 年 4 月版) に則って行った。動物はすべてエーテル深麻酔下で大動脈からの脱血により屠殺し、動物に与える苦痛は最小限に留めた。

C. 研究結果

(1) 各臓器における SXR 発現

hSXR chimeric protein の組織発現パターンを定量 RT-PCR によって検討した。定量 RT-PCR 手法は細胞当たりのコピー数に変換可能な Percellome 手法を用いた。

検討臓器は、脳、胸腺、心、肺、肝、胃、脾、腎、小腸及び精巣の 10 臓器とした。その結果、発現パターンは野生型本来の SXR のパターンと一致し、肺、肝、胃、腎、小腸で 2 コピー以上の発現が認められた (Figure 1)。また、1 コピー以上の発現が精巣において認められた。

(2) 肝、小腸における SXR 応答のヒト型化確認

SXR 応答臓器であることが知られている肝、小腸について、SXR 応答遺伝子である Cyp3a11、Ces6 の定量 RT-PCR 解析 (Percellome 法) を行った。PCN は 20、70、200mg/kg、RIF は 10、30、100mg/kg を単回強制経口投与し、投与 8 時間後に臓器を採取した。その結果、mSXR リガンドである PCN は野生型のみ、hSXR リガンドである RIF は hSXRki のみで発現上昇を示した (Figure 2)。また投与量依存的に発現上昇が見られることも確認された。

よって、少なくとも肝、小腸において SXR が予想通りヒト型反応パターンを示すことが示され、その LBD の種差に応じた活性化が確認された。

(3) *In situ* hybridization による CYP3A11 発現応答解析

Cyp3a11 の発現応答を *in situ* hybridization 法によって可視化した。その結果、定量 RT-PCR 解析の結果通り、*in situ* hybridization 法でも hSXRki マウスがヒト型化反応パターンを示すことが確認された (Figure 3)。

(4) 全身臓器のトランスクリプトーム解析

定量 RT-PCR、*in situ* hybridization 解析により、少なくとも肝、小腸で hSXRki マウスが

ヒト型化パターンを示すことが確認されたので、各種臓器のトランスクリプトーム解析を実施した。PCNは70mg/kg、RIFは30mg/kgについて、Affymetrix社のGeneChip(Mouse Genome 430 2.0)を用いたPercellome法を行った。

その結果、Figure 4Aに示すように、CES6、CYP2B10が肝、小腸の両方でヒト型化反応パターンを示す遺伝子であることが分かった。GSTM3は小腸でヒト型化反応パターンを示すが、肝では応答が見られなかった(Cyp3a11についてはGeneChipでは肝、小腸ともに応答が見られなかったが、その原因はプローブ設計不良であることが知られている)。また、肝、小腸以外でヒト型化反応パターンを示す遺伝子を検索したが、肝、小腸のように明確なパターンを示す遺伝子はほとんど見いだせなかった。Figure 4BにSXR依存的な発現減少パターンを示す3遺伝子の例を示す。HMBOX1は海馬で、HIC2、CYP26B1は腎でSXR活性化に伴い減少している。

(5) 白血球細胞における機能の解析

SXRは炎症性サイトカインによる細胞内シグナル伝達転写因子のNFkBに結合し、相互抑制を行うことが明らかになっている。また、Figure1に示したように、SXRの発現は肝、小腸に留まるものではなく、様々な臓器で確認されている。これらを踏まえるとSXRに肝、小腸以外の臓器での機能があると考えることが妥当であるので、米国カリフォルニア大アーバイン校のBruce Blumbergと共同研究を行い、hSXR選択的な作用を有する化学物質による影響を検討した。BlumbergはPCBの一種がヒトSXRを特異的に抑制すること(マウス

SXRは活性化)、SXRノックアウトマウスがB-1 Cell Lymphomaを発症することを見出している。よって、当該PCBをhSXRkiマウスに投与することでB-1 Cell Lymphomaが生じることが想定されたが、これを支持する事象として、短期実験に於いてB-1 lymphocyteが増加することが確認された(論文投稿準備中のためdata not shown)。

D. 考察

我々が作製したhSXRkiマウスは、野生型マウスのSXR発現と同様の臓器発現分布を示し、少なくとも肝、小腸においてはSXRのヒト型反応パターンを示すことを確認した。その結果は論文として発表しており(J Toxicol Sci. 2012, vol37, No2, p373-380)、今後ヒトへの外挿を念頭に置いた毒性研究で本マウスが有用なツールとなると考えられる。

その際、SXRの機能は当初予想されていた外来化学物質代謝系における機能に留まらないことを考慮する必要がある。NFkBなど他の転写因子との相互作用等を介した未知の機能がありうることを想定すべきである。

E. 結論

本研究により、我々が独自に作製した、SXRのLBDのみをヒトSXRの配列で置き換えたhSXRkiマウスの化学物質反応性がヒト型化していること、全身を対象とした化学物質影響解析に有用であることを確認した。

本マウスは今後ヒトSXRリガンド特異性を反映した毒性研究を行うための有用なツールとして活用されることが期待される。

F. 健康危険情報

なし

G. 研究発表

1. 論文発表

Katsuhide Igarashi, Satoshi Kitajima, Ken-ichi Aisaki, Kentaro Tanemura, Yuhji Taquahashi, Noriko Moriyama, Eriko Ikeno, Nae Matsuda, Yumiko Saga, Bruce Blumberg, and Jun Kanno
Development of Humanized Steroid and Xenobiotic Receptor Mouse by homologous knock-in of the human Steroid and Xenobiotic Receptor Ligand Binding Domain sequence. *J Toxicol Sci.* 2012, Vol37, No2: 373-380.

Fujiki R, Hashiba W, Sekine H, Yokoyama A, Chikanishi T, Ito S, Imai Y, Kim J, He HH, Igarashi K, Kanno J, Ohtake F, Kitagawa H, Roeder RG, Brown M, Kato S. GlcNAcylation of histone H2B facilitates its monoubiquitination. *Nature.* 2011 Nov 27;480(7378):557-560

Matsukura H, Aisaki K, Igarashi K, Matsushima Y, Kanno J, Muramatsu M, Sudo K, Sato N. Genistein promotes DNA demethylation of the steroidogenic factor 1 (SF-1) promoter in endometrial stromal cells. *Biochem Biophys Res Commun.* 412(2011) 366-372.

2. 学会発表

菅野 純、五十嵐勝秀、相崎健一、北嶋 聡、種村健太郎、ヒト型リガンド結合ドメインノックイン PXR マウスの遺伝子発現応答特性、第 29 回内分泌代謝学サマーセミナー(2011.7.8) (仙台)

菅野 純、Percellome 解析:時間軸と用量軸の融合と絶対値による解析精度の向上、第 38 回日本トキシコロジー学会学術年会(2011.7.12) (横浜)

五十嵐勝秀、北嶋 聡、相崎健一、菅野

純、ヒト型 PXR 生理的発現マウス系の全身臓器トランスクリプトーム解析、第 38 回日本トキシコロジー学会学術年会(2011.7.13) (横浜)

菅野 純、相崎健一、北嶋 聡、パーセローム(Percellome)法を用いた定量的トランスクリプトミクスによる遺伝子発現ネットワーク描出による毒性解析の試み、第 34 回日本高血圧学会総会 SHR 学会合同シンポジウム、2011.10.22、宇都宮

Jun Kanno, Humanized Steroid and Xenobiotic Receptor Mouse by homologous knock-in of the human Steroid and Xenobiotic Receptor Ligand Binding Domain." the 51st Annual Meeting of the Society of Toxicology (2012.3.15) San Francisco, USA

H. 知的財産所有権の出願・登録状況(予定も含む)

1. 特許取得

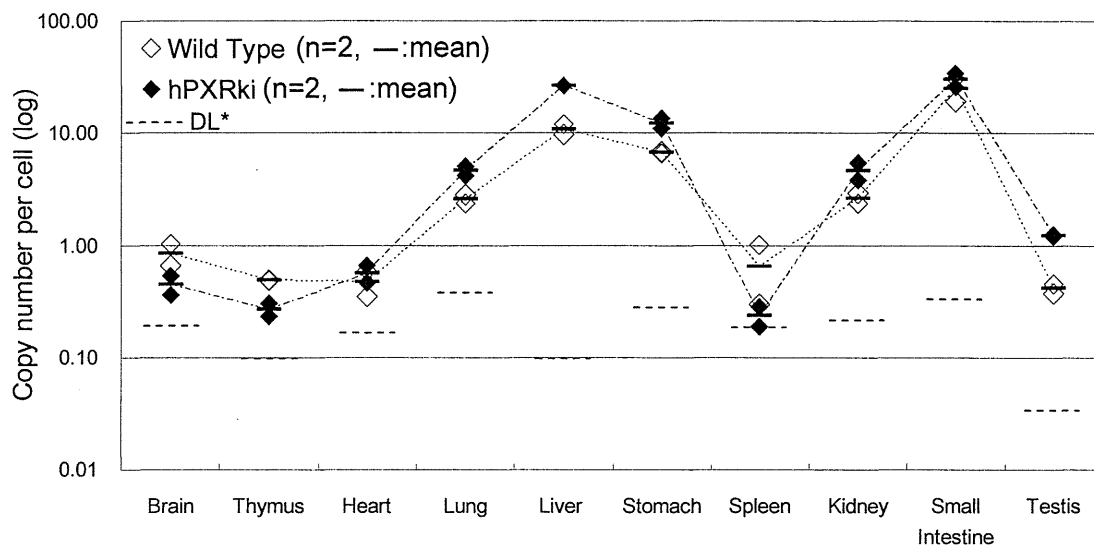
なし

2. 実用新案登録

なし

3. その他

なし



* DL:PCR detection limits of each organ

Figure 1 hSXRki マウスにおける SXR の臓器毎の発現

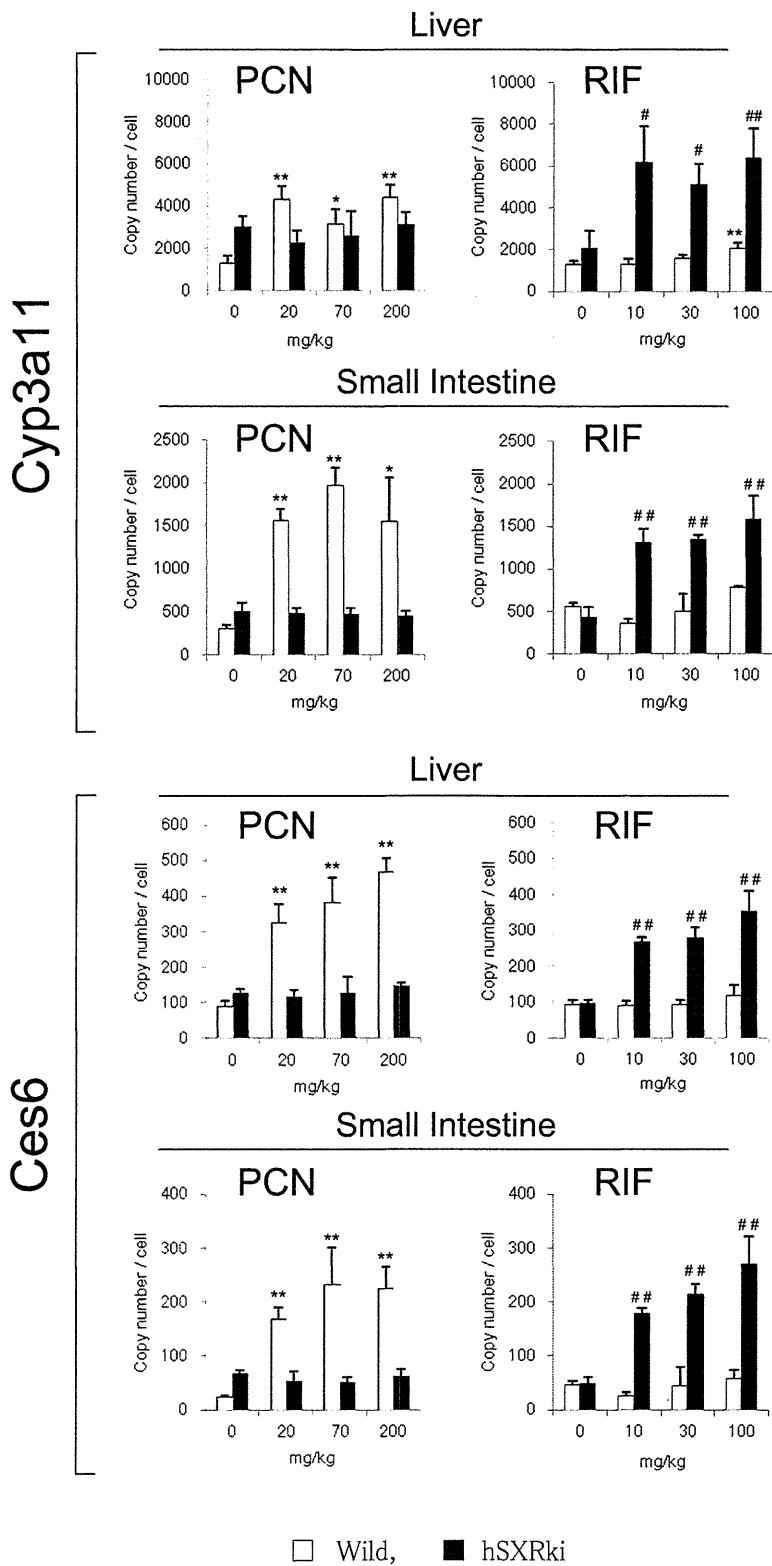


Figure 2 肝、小腸における SXR 応答のヒト型化確認

ISH of Cyp3a11

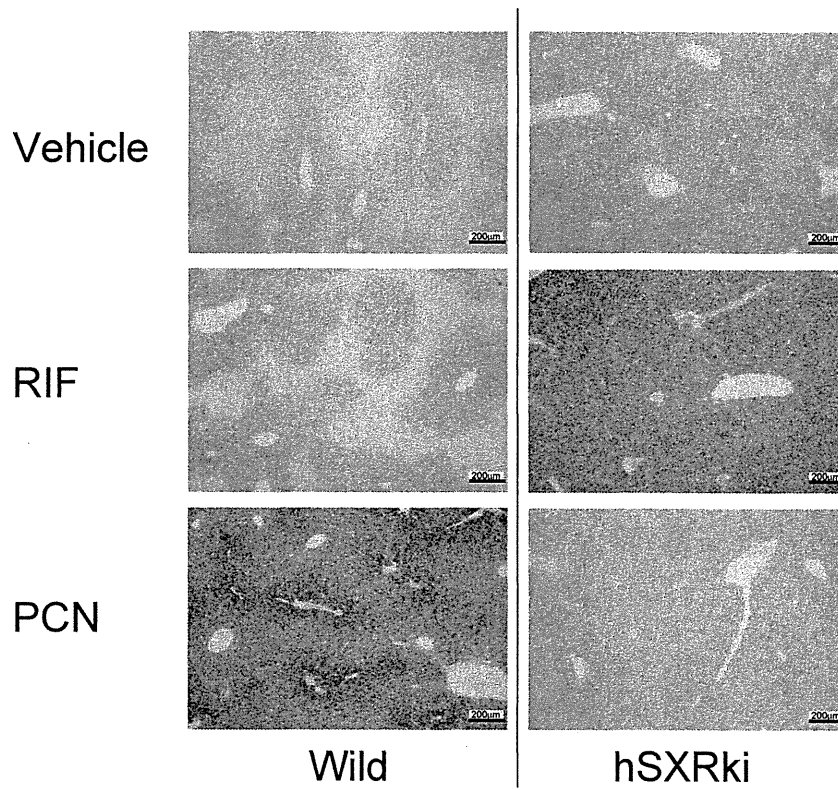
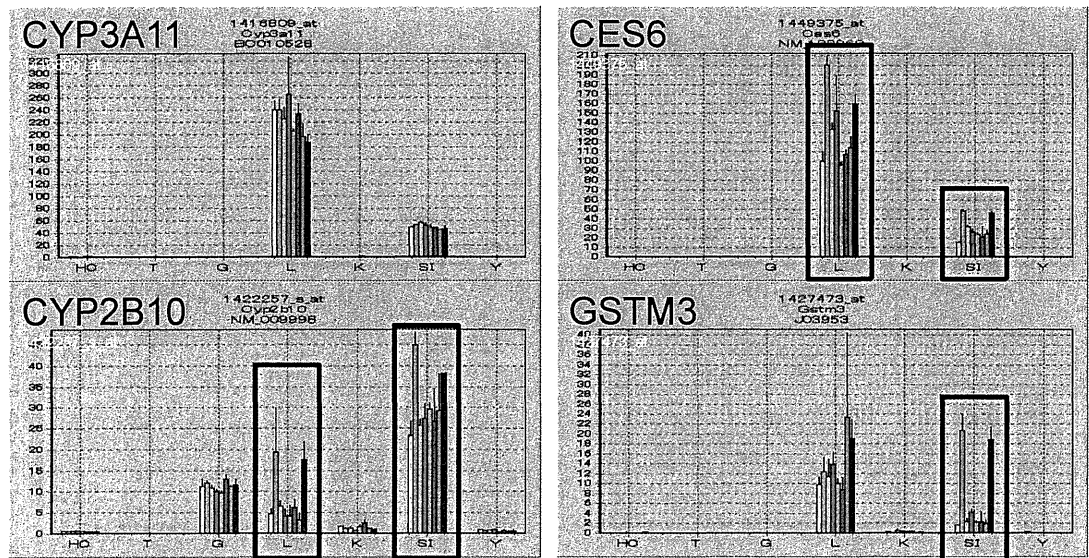


Figure 3 肝における SXR 応答のヒト型化の *in situ* hybridization による確認 (Cyp3a11)

A



B

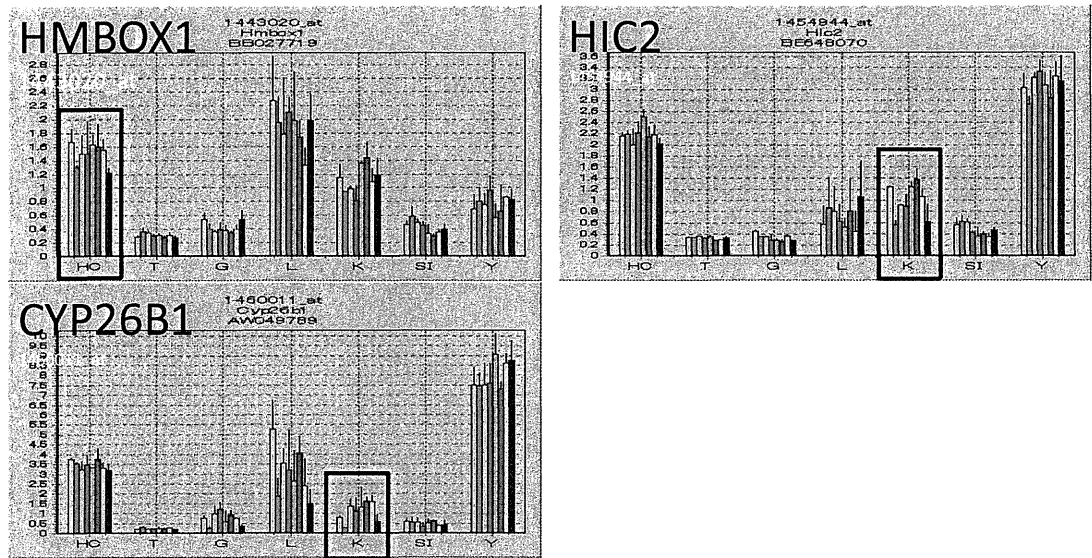


Figure 4 7臓器における遺伝子発現応答検討

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

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.	30(10)	1701-1709	2011
Uehara T, Minowa Y, Morikawa Y, Kondo C, Maruyama T, Kato I, Nakatsu N, Igarashi Y, Ono A, Hayashi H, Mitsumori K, Yamada H, Ohno Y, Urushidani T.	Prediction model of potential hepatocarcinogenicity of rat hepatocarcinogens using a large-scale toxicogenomics database.	Toxicol Appl Pharmacol.	255(3)	297-306	2011

Low Y, Uehara T, Minowa Y, Yamada H, Ohno Y, Urushidani T, Sedykh A, Muratov E, Kuz'min V, Fourches D, Zhu H, Rusyn I, Tropsha A.	Predicting drug-induced hepatotoxicity using QSAR and toxicogenomics approaches.	Chem Res Toxicol.	24(8)	1251-1262	2011
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)	734-42	2011
Yoshida T, Sekine T, Aisaki KI, Mikami T, Kanno J, Okayasu I.,	CITED2 is activated in ulcerative colitis and induces p53-dependent apoptosis in response to butyric acid.	J. Gastroenterol	46(3)	339-49.	2011
Katsuhide Igarashi, Satoshi Kitajima, Ken-ichi Aisaki, Kentaro Tanemura, Yuhji Taquahashi, Noriko Moriyama, Eriko Ikeno, Nae Matsuda, Yumiko Sagar, Bruce Blumberg, and Jun Kanno	Development of Humanized Steroid and Xenobiotic Receptor Mouse by homologous knock-in of the human Steroid and Xenobiotic Receptor Ligand Binding Domain sequence.	J Toxicol Sci.	Vol37, No2	373-380	2012
Fujiki R, Hashiba W, Sekine H, Yokoyama A, Chikanishi T, Ito S, Imai Y, Kim J, He HH, Igarashi K, Kanno J, Ohtake F, Kitagawa H, Roeder RG, Brown M, Kato S.	GlcNAcylation of histone H2B facilitates its monoubiquitination.	Nature	VOL480	557-560	2011

Matsukura H, Aisaki K, Igarashi K, Matsushima Y, Kanno J, Muramatsu M, Sudo K, Sato N.	Genistein promotes DNA demethylation of the steroidogenic factor 1 (SF-1) promoter in endometrial stromal cells.	Biochem Biophys Res Commun.	412	366-372	2011
---	--	--------------------------------------	-----	---------	------

In Silico Toxicology Prediction Using Toxicogenomics Data

Yasushi Okuno¹, Yohsuke Minowa², Hiroshi Yamada², Yasuo Ohno² and Tetsuro Urushidani^{2,3}

¹Department of Systems Bioscience for Drug Discovery, Graduate School of Pharmaceutical Sciences, Kyoto University, Kyoto, Japan, ²Toxicogenomics-Informatics Project, National Institute of Biomedical Innovation, Ibaraki, Osaka, Japan and ³Department 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) knowledge-base (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 [$\text{dim}(\mathbf{x})$: 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

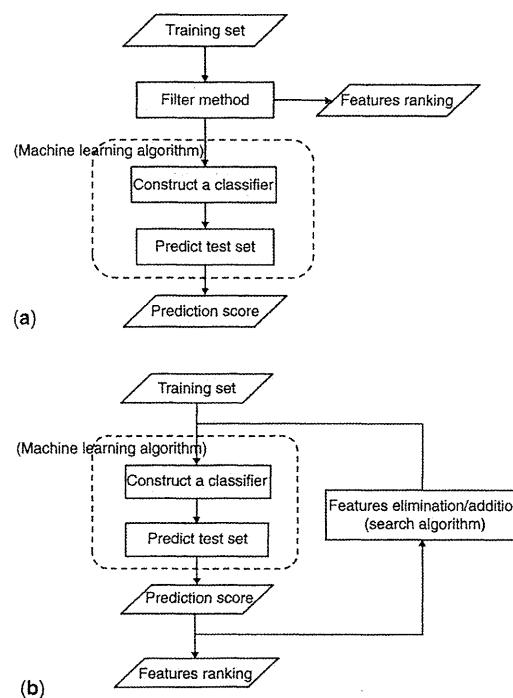


Figure 1. Calculation procedures of filter method (a) and wrapper method (b).

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),

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

where CFS is the score of a feature subset containing k features, \bar{r}_{cf} is the average of the correlation coefficients between features and classes, and \bar{r}_{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 $f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b$ or $f(\mathbf{x}) = \sum_{k=1}^n w_k x_k + b$. 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