

### 資料3 臨床サンプルに関する遺伝子発現解析経過

#### 1 新フィルターアレイによる臨床サンプルでのシグナルの状態

臨床サンプルから得られた RNA での遺伝子発現を新アレイフィルターで検出した場合のシグナルは図1に示すように、良好であった。

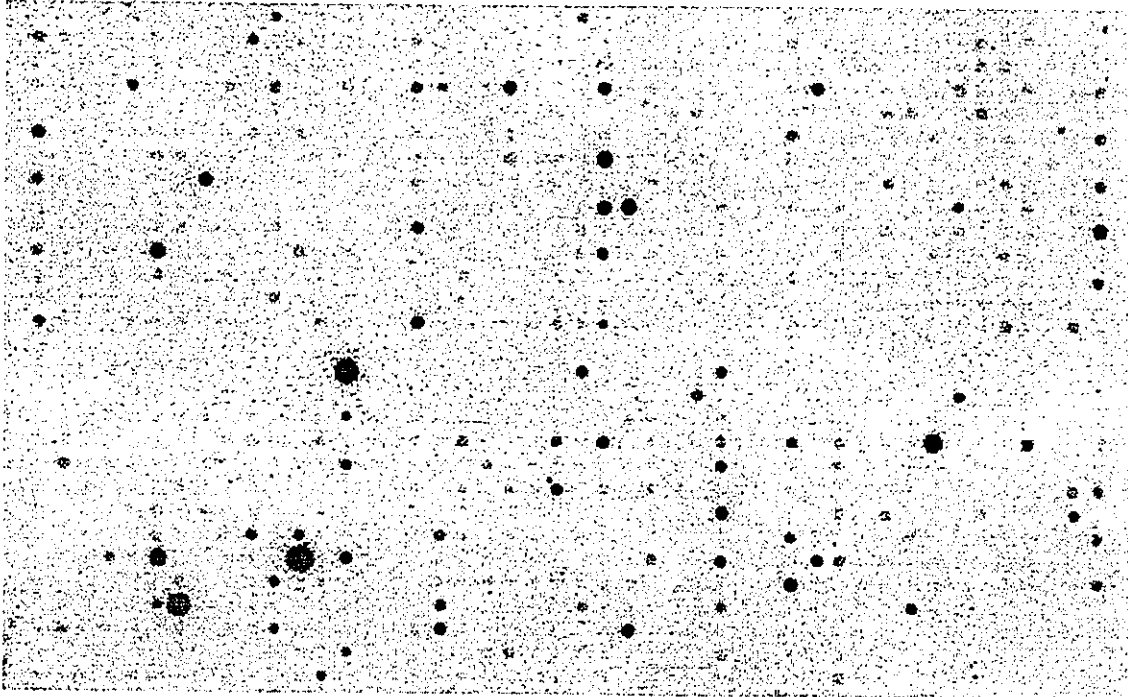


図1

#### 2 末梢血単核球および腫瘍組織を含めたクラスタリング

本プロジェクトでは末梢血単核球と腫瘍組織の異なる組織のアレイが施行されている。よってそれらがクラスタリングにより分けられることで、アレイの検出力の真偽はある程度評価可能と考えられる。施行した結果を図 V-2A に示す。全カテゴリーの遺伝子セットでのクラスタリングでは2群に識別されフィルターのみならず、サンプル管理から RNA 抽出、アレイ施行を含めたアッセイ系が適正に行われていることが確認された (図 V-2A)。

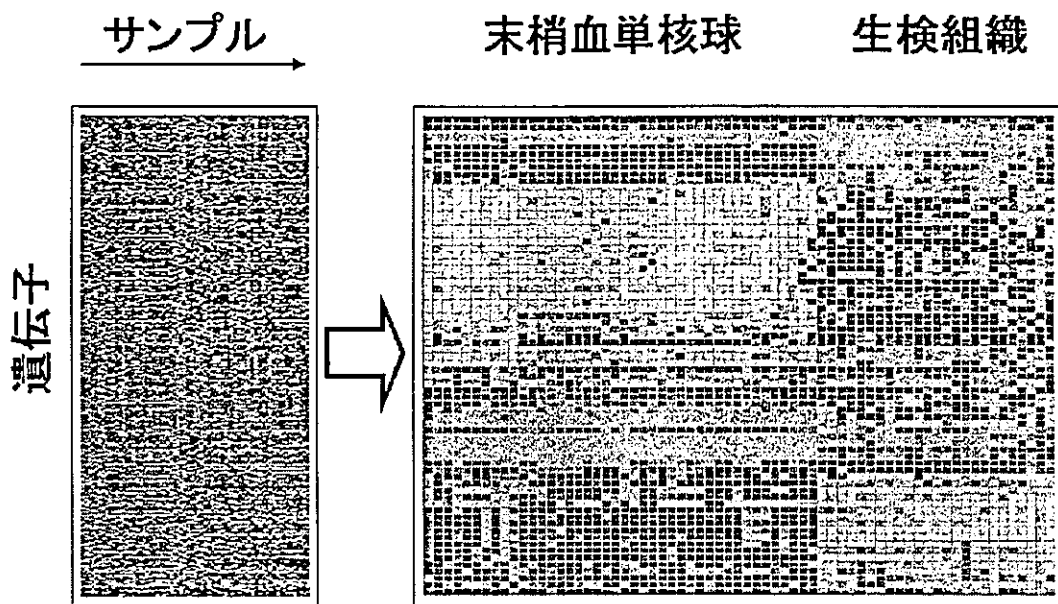


図 2

### 3 ホルモン関連遺伝子セットでの組織別クラスタリング

末梢血単核球および生検組織それぞれのアレイデータを全遺伝子セットでクラスタリングした場合とホルモン関連遺伝子セットでクラスタリングした場合を比較することで選択した遺伝子群の有用性を類推することとした。末梢血リンパ球での場合、全遺伝子セットとホルモン関連遺伝子群セットでのクラスタリング結果は、全く異なる結果を示した。これは図 V-2A で示された 2 群での群分けが単核球としての遺伝子発現に規定されていたことからクラスターを形成する第 1 主成分にはホルモン関連遺伝子が存在していないと類推される。一方、生検組織におけるクラスタリング結果は遺伝子セットが全カテゴリーであってもホルモン関連遺伝子だけであっても相似した群分けがされたことから、乳癌組織の a 亜群分類は選択したホルモン関連遺伝子群によって規定されている可能性が推察される (図 3)。

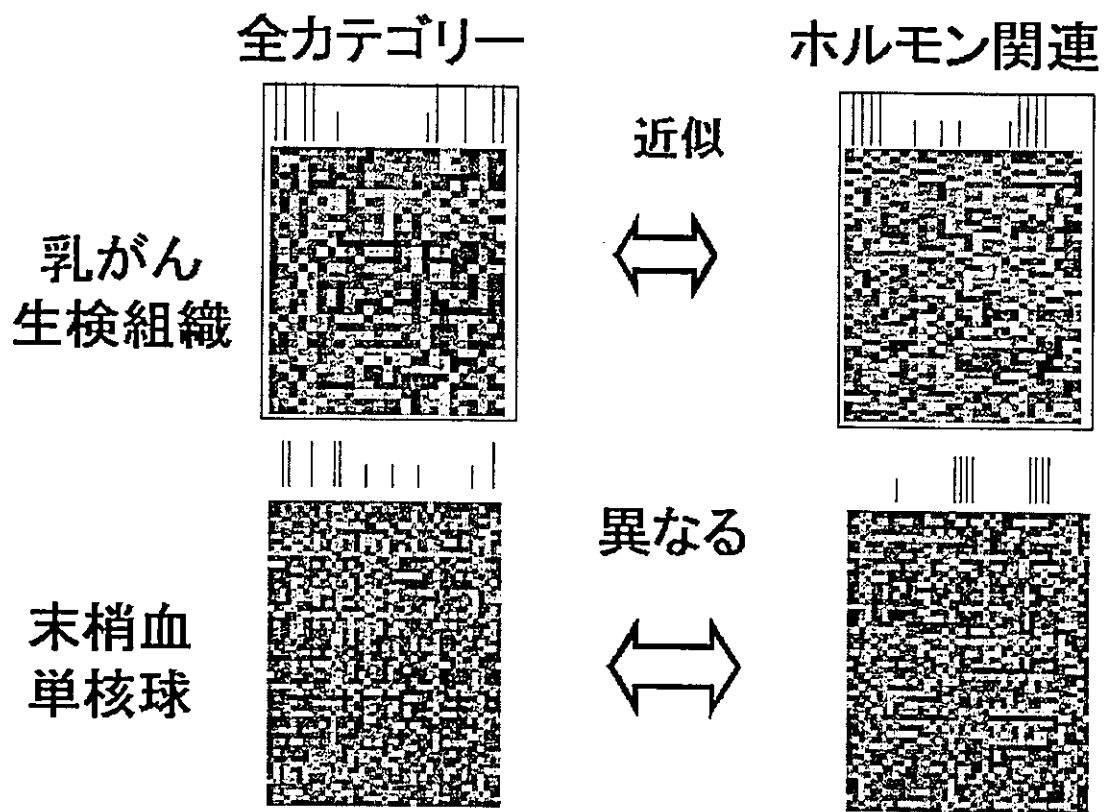


図 3

厚生労働省萌芽的先端医療技術推進研究事業  
分担研究報告書  
新カスタムアレイの評価に関する研究  
分担研究者 大橋靖雄（東京大学医学系研究科教授）

研究要旨：cDNA マクロアレイ実験において、遺伝子発現量を測定する際に生じるばらつき要因の特定と、その要因の分散の推定を行なった。その結果に基づき、ばらつきの大きい要因を考慮した実験デザインを想定し、そのデザインの下で十分な検出力を確保するために必要な繰り返し数の見積もりを行なった。

#### A. 研究目的

cDNA アレイ技術は、多数の遺伝子の発現量を同時に測定することができる技術である。この技術によって得られる情報は、新薬の開発や、各種疾患の解明につながると期待されている。しかし、cDNA アレイによる測定では、測定段階の様々な過程で測定にばらつきが生じる。したがって、このような測定ノイズを含んだデータの中から、いかにして真の情報を得るかが課題となる。そのためにはあらかじめ実験計画を立て、十分な測定の繰り返しを行なうことで、実験で生じるばらつきを制御することが重要である。しかし現在、cDNA アレイ実験の中でも cDNA マクロアレイ実験の実験計画についての検討は十分に行なわれていない。

そこで本研究では、cDNA マクロアレイ実験において、遺伝子発現量を測定する際に生じるばらつき要因の特定と、その要因の分散の推定を行なった。その結果に基づき、ばらつきの大きい要因を考慮した実験デザインを想定し、そのデザインの下で十分な検出力を確保するために必要な繰り返し数の見積もりを行なった。

#### B. 研究方法

今回用いるデータは国立がんセンターにおいて行なわれた cDNA マクロアレイ実験の再現性、信頼性を評価するためのデータである。同一の肺癌細胞株 PC14 を 2 つに分割し、それぞれの細胞より抽出した tRNA に対して T7-RNA 増幅法（T7-RNA ポリメラーゼを利用した RNA を増幅する方法で、臨床研究に用いられている）を行う場合と、増幅法を行わない場合のそれぞれについて cDNA マクロアレイにより遺伝子発現量を測定した。アレイはすべて 1 人の実験者によって測定されている。遺伝子数は 775 個である。具体的な実験手順としては、T7-RNA 増幅法を施行したアレイを 1 日につき 3 枚、3 日間連続して、合計 9 枚作成した。その後、数日間空けてから、T7-RNA 増幅法を施行していないアレイを 1 日につき 3 枚、3 日間連続で合計 9 枚作成した。各遺伝子は 1 枚のアレイ上で、1 回ずつ測定されているため、各遺伝子の測定回数は合計 18 回ずつである。

実データに対して、T7-RNA 増幅法、日、アレイ、遺伝子をばらつき要因としたモデルにより、各ばらつき要因の分散推定値を求めた。研究者が実験作業に慣れることで生じるばらつき

の影響が大きいことが予想されたため、T7-RNA 増幅法を実施した場合と実施していない場合の、それぞれについてサブグループ解析を行った。6日間の ICC の変化を調べるため、1日の3枚のアレイ毎にも解析を行なった。信頼性の検討は、級内相関係数(以下、ICC と表す)を用いて行なった。

この実データに対する解析結果に基づき、日間ばらつきを考慮して、cDNA マクロアレイ実験のデザ

インを3つ想定した。設定1は、同日内で、同じ処理を施行した検体を2枚のアレイを用いて測定する場合。設定2は、同日内で、処理1を施行した検体と処理2を施行した検体をそれぞれ1枚ずつのアレイを用いて測定する場合。設定3は、同日内で、処理1を施行した検体と処理2を施行した検体をそれぞれ2枚ずつのアレイを用いて合計4枚のアレイにて測定する場合である。アレイの枚数は3つのデザイン間で同じである。それぞれのデザインに関して、モデルと分散分析表を作成した。興味があるのは、異なる処理の影響により、有意な遺伝子発現量の変化をしている遺伝子が存在するかどうかである。これに関してF検定を行なう場合に、必要なアレイの繰り返し数について、検出力の観点から検討した。

### C. 研究結果

実データに対する解析によって、各ばらつき要因の分散推定値は、遺伝子が 0.538、日が 0.763、アレイが 0.254、誤差が 0.393 となり、ICC は 0.273 と低かった。測定日間のばらつきが大きいことが確認された。また、T7-RNA 増幅法を実施した3枚のアレイと T7-RNA 増幅法を実施していない3枚のアレイに対するサブグループ解析の結果、前者では、各ばらつき要因の

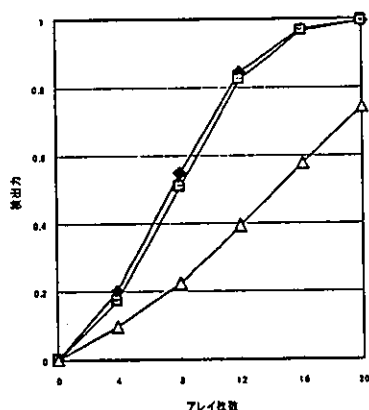
分散推定値は、遺伝子が 0.513、日が 1.281、アレイが 0.129、誤差が 0.154 となり、ICC が 0.247 と低かったのに対し、後者では、遺伝子が 0.951、日が 0.246、アレイが 0.380、誤差が 0.244 となり、日間ばらつきが小さくなったことで、ICC は 0.522 と上昇した。

検出力とアレイ枚数の検討によって、3つのデザインの比較を行なった結果を下図に示す。ただし、対立仮説を、発現量が2倍になる遺伝子が全体の中で1%あるとした場合である。[設定1]、[設定2]、[設定3]の順に検出力は上昇した。つまり、同日内では、異なる処理を施行した2枚のアレイを比較した方がよく、且つ、それぞれの処理を施行したアレイの繰り返しを行なった方がよいことが示された。

### D. 考察

本研究で解析した cDNA マクロアレイ実験では、ばらつきは大きく、信頼性は低かった。日間ばらつきが大きかった原因としては、日間の実験条件の違いのほか、実験を行なった研究者の技術的な向上による影響も大きかったと考えられる。よって、実験を行なう技術者は、実際に実験を始める前に、ある程度高い ICC の値が出るまで練習を行なう必要があると考えられる。その他、十分な信頼性を得られなかった原因として、同じアレイ上での同一遺伝子の繰り返しがなされていなかったことなどが挙げられる。アレイ上における同一遺伝子の繰り返しにより、アレイと遺伝子の交互作用によるばらつきの大さの定量化にもつながる。このように、cDNA マクロアレイ実験では、様々なばらつき要因が存在するため、状況に応じて、ばらつきの大さの要因を考慮した実験デザインの下、実験を行なう必要があることが示唆された。また、実験を行なう前に、予備的検討をして、ある程

度の必要な繰り返し数の目安をつけておくことの重要性が示された。それを考慮せずに実験を行なうと、十分な検出力を得られなかったり、必要以上に繰り返すことで無駄にコストがかかってしまう可能性がある。今回、日間ばらつきに着目し、アレイの繰り返しの検討を行なったが、アレイ以外にも、繰り返しを行なう実験段階は存在するので、今後検討の課題としたい。また、今回信頼性の指標に ICC を用いたが、その他の指標についても今後検討する必要がある。



- ◆ : 同日内で、処理 1 を施行した検体と処理 2 を施行した検体をそれぞれ 2 枚ずつのアレイを用いて合計 4 枚のアレイにて測定する場合 [設定 3]
- : 同日内で、処理 1 を施行した検体と処理 2 を施行した検体をそれぞれ 1 枚ずつのアレイを用いて測定する場合 [設定 2]
- ▲ : 同日内で、同じ処理を施行した検体を 2 枚のアレイを用いて測定する場合 [設定 1]

発現量が 2 倍になる遺伝子が、全体の遺伝子の中に 1%ある場合での、3 つの実験デザインの比較

## E. 結論

新カスタムアレイにおいて、遺伝子発現量を測定する際に生じるばらつき要因の特定と、その要因の分散の推定を行なった。その結果に基づき、ばらつきの大きい要因を考慮した実験デザインを想定し、そのデザインの下で十分な検出

力を確保するために必要な繰り返し数の見積もりを行なった。同結果を基に、臨床サンプルの解析を行うことで、信頼性の向上した遺伝子発現解析の実施が可能である。

## F. 健康危害情報

特になし

## G. 研究発表

- Tominaga, R., Koyama, H.,
1. Toge, T., Miura, S., Sugimachi, K., Ymaguchi, S., Hirata, K., Ohashi, Y., Abe, O. Randomized controlled trial comparing oral doxifluridine plus oral cyclophosphamide with doxifluridine alone in women with node-positive breast cancer after primary surgery. *J. Clin. Oncol.* 21: 991-998, 2003.
  2. Yanaka, N., Kinoshita, T., Asada, T., Ohashi, Y.: Long-linear models for assessing gene-ageinteraction and their application to case-control studies of the apolipoprotein E (apoE) gene in Alzheimer's disease. *J. Human Genet.* 48:520-524, 2003.
  3. Kawado, M., Hinotsu, S., Matsuyama, Y., Yamaguchi, T., Hashimoto, S., Ohashi, Y. A comparison of error detection rates between the reading aloud method and the double date entry method. *Controlled Clinical Trials* 24: 560- 569, 2003.

## H. 知的財産の出願・登録状況

特になし

資料4 フィルターアレイにおいて繰り返し測定された遺伝子の詳細

(address:アレイ内のスポット位置, Gene:遺伝子名, GenBank:GenBank Database における遺伝子 ID, obs1-obs3:3 枚のアレイにおける遺伝子強度(生データ), Type:遺伝子の機能別分類番号)

1. 新フィルターアレイ

NO	address	Gene	GenBank	obs1	obs2	obs3
20	A02f	MRP/organic anion transporter (MOAT-B)	AF071202	75.87	82.46	69.65
953	E13a	MRP/organic anion transporter (MOAT-B)	AF071202	65.45	71.84	66.87
1052	F05b	MRP/organic anion transporter (MOAT-B)	AF071202	61.52	64.17	55.66
1115	F10i	DNA (cytosine-5-)-methyltransferase 2 (DNMT2)	AJ223333	60.18	68.01	56.02
1150	F13b	DNA (cytosine-5-)-methyltransferase 2 (DNMT2)	AJ223333	56.32	61.04	55.73
1151	F13c	DNA (cytosine-5-)-methyltransferase 2 (DNMT2)	AJ223333	57.53	62.85	56.84
26	A02i	protein kinase C delta (PKC-delta)	D10495	77.1	93.89	81.72
564	C13d	protein kinase C delta (PKC-delta)	D10495	67.8	85.08	66.81
1011	F03c	protein kinase C delta (PKC-delta)	D10495	74.1	83.1	69.33
459	C05k	interleukin 2 receptor gamma subunit (IL2R-gamma; IL2RG); cytokine receptor common gamma chain precursor	D11086	62.38	65.2	58.95
757	D13a	interleukin 2 receptor gamma subunit (IL2R-gamma; IL2RG); cytokine receptor common gamma chain precursor	D11086	73.73	74.92	68.93
780	D14j	interleukin 2 receptor gamma subunit (IL2R-gamma; IL2RG); cytokine receptor common gamma chain precursor	D11086	157.49	206.45	227.61
220	B02j	phospholipase C (PLCL)	D42108	127.29	195.93	221.91
511	C09g	phospholipase C (PLCL)	D42108	67.69	67.3	64.36
966	E13n	phospholipase C (PLCL)	D42108	65.12	71.45	60.83
197	B01a	microsomal UDP-glucuronosyltransferase 1-6 precursor (UDPGT; UGT1.6; UGT1F; GNT1)	J04093	64.15	68.74	57.92
392	B14n	microsomal UDP-glucuronosyltransferase 1-6 precursor (UDPGT; UGT1.6; UGT1F; GNT1)	J04093	61.15	66.21	56.32
568	C13h	microsomal UDP-glucuronosyltransferase 1-6 precursor (UDPGT; UGT1.6; UGT1F; GNT1)	J04093	94.39	165.1	115.1
141	A11a	interleukin 7 (IL7)	J04156	72.89	73.47	64.43
770	D13n	interleukin 7 (IL7)	J04156	85.68	99.31	88.92
1155	F13g	interleukin 7 (IL7)	J04156	85.4	105.93	94.92
689	D08c	matrix metalloproteinase 9 (MMP9); gelatinase B; 92-kDa type IV collagenase	J05070; D10051	72.39	77.94	62.45

	precursor (CLG4B)				
1160 F13l	matrix metalloproteinase 9 (MMP9); gelatinase B; 92-kDa type IV collagenase	J05070; D10051	61.48	70.99	58.55
	precursor (CLG4B)				
1167 F14e	matrix metalloproteinase 9 (MMP9); gelatinase B; 92-kDa type IV collagenase	J05070; D10051	136.09	165.36	154.97
	precursor (CLG4B)				
567 C13g	c-fos proto-oncogene; G0S7 protein	K00650	71.09	100.28	77.84
668 D06j	c-fos proto-oncogene; G0S7 protein	K00650	103.46	115.8	92.81
1162 F13n	c-fos proto-oncogene; G0S7 protein	K00650	64.34	75.64	64.46
557 C12k	peroxisome proliferator activated receptor alpha (PPAR-alpha; PPARA)	L02932	60.48	63.12	57.66
573 C13m	peroxisome proliferator activated receptor alpha (PPAR-alpha; PPARA)	L02932	75.68	104.74	89.99
957 E13e	peroxisome proliferator activated receptor alpha (PPAR-alpha; PPARA)	L02932	142.16	161.41	93.97
23 A02i	excision repair cross-complementing rodent repair deficiency complementation group 6 (ERCC6); Cockayne syndrome protein 2 type B (CSB)	L04791	85.37	84.88	80.91
207 B01k	excision repair cross-complementing rodent repair deficiency complementation group 6 (ERCC6); Cockayne syndrome protein 2 type B (CSB)	L04791	119.75	174.54	218.83
1055 F06e	excision repair cross-complementing rodent repair deficiency complementation group 6 (ERCC6); Cockayne syndrome protein 2 type B (CSB)	L04791	63.2	65.67	61.78
221 B02k	heat shock 70-kDa protein 4 (HSPA4); HSP70RY; heat shock 70-related protein	L12723	132	190.84	234.48
	APG-2				
994 F01n	heat shock 70-kDa protein 4 (HSPA4); HSP70RY; heat shock 70-related protein	L12723	143.55	157.78	132.27
	APG-2				
1157 F13i	heat shock 70-kDa protein 4 (HSPA4); HSP70RY; heat shock 70-related protein	L12723	112.29	175.53	142.55
	APG-2				
27 A02m	focal adhesion kinase (FADK); proline-rich tyrosine kinase 2 (PYK2)	L13616	98.92	155.97	183.48
361 B12k	focal adhesion kinase (FADK); proline-rich tyrosine kinase 2 (PYK2)	L13616	141.42	153.08	141.78
1165 F14c	focal adhesion kinase (FADK); proline-rich tyrosine kinase 2 (PYK2)	L13616	71.31	118.2	117.8
581 C14g	retinoblastoma-related protein p107	L14812	123.42	194.77	113.03
632 D04b	retinoblastoma-related protein p107	L14812	72.02	76.69	60.34
976 E14j	retinoblastoma-related protein p107	L14812	174.23	216.38	265.74
469 C06g	cadherin 8 (CDH8)	L34060	60.58	61.79	58.98
955 E13c	cadherin 8 (CDH8)	L34060	65.54	67.99	57.89
1152 F13d	cadherin 8 (CDH8)	L34060	59.3	64.75	55.2
759 D13c	sulfonylurea receptor (SUR); ATP-binding cassette subfamily C (CFTR/MRP) member 8 (ABCC8)	L78207	127.14	110.52	166.1
784 D14n	sulfonylurea receptor (SUR); ATP-binding cassette subfamily C (CFTR/MRP)	L78207	132.51	188.23	197.85



	member 8 (ABCC8)				
888 E08f	sulfonylurea receptor (SUR); ATP-binding cassette subfamily C (CFTR/MRP)	L78207	64.33	68.62	58.56
	member 8 (ABCC8)				
574 C13n	urokinase-type plasminogen activator precursor (U-plasminogen activator; UPA)	M15476; D00244	180.82	270.99	293.62
816 E03d	urokinase-type plasminogen activator precursor (U-plasminogen activator; UPA)	M15476; D00244	314.24	336.07	291.98
1154 F13f	urokinase-type plasminogen activator precursor (U-plasminogen activator; UPA)	M15476; D00244	251.31	375.31	368.19
444 C04j	cytochrome P450 IID6 (CYP2D6)	M20403; M19697	80.82	78.1	81.6
563 C13c	cytochrome P450 IID6 (CYP2D6)	M20403; M19697	76.96	79.29	79.3
758 D13b	cytochrome P450 IID6 (CYP2D6)	M20403; M19697	97.49	96.2	111.2
510 C09f	erbB3 proto-oncogene; HER3	M29366; M34309	75.4	88.42	84.65
777 D14g	erbB3 proto-oncogene; HER3	M29366; M34309	90.43	95.48	83.76
983 F01c	erbB3 proto-oncogene; HER3	M29366; M34309	70.13	76.56	67.94
13 A01m	ras-related C3 botulinum toxin substrate 1 (RAC1); ras-like protein TC25	M29870; M31467	111.13	222.61	293.21
18 A02d	ras-related C3 botulinum toxin substrate 1 (RAC1); ras-like protein TC25	M29870; M31467	106.65	81.38	112.34
675 D07c	ras-related C3 botulinum toxin substrate 1 (RAC1); ras-like protein TC25	M29870; M31467	121.2	132.83	138.24
22 A02h	xeroderma pigmentosum group B complementing protein (XPB); excision repair cross-complementing rodent repair deficiency complementation group 3 (ERCC3); basal transcription factor 2 89-kDa subunit (BTF2p89; TFIIH 89-kDa subunit)	M31899	135.83	150.61	185.9
281 B07a	xeroderma pigmentosum group B complementing protein (XPB); excision repair cross-complementing rodent repair deficiency complementation group 3 (ERCC3); basal transcription factor 2 89-kDa subunit (BTF2p89; TFIIH 89-kDa subunit)	M31899	104.33	96.49	92.15
583 C14i	xeroderma pigmentosum group B complementing protein (XPB); excision repair cross-complementing rodent repair deficiency complementation group 3 (ERCC3); basal transcription factor 2 89-kDa subunit (BTF2p89; TFIIH 89-kDa subunit)	M31899	140.78	321.45	178.46
199 B01c	insulin-like growth factor-binding protein 2 (IGF-binding protein 2; IGFBP2; IBP2)	M35410	141.3	176.33	162.17
772 D14b	insulin-like growth factor-binding protein 2 (IGF-binding protein 2; IGFBP2; IBP2)	M35410	154.32	151.87	168.2
788 E01d	insulin-like growth factor-binding protein 2 (IGF-binding protein 2; IGFBP2; IBP2)	M35410	71.89	79.45	65.53
9 A01i	X-ray repair-complementing defective repair in Chinese hamster cells 1 (XRCC1)	M36089	139.63	134.87	177.46

235 B03k	X-ray repair-complementing defective repair in Chinese hamster cells 1 (XRCC1)	M36089	83.68	86.18	82.14
561 C13a	X-ray repair-complementing defective repair in Chinese hamster cells 1 (XRCC1)	M36089	61.57	78.62	70.25
162 A12h	interleukin 11 (IL11); adipogenesis inhibitory factor (AGIF)	M57765	81.65	85.65	74.8
203 B01g	interleukin 11 (IL11); adipogenesis inhibitory factor (AGIF)	M57765	68	71.24	61.42
217 B02g	interleukin 11 (IL11); adipogenesis inhibitory factor (AGIF)	M57765	70.21	70.02	61.52
3 A01c	interleukin 3 receptor beta subunit precursor (IL3R-beta; IL3RB); interleukin 5 receptor beta subunit precursor (IL5R-beta; IL5R); cytokine receptor common beta chain precursor; CSF2RB	M59941	114.43	104.14	145.52
896 E08n	interleukin 3 receptor beta subunit precursor (IL3R-beta; IL3RB); interleukin 5 receptor beta subunit precursor (IL5R-beta; IL5R); cytokine receptor common beta chain precursor; CSF2RB	M59941	67.61	71.7	62
1153 F13c	interleukin 3 receptor beta subunit precursor (IL3R-beta; IL3RB); interleukin 5 receptor beta subunit precursor (IL5R-beta; IL5R); cytokine receptor common beta chain precursor; CSF2RB	M59941	76.49	81.61	80.39
118 A09f	colorectal mutant cancer protein (MCC)	M62397	72.84	72.57	59.27
211 B02a	colorectal mutant cancer protein (MCC)	M62397	63.55	65.92	59.59
216 B02f	colorectal mutant cancer protein (MCC)	M62397	82	86.71	86.12
222 B02i	integrin beta 7 precursor (ITGB7)	M62880; S80335	65.23	73.37	65.79
895 E08m	integrin beta 7 precursor (ITGB7)	M62880; S80335	67.57	72.59	61.91
956 E13d	integrin beta 7 precursor (ITGB7)	M62880; S80335	78.34	91.24	64.55
24 A02j	ERBB2 proto-oncogene; NEU proto-oncogene; HER2	M95667; M11730	71.09	73.47	59
174 A13f	ERBB2 proto-oncogene; NEU proto-oncogene; HER2	M95667; M11730	67.07	73.1	62.32
565 C13e	ERBB2 proto-oncogene; NEU proto-oncogene; HER2	M95667; M11730	60.91	66.39	59.13
577 C14c	tumor necrosis factor type 1 receptor-associated protein (TRAP1)	U12595	108.77	258.2	157.02
585 C14k	tumor necrosis factor type 1 receptor-associated protein (TRAP1)	U12595	157.97	266.38	206.73
840 E04n	tumor necrosis factor type 1 receptor-associated protein (TRAP1)	U12595	192.3	263.6	222.62
219 B02i	180-kDa secretory phospholipase A2 receptor precursor (PLA2 receptor; PLA2R)	U17033	76.65	94.62	95.1
782 D14l	180-kDa secretory phospholipase A2 receptor precursor (PLA2 receptor; PLA2R)	U17033	125.89	173.79	124.9
1103 F09k	180-kDa secretory phospholipase A2 receptor precursor (PLA2 receptor; PLA2R)	U17033	61.87	73.02	61.94
500 C08j	ionizing radiation resistance-conferring protein + death-associated protein 3	U18321 + X83544	69.92	74.54	73.14

	(DAP3)					
569	C13i	ionizing radiation resistance-conferring protein + death-associated protein 3	U18321 + X83544	73.67	103.73	87.32
	(DAP3)					
760	D13d	ionizing radiation resistance-conferring protein + death-associated protein 3	U18321 + X83544	110.06	109.66	123.14
	(DAP3)					
15	A02a	ras-related protein RAB5C	U18420	113.58	112	128.61
762	D13f	ras-related protein RAB5C	U18420	169.84	164.24	222.5
939	E12a	ras-related protein RAB5C	U18420	154.69	171.68	163.67
113	A09a	EB1 protein	U24166	81.68	123.55	113.77
572	C13l	EB1 protein	U24166	133.49	274.14	231.58
978	E14l	EB1 protein	U24166	132.61	150.66	145.88
25	A02k	histone deacetylase 2 (HD2)	U31814	80.04	84.08	82.32
266	B05n	histone deacetylase 2 (HD2)	U31814	84.16	85.07	76.22
961	E13i	histone deacetylase 2 (HD2)	U31814	160.96	208.35	240.82
602	D01n	ataxia telangiectasia (ATM)	U33841	89.41	92.17	83.38
766	D13j	ataxia telangiectasia (ATM)	U33841	106.09	119.22	114.73
963	E13k	ataxia telangiectasia (ATM)	U33841	82.43	93.85	90.41
92	A07h	beta 3-endonexin	U37139	87.36	90.17	84.95
768	D13l	beta 3-endonexin	U37139	79.14	95.96	76.25
965	E13m	beta 3-endonexin	U37139	67.59	73.48	64.15
201	B01e	DNA topoisomerase III alpha (TOP3A)	U43431	177.93	209.31	225.42
798	E01n	DNA topoisomerase III alpha (TOP3A)	U43431	72.79	75.89	61.45
968	E14b	DNA topoisomerase III alpha (TOP3A)	U43431	66.81	70.44	58.64
42	A03n	serine/threonine protein kinase PRP4 homolog	U48736	75.59	81.6	67.81
566	C13f	serine/threonine protein kinase PRP4 homolog	U48736	66.77	79.83	73.28
571	C13k	serine/threonine protein kinase PRP4 homolog	U48736	67.29	84.64	74.76
575	C14a	gamma-glutamyl hydrolase precursor (GGH; GH); folic polygamma-glutamyl hydrolase; gamma-glu-X carboxypeptidase; conjugase	U55206	59.37	68.58	62.47
897	E09a	gamma-glutamyl hydrolase precursor (GGH; GH); folic polygamma-glutamyl hydrolase; gamma-glu-X carboxypeptidase; conjugase	U55206	65.53	69.25	57.71
964	E13l	gamma-glutamyl hydrolase precursor (GGH; GH); folic polygamma-glutamyl hydrolase; gamma-glu-X carboxypeptidase; conjugase	U55206	64.81	72.13	59.79
212	B02b	MHC class I molecule; MHC class I subunit-related protein A (MICA); PERB11.1	U65416; U69631; X92841	86.24	82.25	77.85
302	B08h	MHC class I molecule; MHC class I subunit-related protein A (MICA); PERB11.1	U65416; U69631; X92841	124.39	148.22	127.81
1169	F14g	MHC class I molecule; MHC class I subunit-related protein A (MICA); PERB11.1	U65416; U69631; X92841	171.06	231.64	216.38

562	C13b	TSG101 tumor susceptibility protein	U82130	64.86	76.95	70.26
731	D11c	TSG101 tumor susceptibility protein	U82130	96.61	100.56	89.87
980	E14n	TSG101 tumor susceptibility protein	U82130	71.91	78.62	72.49
7	A01g	RHO GDP-dissociation inhibitor 3 (RHO-GDI 3); RHO-GDI gamma; ARHGDIG	U82532	110.6	136.29	146.5
858	E06d	RHO GDP-dissociation inhibitor 3 (RHO-GDI 3); RHO-GDI gamma; ARHGDIG	U82532	76.33	78.61	71.7
959	E13g	RHO GDP-dissociation inhibitor 3 (RHO-GDI 3); RHO-GDI gamma; ARHGDIG	U82532	67.52	70.54	61.83
16	A02b	glutathione-S-transferase homolog	U90313	231.98	182.33	291.84
19	A02e	glutathione-S-transferase homolog	U90313	194.07	150.74	243.03
536	C11d	glutathione-S-transferase homolog	U90313	195.07	551.81	268.61
5	A01e	kappaB epsilon (IKBE)	U91616	77.69	84.2	70.84
209	B01m	kappaB epsilon (IKBE)	U91616	70.27	73.51	62.86
320	B09l	kappaB epsilon (IKBE)	U91616	65.39	70.65	66.3
215	B02e	dihydrofolate reductase (DHFR)	V00507	86.61	100.09	93.06
630	D03n	dihydrofolate reductase (DHFR)	V00507	75.38	84.17	67.08
972	E14f	dihydrofolate reductase (DHFR)	V00507	171.93	207.6	106.1
765	D13i	transforming growth factor beta (TGF-beta; TGFB)	X02812; J05114	139.83	182.93	194.7
1112	F10f	transforming growth factor beta (TGF-beta; TGFB)	X02812; J05114	60.39	65.64	59.67
1163	F14a	transforming growth factor beta (TGF-beta; TGFB)	X02812; J05114	67.48	64.78	59.12
224	B02n	c-kit proto-oncogene; mast/stem cell growth factor receptor precursor (SCFR); X06182 CD117 antigen		66.01	71.7	61.33
303	B08i	c-kit proto-oncogene; mast/stem cell growth factor receptor precursor (SCFR); X06182 CD117 antigen		588.74	659.9	650
764	D13h	c-kit proto-oncogene; mast/stem cell growth factor receptor precursor (SCFR); X06182 CD117 antigen		89.05	100.33	89.18
205	B01i	retinoic acid receptor alpha 1 (RAR-alpha 1; RARA); PML-RAR protein	X06538; X06614; M73779	102.91	185.33	174.8
457	C05i	retinoic acid receptor alpha 1 (RAR-alpha 1; RARA); PML-RAR protein	X06538; X06614; M73779	64	69.4	59.8
587	C14m	retinoic acid receptor alpha 1 (RAR-alpha 1; RARA); PML-RAR protein	X06538; X06614; M73779	138.79	226.93	245.2
213	B02c	ras homolog gene family member B (RHOB; ARHB); ARH6; H6	X06820	101.24	103.31	102.71
314	B09f	ras homolog gene family member B (RHOB; ARHB); ARH6; H6	X06820	85.33	96.61	76.33
763	D13g	ras homolog gene family member B (RHOB; ARHB); ARH6; H6	X06820	94.41	104.25	99.02
21	A02g	interferon regulatory factor 1 (IRF1)	X14454	86.89	93.53	92.64
769	D13m	interferon regulatory factor 1 (IRF1)	X14454	120.03	197.25	150.23
929	E11e	interferon regulatory factor 1 (IRF1)	X14454	75.73	76.49	72.8
1	A01a	interleukin 9 precursor (IL9); T-cell growth factor p40	X17543; M30134	111.11	104.81	115.86
187	A14e	interleukin 9 precursor (IL9); T-cell growth factor p40	X17543; M30134	64.97	66.63	55.03

1159	F13k	interleukin 9 precursor (IL9); T-cell growth factor p40	X17543; M30134	66.93	81	62.14
28	A02n	heat shock 70-kDa protein 6; heat shock 70-kDa protein B	X51757; M11236	126.21	169.37	191.29
142	A11b	heat shock 70-kDa protein 6; heat shock 70-kDa protein B	X51757; M11236	71.23	68.08	57.36
579	C14e	heat shock 70-kDa protein 6; heat shock 70-kDa protein B	X51757; M11236	64.93	79.25	64.61
738	D11j	rho-related GTP-binding protein (RHOG); ARHG	X61587	78.17	80.87	64.14
954	E13b	rho-related GTP-binding protein (RHOG); ARHG	X61587	67.92	73.07	63.66
970	E14d	rho-related GTP-binding protein (RHOG); ARHG	X61587	178.44	204.74	93.84
115	A09c	DNA cytosine-5-methyltransferase (DNA metase; MCMT)	X63692	70.74	73.32	58.5
223	B02m	DNA cytosine-5-methyltransferase (DNA metase; MCMT)	X63692	67.18	69.46	57.45
1158	F13j	DNA cytosine-5-methyltransferase (DNA metase; MCMT)	X63692	91.46	142.56	101.81
615	D02m	serine/threonine-protein kinase PCTAIRE 1 (PCTK1)	X66363	74.98	82.23	61.03
958	E13f	serine/threonine-protein kinase PCTAIRE 1 (PCTK1)	X66363	86.58	96.41	69.26
960	E13h	serine/threonine-protein kinase PCTAIRE 1 (PCTK1)	X66363	84.55	94.2	94.69
105	A08g	RHO GDP-dissociation inhibitor 1 (RHO-GDI 1); RHO-GDI alpha (GDIA1); ARHGDI1A	X69550	99.79	110.49	113.17
761	D13e	RHO GDP-dissociation inhibitor 1 (RHO-GDI 1); RHO-GDI alpha (GDIA1); ARHGDI1A	X69550	186.22	189.61	266.08
767	D13k	RHO GDP-dissociation inhibitor 1 (RHO-GDI 1); RHO-GDI alpha (GDIA1); ARHGDI1A	X69550	90.64	115.17	111.13
218	B02h	uridine phosphorylase (UDRPase; UP)	X90858	66.14	71.3	62.84
513	C09i	uridine phosphorylase (UDRPase; UP)	X90858	64.04	67.69	68.57
1149	F13a	uridine phosphorylase (UDRPase; UP)	X90858	62.82	63.46	55.74
400	C01h	bleomycin hydrolase (BLM hydrolase)	X92106	64.69	80.84	69.24
774	D14d	bleomycin hydrolase (BLM hydrolase)	X92106	222.29	232.16	248.93
776	D14f	bleomycin hydrolase (BLM hydrolase)	X92106	244.87	276.91	253.14
570	C13j	ras-related protein RAB28	X94703	58.4	69.19	57.86
917	E10g	ras-related protein RAB28	X94703	65.02	69.33	58.32
974	E14h	ras-related protein RAB28	X94703	146.9	163.41	188.01
11	A01k	adenine phosphoribosyltransferase (APRT)	Y00486	74.91	76.97	60.88
214	B02d	adenine phosphoribosyltransferase (APRT)	Y00486	145.4	156.6	190.18
456	C05h	adenine phosphoribosyltransferase (APRT)	Y00486	62.87	67.28	66.49
99	A08a	nm23-H4; nucleoside-diphosphate kinase; nucleoside 5'-diphosphate phosphotransferase (NDK)	Y07604	84.32	107.96	121.88
778	D14h	nm23-H4; nucleoside-diphosphate kinase; nucleoside 5'-diphosphate phosphotransferase (NDK)	Y07604	267.96	333.77	329.79

1161 F13m	nm23-H4; nucleoside-diphosphate kinase; nucleoside 5'-diphosphate phosphotransferase (NDK)	Y07604	103.54	135.49	127.06
292 B07i	rab geranylgeranyltransferase alpha subunit (rab GG transferase; rab GGTase)	Y08200	67.23	65.72	59.12
962 E13j	rab geranylgeranyltransferase alpha subunit (rab GG transferase; rab GGTase)	Y08200	90.68	107.68	101.27
1156 F13h	rab geranylgeranyltransferase alpha subunit (rab GG transferase; rab GGTase)	Y08200	98.43	129.3	111.75

## 2. 旧フィルターアレイにおける対応遺伝子

NO	address	Gene	GenBank	obs1	obs2	obs3	TYPE
1056 F06f	MRP/organic anion transporter (MOAT-B)		AF071202	49.93	59.67	59.45	12
918 E10h	DNA (cytosine-5-)-methyltransferase 2 (DNMT2)		AJ223333	50.45	54.29	57.85	9
244 B04f	protein kinase C delta (PKC-delta)		D10495	52.38	56.22	60.33	2
732 D11d	interleukin 2 receptor gamma subunit (IL2R-gamma; IL2RG); cytokine receptor common gamma chain precursor		D11086	47.13	51.78	49.05	7
256 B05d	phospholipase C (PLCL)		D42108	44.98	50.79	50.36	2
958 E13f	microsomal UDP-glucuronosyltransferase 1-6 precursor (UDPGT; UGT1.6; UGT1F; GNT1)		J04093	47.25	51.76	53.94	10
748 D12f	interleukin 7 (IL7)		J04156	50.76	54.83	51.86	7
536 C11d	matrix metalloproteinase 9 (MMP9); gelatinase B; 92-kDa type IV collagenase precursor (CLG4B)		J05070; D10051	50.37	54.57	59.45	5
123 A09k	c-fos proto-oncogene; GOS7 protein		K00650	51.5	60.26	54.75	2
482 C07f	peroxisome proliferator activated receptor alpha (PPAR-alpha; PPARA)		L02932	50.77	54.3	58.8	4
838 E04i	excision repair cross-complementing rodent repair deficiency complementation group 6 (ERCC6); Cockayne syndrome protein 2 type B (CSB)		L04791	46.31	53.18	48.47	9
1033 F04k	heat shock 70-kDa protein 4 (HSPA4); HSP70RY; heat shock 70-related protein APG-2		L12723	109.3	162.89	117.49	11
627 D03k	focal adhesion kinase (FADK); proline-rich tyrosine kinase 2 (PYK2)		L13616	58.12	67.14	56.48	5
107 A08i	retinoblastoma-related protein p107		L14812	49.59	55.37	48.67	1
626 D03j	cadherin 8 (CDH8)		L34060	48.53	53.42	49.44	5
1060 F06j	sulfonylurea receptor (SUR); ATP-binding cassette subfamily C (CFTR/MRP) member 8 (ABCC8)		L78207	47.85	60.56	56.91	12
634 D04d	urokinase-type plasminogen activator precursor (U-plasminogen activator; UPA)		M15476; D00244	109.25	158.12	88.21	5
947 E12i	cytochrome P450 IID6 (CYP2D6)		M20403; M19697	62.59	66.19	66.73	10
707 D09g	erbB3 proto-oncogene; HER3		M29366; M34309	98.72	122.03	120.47	6

264	B05f	ras-related C3 botulinum toxin substrate 1 (RAC1); ras-like protein TC25	M29870; M31467	71.53	102.13	99.2	3
835	E04i	xeroderma pigmentosum group B complementing protein (XPB); excision repair cross-complementing rodent repair deficiency complementation group 3 (ERCC3); basal transcription factor 2 89-kDa subunit (BTF2p89; TFIIH 89-kDa subunit)	M31899	61.94	69.94	79.37	9
682	D07j	insulin-like growth factor-binding protein 2 (IGF-binding protein 2; IGFBP2; IBP2)	M35410	54.76	55.93	54.53	6
878	E07j	X-ray repair-complementing defective repair in Chinese hamster cells 1 (XRCC1)	M36089	58.85	64.71	65.04	9
759	D13c	interleukin 11 (IL11); adipogenesis inhibitory factor (AGIF)	M57765	50.13	52.49	49.78	7
631	D04a	interleukin 3 receptor beta subunit precursor (IL3R-beta; IL3RB); interleukin 5 receptor beta subunit precursor (IL5R-beta; IL5R); cytokine receptor common beta chain precursor; CSF2RB	M59941	45.68	54.16	47.26	5
330	B10h	colorectal mutant cancer protein (MCC)	M62397	45.35	52.22	50.13	2
511	C09g	integrin beta 7 precursor (ITGB7)	M62880; S60335	48.56	50.49	54.54	5
218	B02h	ERBB2 proto-oncogene; NEU proto-oncogene; HER2	M95667; M11730	46.39	48.22	49.65	2
744	D12b	tumor necrosis factor type 1 receptor-associated protein (TRAP1)	U12595	123.01	128.36	123.5	7
252	B04n	180-kDa secretory phospholipase A2 receptor precursor (PLA2 receptor; PLA2R)	U17093	44.51	49.65	51.3	2
832	E04f	ionizing radiation resistance-conferring protein + death-associated protein 3 (DAP3)	U18321 + X83544	55.81	66.14	68.87	8
291	B07k	ras-related protein RAB5C	U18420	93.66	125.57	113.8	3
332	B10j	EB1 protein	U24166	73.77	115.63	98.15	2
886	E08d	histone deacetylase 2 (HD2)	U31814	49.42	59.45	58.51	9
135	A10i	ataxia telangiectasia (ATM)	U33841	134.58	226.59	157.96	2
649	D05e	beta 3-endonoxin	U37139	51.5	56.14	50.77	5
888	E08f	DNA topoisomerase III alpha (TOP3A)	U43431	50.53	57.87	55.38	9
368	B13d	serine/threonine protein kinase PRP4 homolog	U48736	47.51	53.66	54.03	2
1005	F02k	gamma-glutamyl hydrolase precursor (GGH; GH); folic polygamma-glutamyl hydrolase; gamma-glu-X carboxypeptidase; conjugase	U55206	51.55	61.02	57.28	10
773	D14c	MHC class I molecule; MHC class I subunit-related protein A (MICA); PERB11.1	U65416; U69631; X92841	59.07	65.06	57.26	7
343	B11g	TSG101 tumor susceptibility protein	U82130	47.86	53.9	54.66	2
440	C04f	RHO GDP-dissociation inhibitor 3 (RHO-GDI 3); RHO-GDI gamma; ARHGDI3	U82532	43.18	49.86	54.6	4

1025 F04c	glutathione-S-transferase homolog	U90313	158.28	387.66	332.95	11
858 E06d	B kappaB epsilon (IKBE)	U91616	50.96	54.03	54.2	9
915 E10c	dihydrofolate reductase (DHFR)	V00507	51.46	60.56	56.53	9
698 D08l	transforming growth factor beta (TGF-beta; TGFB)	X02812; J05114	48.33	54.57	49.24	6
131 A10c	c-kit proto-oncogene; mast/stem cell growth factor receptor precursor (SCFR); CD117 antigen	X06182	50.01	54.86	51.94	2
304 B08j	retinoic acid receptor alpha 1 (RAR-alpha 1; RARA); PML-RAR protein	X06538; X06614; M73779	43.83	53.11	49.08	2
301 B08g	ras homolog gene family member B (RHOB; ARHB); ARH6; H6	X06820	63.67	72.7	64.22	3
771 D14a	interferon regulatory factor 1 (IRF1)	X14454	46.12	55.7	49.83	7
737 D11i	interleukin 9 precursor (IL9); T-cell growth factor p40	X17543; M30134	49.12	52.53	50.99	7
1032 F04j	heat shock 70-kDa protein 6; heat shock 70-kDa protein B	X51757; M11236	51.67	62.82	59.96	11
428 C03h	rho-related GTP-binding protein (RHOG); ARHG	X61587	49.43	54.38	56.89	3
894 E08i	DNA cytosine-5-methyltransferase (DNA metase; MCMT)	X63692	48.84	55.42	52.67	9
345 B11i	serine/threonine-protein kinase PCTAIRE 1 (PCTK1)	X66363	52.6	61.98	60.95	2
425 C03e	RHO GDP-dissociation inhibitor 1 (RHO-GDI 1); RHO-GDI alpha (GDIA1); ARHGDIA	X69550	48.1	60.51	59.95	4
983 F01c	uridine phosphorylase (UDRPase; UP)	X90858	64.36	81	89.91	10
1042 F05f	bleomycin hydrolase (BLM hydrolase)	X92106	51.85	67.24	66.76	11
279 B06m	ras-related protein RAB28	X94703	47.43	50.13	50	3
962 E13j	adenine phosphoribosyltransferase (APRT)	Y00486	47.29	57.67	53.84	10
982 F01b	nm23-H4; nucleoside-diphosphate kinase; nucleoside 5'-diphosphate phosphotransferase (NDK)	Y07604	112.79	135.9	131.1	10
448 C04n	rab geranylgeranyltransferase alpha subunit (rab GG transferase; rab GGTase)	Y08200	57.14	59.58	58.83	4



フィルターアレイの信頼性および抗癌剤感受性遺伝子に関する研究  
分担研究者 竹内 正弘 北里大学大学院 薬学研究科 臨床統計部門 教授

研究要旨 フィルターアレイの信頼性に関する検討と、エストロゲンおよびタモキシフェン接触処理によって乳癌細胞株において発現に変化が生じる遺伝子の抽出を行なった。その結果、フィルターアレイの信頼性が比較的良好であることが明らかになった。また、タモキシフェン感受性が示唆される遺伝子が抽出された。

#### A. 研究目的

マイクロアレイを用いた遺伝子発現解析を臨床応用する際には、その結果の信頼性を評価することが重要である。なぜならば、マイクロアレイを用いた遺伝子発現解析の信頼性を疑問視する意見が多く、また臨床応用においては、解析結果に混入する偏りや変動が過度に大きいと、患者の治療に悪影響を与えかねないからである。本解析の目的のひとつは、クロンテック社のフィルターアレイの信頼性に関する検討を行うことにある。本研究でクロンテック社のフィルターアレイを使用した理由は、他のアレイに比べてフィルターアレイが比較的大型であり、またアレイ上にスポットされているプローブ数が約 1000 個と少ないことから、より高い結果の再現性が期待されるからである。

また、乳癌細胞株に対する抗癌剤の影響はまだ未知の部分が多く、その遺伝子レベルでの解明が望まれている。特にホルモン治療薬における乳癌細胞株への遺伝子発現シグナル解析は重点的に研究が進められている領域であり、乳癌における抗癌剤の作用機序を理解する上で重要である。本研究のもうひとつの目的は、エストロゲンおよびタモキシフェン接触処理によって乳癌細胞株の発現に変化が生じる遺伝子を統計学

的に抽出することである。

#### B. 研究方法

##### 1. 実験方法

本研究ではクロンテック社のアトラスフィルターアレイを使用して、代表的な乳癌細胞株に対してエストロゲンおよびタモキシフェン接触を行ったときの遺伝子発現変動を測定した。今回使用したアトラスフィルターアレイは 825 種類の遺伝子に関する発現情報を一度に測定することが可能である。本実験では BT474、MCF7、MDA-MB-231、SK-BR-3 という 4 種類の代表的な乳癌細胞株に対して、それぞれ 4 種類のエストロゲン接触処理を行った。4 種類の処理はそれぞれ、エストロゲンのみ接触、エストロゲンとタモキシフェン低用量に接触、エストロゲンとタモキシフェン高用量に接触、およびエストロゲンフリー（エストロゲン枯渇状態）である。従って、細胞株と処理の組み合わせは計 16 通りである。本実験ではアレイデータの信頼性を評価するために、この 16 通りの処理を 2 回繰り返して実施した。通常、アレイデータ解析における繰り返しは推定値の精度を高めるために行われるため、同一実験日に同一条件の実験を繰り返すが、今回の目的は信頼性の評価であることから、2 回の実

験はあえて実験日を変えて行った。アレイデータの信頼性はこれら 2 回の繰り返し実験における結果間のピアソン相関係数を計算することによって評価する。フィルターアレイは RI による単独標識を用いるため、本実験で使用したアレイの数は 32 枚である。実験の順序に伴う系統的変動を除去する目的で、本実験では各 16 種類の組み合わせに対するハイブリダイゼーション処理はランダム化した。従って、統計学的には、本実験は繰り返し数 2 の 4 細胞株×4 処理の要因実験とみなすことができる。表 1 は細胞株と処理の組み合わせを示したものである。なお、実験は単独の実験者によって行われ、RI 使用に関する制約から実験は 8 回に分けて行われた。

## 2. データの前処理

図 1 は細胞株と処理の 16 種類の組み合わせにおける遺伝子発現強度の散布図である。横軸は 1 回目の実験における遺伝子発現強度、縦軸は 2 回目の実験における発現強度である。この図より、実験間で発現強度の位置と尺度が大きく変化していること、および発現強度の分布が右上に大きく裾を引くことがわかる。このようなデータに対してはピアソンの相関係数を用いても信頼性を妥当に評価することはできず、また、後述する薬剤感受性遺伝子の抽出においても、統計解析上困難が生じる。従って、統計解析のためには位置・尺度の調整だけではなく、適当な変数変換が必要である。

図 2 は遺伝子発現強度の遺伝子毎の平均と標準偏差について、散布図をとったものである。大きな点はデータ系列に対する移動平均である。図より、平均強度の増加に

伴って標準偏差が著しく増加することがわかる。

各組み合わせの発現強度が比較可能になるように尺度を調整し、かつ平均強度とその標準偏差が無関係になるようにデータを変換することにより、エストロゲン接触関連遺伝子を統計学的に抽出することが比較的容易になる。通常のマイクロアレイ解析では、データに対して対数変換を施し、平均値あるいは中央値を用いて正規化を行なうのが一般的である。しかし、本データにおいて上記の正規化を試みたところ、以前として平均強度とその標準偏差に強い相関が認められるため、従来の方法とは異なる正規化の手法が必要になる。

本解析では、Huber(2002)による正規化および分散安定化変換の手法をフィルターアレイのデータに対して適用した。アレイ  $i(=1, \dots, m)$  における遺伝子  $j(=1, \dots, n)$  の発現強度を  $x_{ij}$  とする。アレイ  $i$  における位置および尺度の調整用パラメータをそれぞれ  $o_i$  と  $s_i$  とする。このときデータの変換を、

$$x_{ij} \propto z_{ij} = o_i + s_i x_{ij}$$

で定義する。一般性を失わず  $o_i=0$  と  $s_i=1$  とすることができるので、推定すべきパラメータ数は  $2m-2$  個である。次に分散を安定化させるために、発現強度の平均と標準偏差にモデルを仮定する。図 2 より、平均と標準偏差には線形関係が仮定できるので、この場合の分散安定化変換は、

$$h(y_{ij}) = \operatorname{arcsinh}(a_i + b_i y_{ij})$$

として得ることができる。ここで  $a_i = a + b o_i$ 、 $b_i = b s_i$  であり、 $a$  と  $b$  は分散安定化に用いるパラメータである。 $a_i$ 、 $b_i$  の推定はプロファイル尤度法で行なった。ここで、当フィルターアレイデータの解析にはオープンソー

スの統計解析ソフトウェアである R、およびその遺伝子データ解析パッケージである Bioconductor を使用した。

### 3. 薬剤感受性遺伝子の抽出

次に、エストロゲン接触によって発現強度に違いが生じる遺伝子を統計的に抽出した。解析は Scholtens and Gentleman のアプローチに基づいて行なった。当実験における発現強度データを  $y_{ijkl}$  とする。ここで  $i$  は遺伝子 ( $i=1, \dots, 825$ )、 $j$  は細胞 ( $j=1, \dots, 4$ )、 $k$  は処理 ( $k=1, \dots, 4$ )、 $l$  は繰り返し ( $l=1, 2$ ) とする。当実験の発現強度データに対して、下記の分散分析モデルを遺伝子毎に適用した。

$$y_{ijkl} = \mu_i + C_{ij} + T_{ik} + CT_{ijk} + e_{ijkl}$$

ここで  $\mu_i$  は主効果、 $C_{ij}$  は細胞の効果、 $T_{ik}$  は処理の効果、 $CT_{ijk}$  は細胞と処理の交互作用である。次に、処理効果  $T_{ik}$  の検定を行なうために次の帰無モデルを用いた。

$$y_{ijkl} = \mu_i + C_{ij} + e_{ijkl}$$

処理効果の  $p$  値は  $F$  検定で導出した。また、細胞と処理の交互作用  $CT_{ijk}$  の検定を行なうために、次の帰無モデルを当てはめた。

$$y_{ijkl} = \mu_i + C_{ij} + T_{ik} + e_{ijkl}$$

発現強度データには依然として実験間で系統的な影響が生じていたことが探索的な検討から判明したため、実際の解析では上記の全てのモデルに実験の効果  $E_m$  ( $m=1, \dots, 8$ ) を加えて解析を行なった。従って、解析に用いたモデルはそれぞれ下記の通りである。

$$y_{ijkl} = \mu_i + E_m + C_{ij} + T_{ik} + CT_{ijk} + e_{ijkl}$$

$$y_{ijkl} = \mu_i + E_m + C_{ij} + e_{ijkl}$$

$$y_{ijkl} = \mu_i + E_m + C_{ij} + T_{ik} + e_{ijkl}$$

上記の解析では一度に 825 回の  $F$  検定を

行なうため、検定の多重性の問題が当然生じる。多重性の問題を考慮するために、本解析では Benjamini and Yekutieli (2001) の多重比較法を用いた。Benjamini and Yekutieli 法（以下 BY 法と略す）は任意の従属関係を持つ検定統計量から得られた  $p$  値に対して、試験全体の第一種の過誤ではなく、False Discovery Rate（以下 FDR）を調整する多重比較法である。V を誤って棄却した真の仮説の数、R を棄却した仮説の数としたとき、確率変数  $Q$  を次のように定義する。

$$Q = \begin{cases} V/R & R > 0 \text{ のとき} \\ 0 & \text{その他} \end{cases}$$

このとき、FDR は  $Q$  の期待値  $E(Q)$  である。つまり、棄却された仮説に真の仮説が含まれる割合の期待値である。

任意の従属関係をもつ  $n$  個の検定統計量から得られた  $n$  個の  $p$  値を  $p_1, p_2, \dots, p_n$  とする。これらの  $p$  値を小さい順に並べ変えたものを  $p^{(1)}, p^{(2)}, \dots, p^{(n)}$  としたとき、

$$k = \max \left\{ i : p^{(i)} \leq \frac{i}{n} \frac{1}{\sum_{l=1}^n (1/l)} q \right\}$$

を計算し、 $p^{(1)}, p^{(2)}, \dots, p^{(k)}$  に対応する仮説を棄却する。この検定方式によって、FDR は  $q$  以下に調整される (Benjamini and Yekutieli, 2001)。なお本研究では FDR の基準値として 0.05 を選択した。

次に探索的な目的（仮説発見）のために、タモキシフェン介入に伴って系統的に変動する遺伝子発現プロファイルを各細胞株毎に解析する。解析においては平均値に関する線形傾向性検定を用いた。線形対立仮説を表現する線形対比、

$$L = \sum_k^K c_k \bar{Y}_k$$

に対する検定を考える。ここで  $k$  は処理を示す因子の水準数であり、線形対比はそれぞれ  $c_1=-3$ 、 $c_2=-1$ 、 $c_3=1$ 、 $c_4=3$  である。このとき検定統計量は、

$$T = \frac{\sum_{k=1}^4 c_k \bar{Y}_k}{\sqrt{10\sigma^2}}$$

であり、帰無仮説のもとで自由度 4 の  $t$  分布に従う。この検定を細胞株毎、遺伝子毎に適用する。なお本解析では仮説発見の目的のため、多重性の調整は行っていない。

上記の解析結果を視覚的に明確化するために、ボルケーノプロットを描画した。ボルケーノプロットは横軸に検定統計量の分子、縦軸に  $p$  値の対数の負値をプロットしたものである。本解析においては、図の右側に位置する遺伝子は介入に伴って発現が増加する傾向を、左側は減少する傾向を示す。

## C. 研究結果

### 1. データの前処理

図 3 は分散安定化変換後の各遺伝子の平均強度と標準偏差の関係を示す図である。図 3 より、分散安定化変換を行うことで、標準偏差は平均強度によらずほぼ一定になっていることがわかる。これより、本データに Huber の分散安定化手法を用いることで、通常の分散分析など、当分散性を仮定する解析手法を適用することが妥当であることが分かった。

図 4 は分散安定化変換後の実験間の散布図である。図 4 より、データの重心は各散布図の中心にほぼ位置していることがわか

り、変数変換前の右上への裾の広さが解消されていることがわかる。これにより、分布の非対称性に影響されやすい、正規分布を仮定した統計解析手法を本データに適用することが妥当であることが示された。

### 2. フィルターアレイの信頼性検討

本実験で使用したフィルターアレイの信頼性を検討するために、上記の分散安定化変換後のデータに対して、ふたつの実験間における発現強度のピアソン相関係数を各組み合わせ毎に計算した（表 2）。なお、分布の非対称性はピアソン相関係数に大きく影響することから、分散安定化前のデータに対してピアソン相関係数を計算することは望ましくないことを注記する。解析の結果、いずれの細胞と処理の組み合わせにおいても、ピアソン相関係数は 0.8 以上の値を示した。今回の実験のように、マイクロアレイ間で別々のハイブリダイゼーションを行なった場合、対応する発現強度間の相関は 60% から 80% であるという報告があり（Churchill, 2002）、それを考慮すると、この結果はフィルターアレイの信頼性が他のアレイに比べて高水準であることを示唆している。

### 3. 薬剤感受性遺伝子の抽出

Huber の分散安定化変換によってアレイデータの非対称性と不等分散性がほぼ解消されたので、先に紹介した分散分析法と対比を用いた方法を本データに対して適用した。分散分析に関しては、BY 法によって検定の多重性を調整した場合、主効果および交互作用の検定のいずれにおいても  $k=0$  であり、残念ながら、処理および処理と細胞