

Microarray Data Analysis Tool データ解析の進め方 Vol.5 - 2グループ間の比較解析 –

疾患グループ・正常グループなどのように、繰り返し実験を含む多数のサンプルを、2 つのグループに分け、統計検定も使用して、グループ間の発現量を比較する場合の 解析例を、Signal比較解析を使用して説明します。

解析テーマは

「正常グループ-腫瘍グループ間の、遺伝子発現プロファイルの探索」です。



グループ間で比較する。

上記のテーマについて解析を行うには、まず各サンプルを用いて、単色のマイクロアレイ実験を行い、 その後、本ソフトウェアのSignal比較解析機能を用いて、下記のような処理を行う必要があります。

①全実験データの正規化(ノーマライズ)

②各サンプルデータのグループ分け

③統計検定機能を用いて、グループ間の発現量の差の有意確率を計算

④グループ間で発現が変動している遺伝子を抽出

Vol.4で説明したように、Signal比較解析では、全実験データの正規化およびグループ分けを行うこと ができます。そしてグループ情報に従い、グループ間発現量差の有意確率を計算し、それらの値を用 いて、有意に発現が変動している遺伝子を抽出します。

1. ソフトウェアの起動と実験データの読み込み

<u>1-1. ソフトウェアの準備</u>

CDIこ入っている「Microarray Data Analysis Tool」のフォルダを、自分のコンピュータ にコピーしてください。

1-2. ソフトウェアの起動

コピーしたフォルダを開き、「MAtool.exe」の アイコンをダブルクリックしてください。 右図のタイトル画面が表示され、ソフトウェア が起動します。



(10) 22342(A 2010) 1	8820 1422			
0 🔂 🖬 💷			ý l	
ulmar Neter	Cowiner Name	Ter	Curtai	
_				
_				

ソフトウェアが起動後は右図のメイン画面が 表示されます。

<u>1-3. 実験データ読み込み</u>

アイコンをクリックすると、右の画面が表示されます。「Add」ボタンをクリックし、対象となる 実験データの選択・読み込みを行います。 なおSignal比較解析用の実験データファイルは、 1サンプル分の実験データしかもっていません。 そのためファイル名も「SampleA vs SampleB」で はなく、「SampleA」、「SampleB」のように1サン プルずつに分かれており、サンプル名の後ろに 「Signal比較解析用」と入力されています。



実験データを読み込むと、右図のように選択 したファイルの使用マイクロアレイ名、サンプ ル名等が表示されます。 ここでは「腫瘍サンプル1~3」と、「正常サン プル1~3」の、合計6つの実験データを読み 込みます。表示の確認が終わったら、「Next」 ボタンをクリックして次へ進みます。



2. 実験データの正規化とグループ分け

<u>2-1. 正規化手法の指定</u>

「Next」ボタンをクリックすると、「Data Edit」(下図)の画面が現れます。この画面では正 規化の手法や、比較の組み合わせの選択を行います。各項目の詳細については、本ソ フトウェアのユーザーマニュアル、またはVol.4を参照してください。

🧏 Data Edit
Low data treatment
> Net intensity -> Net intensity
*0.001以下の数値は自動的に0.001に切り上げられます。
Per Array Normalization 🔽 Per Gene Normalization
C Median
C Mean
Quantile C Control experiment
C Control gene コントロールの指定をする
コントロールの指定をする
Ratio calculation 比較の組み合わせを指定する
Next Cancel

ここでは「Per Array Normalization」に「Quantile」を選択し、「Per Gene Normalization」 に「Median」を選択しました。本解析のように、クラスタリングや発現量の統計検定を行 う場合は、必ずPer Gene Normalizationの設定を行ってください。 また本解析のように、サンプル同士の直接の比較を行わず、グループ間で比較を行う 場合は、「Ratio calculation」の設定は必要ありません。

<u>2-2. グループ化</u>

Data Edit画面から進むと、「Group Edit」画面が現れます。ここで実験データをグループ分けすることで、グループ単位の検索や統計検定が実行可能になります。

🧱 Group Edit	
グループ機能 ⓒ 使う ○ 使わない	
グループ編集 新規登録 新規登録 新し、ヴループを登録します。 名称変更 選択中のグループの名称を変更します。 削除 グループを削除します。	<u>業集</u> 正常
グループ登録 グループ未登録実験: 構成サンフルー 本語サンプルー3 正常サンプルー2 正常サンプルー3 (< Remove 正常サンプルー3	グルーブ: 腫瘍
) 色: 100000000
「Group Ratio calculation 比較の組み合わせを指定する」	
	Next Cancel

ここでは2つのグループを作成し、一方に「腫瘍サンプル-1~3」を、もう一方に「正常サンプル-1~3」を設定しました。

グループ分けの設定を行ったら、次に「Group Ratio caluculation」の設定を行います。 本解析のように、2つ以上のグループを作成した場合、各グループ内の平均発現量 (Per Array Normalized intensity)を用いて、各遺伝子のグループ間発現比を計算する ことができます。

aroup Ratio calculation 比較の組み合わせを指定する		
クリック!	Select デストグループ名 ○ 推移 正常 All 全てにチェックをみる	 ユントロールグルーブ名 V3 正常 V3 厳痛 N4ます。
	Delete 全てのチェックを外	NUます。 OK Cancel

「Ratio calculation」と同様、「Group Ratio caluculation」のチェックボックスにチェックを 入れた後、「比較の組み合わせを指定する」という青色の文章をクリックすると上図の ような画面が表示されますので、比較をしたいグループの組み合わせを選んでくださ い。本解析では、「腫瘍グループ vs 正常グループ」の組み合わせで、発現比を計算す るように設定します。

「Next」ボタンをクリックすると、2つ目の「Group Edit」画面(下図)が現れます。ここでは、 計算したRatioデータをグループ分けできるのですが、本解析では「腫瘍グループ vs 正常グループ」の1つしかRatioを計算していないため、何も設定せず次に進みます。

🧱 Group Edit	
Radio dataグループ機能	
 使う で使わない 	
Ratioグループ編集	
新規登録 新しいグループを登録します。	
名称変更 選択中のグルーブの名称を変更します。	
<u> </u>	
Ratioグループ登録	
グループ未登録Ratio data	Ratioグループ:
Add ->	
<- Remove	
	色:
Group Ratioグループ編集	
新規登録 新しいヴルーブを登録します。	
名称変更 選択中のグループの名称を変更します。	
Group Ratioグループ登録	
グループ未登録Group Ratio data 睡痒 vo 正常	Group Ratioグループ:
Add ->	
<- Remove	
	<u>ا</u>
	Next Gancel

2つ目のGroup Editから進むと、表示するフィールドの選択画面が現れます。ここでは、以下の4項目を表示するように、設定します。

Per Array Normalized intensity: Per Gene Normalized intensity: Quality flag: Ignored cells: Per Array Normalization後のintensity Per Gene Normalization後のintensity 各スポットにつけられたフラグ 0:解析対象スポット、1:解析除外スポット

表示するフィ	ィールド名を選択してください
Field1:	Per Array Normalized intensity
🔽 Field2:	Per Gene Normalized intensity
🔽 Field3:	Quality flag
🔽 Field4:	Ignored cells
Field5:	Raw intensity
🔲 Field6:	Raw intensity
	OK Cancel

「OK」ボタンをクリックすると、選択したフィールドのデータが、サンプルごとにメイン画面に表示されます。またグループ設定を行っている場合は、実験データはグループごとに分類されて表示されます。

ファイル(火) グラ:	Data ana) 2表示(W) 道	HR© ≯	として と で の の の の の の の の の の の の の	3.2 AJU	プロン	n,una												
001	d	Contract of	7		, I	P	- E		휬 🛅		•	検索結果	R: 54359 (4					
MicroArray Nam	e CodeLink(TM) Hum	an Whole	G E	xperiment	Name			Test			Control			_			
	[1				腫瘍サンブル	腫瘍サンプル	腫瘍サンブル	経病サンプル	腫瘍サンブル	腫瘍サンブル	腫瘍サンプル	腫瘍サンプル	腫瘍サンプル	腫瘍サンプル	腫瘍サンブル	腫瘍サンブル	正常サンプル
No.	Select	Block	Row	G	olumn	Per Array No	Per Gene No	Quality flag	Ignored cells	Per Array No	Per Gene Nor	Quality flag	Ignored cells	Per Array No	Per Gene Nor	Quality flag	Ignored cells	Per Array No
1	0		1	1	1	8749.8671	0.04686	0	0	9958.8478	0.23358	G	0	8367.97223	-0.01752	G	C	8157.38517
2	0	1	1	1	2	643.77835	-0.40489	0	1	1011.52817	0.24701	G	0	908.34463	0.09178	G	0	684.14382
3	0		1	1	3	102.01192	-0.10928	6	0	233.79253	1.08722	1	1	118.0658	0.10158	G	C	63.36395
4	0		1	1	4	119.5662	0.18301	6	0	280.81735	1.41483	G	0	82.08082	-0.35969	G	C	91.07723
5	0		1	1	5	70.59253	-0.24062	l	. 0	187.09183	1.16554	G	0	108.78235	0.38324	G	C	42.01358
6	0		1	1	6	116.9937	-0.01278	G	0	166.5947	0.49713	G	0	119.07563	0.01267	G	C	72.10218
7	0		1	1	7	90.97487	-0.39177	l	0	205.98362	0.78721	G	0	128.98348	0.11187	G	C	49.546
8	0		1	1	8	8115.02488	-0.04432	0	0	8621.30263	0.04299	G	0	7916.39753	-0.08007	G	0	7629.60083
9	0		1	1	9	213.19323	0.06983	0	0	222.1442	0.12917	G	0	238.86552	0.23387	G	0	74.93637
10	0		1	1	10	457.1823	0.03596		0	654.70012	0.55403	G	0	498.01488	0.15938	G	0	363.23773
11	0		1	1	11	74.54337	-0.19831		. 0	52.12703	-0.71435	G	0	76.38762	-0.16305	L	C	94.66698
12	0		1	1	12	118.98725	-0.71028	6	0	228.4202	0.23061	G	0	253.36588	0.38014	G	0	199.3143
13	0		1	1	13	97.971	0.07363		. 0	88.2223	-0.07759	G	0	114.54962	0.29917	G	C	86.11587
14	0		1	1	14	142.11725	0.04617	6	0	228.72418	0.7327	G	0	156.23525	0.18281	G	0	103.17162
15	0		1	1	15	7587.06307	0.05857	6	0	7243.34913	-0.00831	G	0	6146.34783	-0.24524	G	C	6329.2321
16	0		1	1	16	158.64982	-0.08796	0	0	192.78053	0.19315	G	0	181.48937	0.10608	G	C	99.94723
17	0		1	1	17	152,72017	0.06209		0	171.53138	0.22968	G	0	139.85283	-0.06489	G	C	109.34903
18	0		1	1	18	122.12018	0.097		0	114.0953	-0.00106	G	0	114.2628	0.00106	G	C	102.58497
19	0		1	1	19	52.31333	0.00937	1	. 0	55.32192	0.09004	G	0	46.34333	-0.16545	L	C	51.63607
20	0		1	1	20	92,43698	-0.03383		. 0	96.82397	0.03306	G	0	48.00707	-0.97906	L	C	120.77595
21	0		1	1	21	589.32775	0.13188		0	564.93425	0.07089	G	0	495.30008	-0.11889	G	C	391.1164
22	0		1	1	22	7307.53295	0.10656	0	0	6747.2308	-0.00853	G	0	5747.86492	-0.2398	G	C	6827.28108
23	0		1	1	23	425.50772	0.09391	0	0	512,93318	0.3635	G	0	371.87537	-0.10045	G	0	263.06582
24	0		1	1	24	32.7635	-0.8078	l	. 0	108.22402	0.91606	G	0	103.71152	0.85461	G	0	31.9568
26	0		1	1	25	54.50505	-0.48114	· .	0	82.75783	0.12136	G	0	134.70053	0.82415	G	0	69.40373
26	0		1	1	26	78.20147	-0.00012		. 0	78.2147	0.00012	G	0	47.97878	-0.70492	L	0	79.41422
27	0	-	1	1	27	1566.95953	-0.05839		0	1825.72568	0.16211	G	0	1649.05027	0.01528	G	0	1171.79068
28	0	-	1	1	28	71.46835	0.36533		0	32.30387	-0.78027	1	0	58.06245	0.06563	L	0	52.89787
29	0	-	1	1	29	8547.63728	0.02075	0	0	8303.5764	-0.02105	G	0	7712.56227	-0.12757	G	0	11628.38915
30	0	-	1	1	32	2229.96838	-0.05075	0	0	2007.1503	-0.20262	G	0	1696.9663	-0.45342	G	0	2389.63443
21	0	-	1	1	33	57.94953	-0.1102		0	12.45948	-2.32775	L	0	28.6047	-1.12875	L	0	74.62293
01								-		100000		-	-				-	

3. 統計検定の実行 3-1. 検定手法の選択	(愛 Microarray D ファイルの グラフ表 () () () () MicroArray Name	ata Analy 示W 選	<mark>ysis Too</mark> 訳② 検 ■ ■ M) Humar	I Ver3.: 索① ^ I Whole G	2 Filgo ルプロ Experiment	e n,Inc	<u> </u>		Gio Test
をクリックしてください。						腫瘍サンプル	腫瘍サンプル	腫瘍サンブル	腫瘍!
	No.	Select	Block	Row	Column	Per Array No	Per Gene Nor	Quality flag	Ignore
	1	0	1	1	1	8749.8671	0.04686	G	i
	2	0	1	1	2	643.77835	-0.40489	C	
※ 統計検定 「 統計検定 「 Student's t-test Cone-Way ANOVA Cone-Sample t-test 対象とならグループ: 「 Ignored cells 老時外する) 「 FID Geniamiai and Hochere/設計算する。									
OK 1. Lou rocitatini all'i loci rocitata 8.2.2	Cancel								

すると上図のように、実行する統計検定手法とオプション項目の選択画面が表示されます。

使用する検定手法は、解析を行うデータのプラットフォームやグループ数によって、以下の ように使い分けます。

- •Student's t-test : 2グループ間比較
- •One-Sample t-test : 1グループ比較(2色法の場合)
- •One-Way ANOVA : 3グループ以上の比較

また2つのオプション項目は、それぞれのチェックボックスにチェックを入れることによって、 以下の機能が使用できるようになります。

 Ignored cellsを除外する: P-value計算時に、解析除外スポット(Ignored cellsが1の スポット)のデータを使用しない。

•FDR (Benjamini & Hochberge)を計算する: P-valueと同時に、FDR (False Discovery Rate)も計算する。

なお、本解析では2グループの比較解析になるので、使用する検定手法として「Student's t-test」を、オプション項目として「Ignored cellsを除外する」を選択しました。

上記のように設定後、「OK」ボタンをクリックすると、メイン画面の「P-value」列に、各遺伝子のグループ間発現量の差を表す、P-valueが表示されます(下図)。

🎆 Microarray Da	ata Analysis '	Tool Ver3.	2 Filge	n,Inc				
ファイル(V) グラフ表:	示──── 選択──	検索(⊻) ^	リレプ(乙)					
	5		F	P	E÷		👷 📴 📴 📀	検
MicroArray Name 🛛	CodeLink(TM) H	uman Whole G	Experiment	Name 🗌			Test	Co
	正常サンブル	正常サンブル	正常サンプル	正常サンブル	腫瘍 vs 正常	腫瘍 vs 正常	Student's t-test	Student's t-test
No.	Per Array No	Per Gene Nor	Quality flag	Ignored cells	ratio	log2ratio	P-value	DR
1	8572.50823	0.01731	G	0	1.07964	0.11055	0.242953311756289	
2	916.77918	0.10512	G	0	1.20128	0.26457	0.202342912595338	
3	131.35457	0.25545	G	0	1.25724	0.33026	0.421228263974263	
4	140.12423	0.4119	G	0	1.80163	0.8493	0.332858890758491	
5	96.21778	0.20617	L	0	1.77299	0.82619	0.212607603526305	
6	157.45542	0.41573	G	0	1.49626	0.58136	0.258609568543353	
7	169.07073	0.50231	G	0	1.29721	0.37541	0.495648652537351	
8	8943.11052	0.09587	G	0	0.9643	-0.05244	0.597084855732903	
9	193.04647	-0.07338	G	0	1.57124	0.6519	0.145579454953325	
10	434.67242	-0.03688	G	0	1.30659	0.38581	0.104383004885151	

4. 発現変動遺伝子の抽出

<u>4-1. Filter Options検索画面の表示</u>

メイン画面のFilter Optionsアイコン 🎤 を クリックし、Filter Options検索画面を開き ます。



Filter Options検索画面では各種検索条件で データの抽出を行うことが出来ます。

> keyword 機能カテゴリー 実験データ 染色体番号 遺伝子セット

fite	e Optiere				
earch (Ceton 油G子セット支持				
tay ki	ad				
Γ.	NGTOTES File	2		C SU C SIGUE	# AND # CO
E	\$51071:00 Fine	2			@ AND @ OR
	NOTION FILE	2		 @ St: C Skin: 	R AND C CR
E.	NOTION FILE	2			R AND C 08
г	WHEPER Five	2		 e su e suu : 	12 大文中小文中世区的LAL1
- 11	unction MRES:	MONOTESTO - MO	110000000000000000000000000000000000000		I F BEFER
No	177:37		937730		So Select db27trst
-					1
-					
Osolit,	y flet				
Π.		(C 301.1)			
AM 2	7-3				
г	Pare Mensily Boot	2 0	NE CP	LIT & AND C OR	
					Choeccane
		2 0 0	NE CP	LITE & AND C OR	No Select Oracessone
					0
			KLE C P	LIF & MD C OR	1 1
Ε.	Page Manufacture Read	- C D	HE CO	IT & AND C OR	
	Description in the second	1.05	- MA	100	
		21 - 1	NUE - P	T ILDREAD	F Frahe
					C NF C N2
T)					E 108.
		clo n	2 () 1	IT GAND COR	C 121 C 137
	Correction of the later				- lencend cells
	L. Unite runs and there	ep u	E CP I	17	E largered cells \$3875.7.5
	level				Scaturaler## 05 Cancel
					200 million 1

<u>4-2. 検索条件を設定</u>

ここでは下記を目的を達するための条件設定を行います。 ①コントロール遺伝子などを除去し、解析対象遺伝子のみにする。 ②グループ間で発現差のあるデータを抽出する。 ③ノイズレベルの信頼性の低いデータを除外する。(CutOff)

条件設定の方法

①コントロール遺伝子などを除去し、解析対象遺伝子のみにする。

ここでは、keyword検索を使用します。事前にアレイに搭載されているプローブは 「0、2」という数字で識別をしています。 「0」は解析対象遺伝子 「2」は解析対象外遺伝子(Negative Control,Blank Spot,マーカーコントロールなど)

まずは、左のチェックボックスをクリックし、 該当項目で「解析対象 Flag」と、「0}を選択してください。これで、解析対象遺伝子 のみを抽出する条件設定が出来ました。

	Filter	Options					
S	earch Op	otion 遺伝子セット登録					
	KeyWor	d					
	•	解析対象 Flag	•	0	● 含む ○ 含まない	AND C OR	
	Г	解析対象 Flag	*			🖸 AND 🔿 OR	
	Γ	解析対象 Flag	Y		€ 含む € 含まない	C AND C OR	
	Γ	解析対象 Flag	v		€ 含む € 含まない	C AND C OR	
	Г	解析対象 Flag	-		€ 含む € 含まない		▼ 大文字小文字を区別しない
				_			

<u>②グループ間で発現差のあるデータを抽出する。</u>

グループ間で発現差のあるデータを抽出する場合、グループ間のFold Changeと、統計検定のP-value(or FDR)で抽出する場合の2種類があり、この2つは同時に使用することができます。以下、それぞれの設定方法を紹介します。

<u>②-1 Fold Changeで検索する。</u>

4ページで指定したグループ間の発現比の検索には、「Fold Change」検索を使用します。 「Ratio data 比較」ではサンプル間の発現比、「Group Ratio data 比較」では実験グルー プ間の発現比の検索条件の指定を行います。本解析ではサンプル間発現比の計算を行っ ていませんので、「Ratio data 比較」の項目は指定できないようになっています。

左のチェックボックスをクリックし、検索条件を入力します。(下図は「Group Ratio data 比較」を2以上にした例です。)

続いて、「複数データの検索条件を指定してください」の項目を設定してください。本解析で はGroup Ratioとして、「腫瘍グループ vs 正常グループ」だけしか計算していませんが、こ のFold Change検索では、Ratioデータが1つだけしかなくても、複数データの検索条件の指 定を求められるようになっているので、必ず指定を行ってください。

~	Fold Change	cp	以上	с <mark>0</mark>	以下	AND	C OR	
	✓ Group Ratio data 比較 すべての実験において条件を	○ 2 満たす場合	以上	с <u>Р</u>	以下			

<u>(2)-2 P-value(FDR)で検索する。</u>

6ページで計算したグループ間発現量差のP-value(or FDR)の検索には、「P-value」「FDR」 検索を使用します。それぞれの入力欄に数値を入力し、検索を実行すると、メイン画面上 の「P-value」または「FDR」のデータに対して、指定の条件で抽出を行うことができます。な お、本解析ではP-valueのみ計算を行っているので、ここでは「P-value」が「0.01」以下と設 定します。

▼ P-value			
C 0	以上	 0.01 	以下
FDR			
C D	以上	C 🛛	以下

なお本解析では、Vol.2~4では設定を行った、Filter Optionsの「Ignored cellsを除外す る」機能を使用しません。この機能を使用すると、全実験データ(本解析では6つ)のうち 1つでも解析除外スポットがある遺伝子は、検索時に除去されてしまうからです。 またGroup Ratioの計算時には、自動的に解析除外スポットのデータは、計算に使用し ないようになっているうえ、統計検定のP-value(or FDR)においては、6ページで説明し たように、計算時に解析除外スポットを除去することができます。そのためグループ間 比較において、「Group Ratio」や「P-value(or FDR)」で検索を行う場合は、Filter Options の「Ignored cellsを除外する」に、チェックを入れないことを推奨しています。 <u>③ノイズレベルの信頼性の低いデータを除外する。(CutOff)</u>

ここでは1比較解析のときと同様、1. Quality flagを用いる場合 2. Intensityを用いる場合 の、2通りの解析手順を紹介します。 1・複数比較では、「G,G」のように2サンプル分のflagが、カンマで区切って記入されていま

すが、Signal比較解析用の実験データでは、「G」のように1サンプル分のflagしか記入されていません。そのためSignal比較解析では、検索文字列としてカンマを使用しないで下さい。

<u>③-1 Quality flagを用いてCutOffを行う。</u>

<u>例1. 少なくとも1つの実験データで、Quality flagが「G」の遺伝子を抽出する場合</u>

項目「Quality flag」で「Gと等しい」と指定し、複数データの検索条件の選択において、「1つの実験でも条件を満たす場合」を選択します。

Quality flag 「 G (等し) C含む 1つの実験でも条件が満たす場合

例2. すべての実験データで、Quality flagが「G」の遺伝子を抽出する場合

項目「Quality flag」で「Gと等しい」と指定し、複数データの検索条件の選択において、「すべての実験において条件を満たす場合」を選択します。

 G
 G
 ○ 等しい ○含む

 すべての実験において条件を満たす場合

例3.1つのグループでも、Quality flagが「G」の遺伝子を抽出する場合

項目「Quality flag」で「Gと等しい」と指定し、複数データの検索条件の選択において、グルー プ単位(「内1つの実験でも条件を満たすかどうか」 or 「内すべての実験で条件を満たすか どうか」)の、「1つのグループでも条件を満たす場合」を選択します。

Quality flag

*この検索条件は、実験データをグループ化していないと、使用できません。

上記のようにCutOffを設定し、さらに②の発現変動遺伝子抽出の設定を同時に行うことで、発現変動を示し、なおかつノイズレベルのデータを除去した遺伝子のみを抽出することができます。

6サンプル全ての実験データで、Quality flagが「G」のデータのみを解析に使用すれば、 データの信頼性は上がりますが、場合によっては条件が厳しすぎて、抽出される遺伝子 が少なくなってしまうことがあります。例1~3を参考に、必要に応じて抽出条件の変更を 行ってください。

<u>③-2 intensityを用いてCutOffを行う。</u>

Signal比較解析でintensityを用いてCutOffを行う場合は、複数比較の場合と同様に、サン プルごとのネガティブコントロールの値が違うため、実験データ共通のCutOff値を決めなけ ればいけません。以下にその手順を示します。

まずSummarizationを行い、データ内のネガティブコントロールスポット蛍光強度の平均 値を求めます。

この処理を行うには、メイン画面上のSummarizationアイコンをクリックし、対象とする アノテーション項目に、「Negative Control」を選択してください(下図)。この状態で「OK」ボ タンをクリックすると、計算が開始され、処理が終了するとメイン画面が表示されます。 なおSummarizationを実行すると、統計検定で計算したP-value(or FDR)のデータは一度 リセットされます。その場合は、Summarization実行後に、再び統計検定を実行し、Pvalue(or FDR)を計算してください。

🕅 Summarization 📃 🗖 🔰
基準とするアノテーション項目を選択してください。 選択された項目列における同一文字列の実験データのまとめを行います。
Negative Control
Summarize Options
Mean
C Median
▼ Log2RationからRatioを再計算する。
OK Cancel

次にメイン画面のデータの中から、Summarizationを行ったネガティブコントロールのデータを抽出します。

メイン画面上のFilter Optionsアイコン 「Negative Control」、右側の入力欄には「O」を指定します。指定を行うには、▼(下図)の 部分をクリックし、表示される項目の中から目的の項目名をクリックしてください。項目の 指定が終わったら、「OK」ボタンをクリックし、抽出を実行します。

🧱 Filte	r Options		
Search C	Dption 遺伝子セット登録		
⊢KevWa	ord		
	Negative Control	(_)	⊙ 含む ○ 含まない
	解析対象:0	<u>_</u>	

それぞれのサンプルデータについて、ネガティブコントロールの値が表示されます。通常 CutOffは、正規化を行った後のデータに対して行うことが多いので、基準とするネガティ ブコントロール値も、各サンプルのPer Array Normalized intensityの値を用います(下図)。 また本解析では、6サンプルのデータがあるので、この中から最も小さい値をCutOffに使 用します。

	腫瘍サンプル	腫瘍サンブル	腫瘍サンブル	腫瘍サンブル	腫瘍サンプル	腫瘍サンプル	腫瘍サンプル	腫瘍サンプル	腫瘍サンプル	腫瘍サンブル	腫瘍サ
No.	Per Array No	Per Gene Nor	Quality flag	Ignored cells	Per Array No	Per Gene Nor	Quality flag	Ignored cells	Per Array No	Per Gene Nor	Quality
216	48.0613	-0.37204	L L L L L L L L	0	56.68035	-0.23356	ililingigilig	0	57.65342	-0.1545	G L L G
Mean value	\smile	-	-	-		-	-	-	\smile	-	
Standard deviation	-	-	-	-	-	-	-	-	-	-	

最後に、前ページで計算したネガティブコントロールの値を、CutOff値として設定を行います。

Filter Optionsの「実験データ」項目で、左のチェックボックスをクリックし、項目「Per Array Normalized intensity」でCutOff値を入力し、かつ複数データの検索条件の選択に おいて、「すべての実験において条件を満たす場合」、または「1つの実験でも条件を満 たす場合」などを選択します。

下図は、ノイズ除去の設定例です。

前ページの例では、Per Array Normalized intensityにおける6サンプルでの最小値が 「48」だったので、このように設定を行いました。



以上が検索条件の設定例です。ただし、「③ノイズレベルの信頼性の低いデータを除外 する」で、③-2 intensityを用いてCutOffを行う場合は、Summarizationによるネガティブ コントロール値の計算、および統計検定のやり直しを、変動遺伝子の抽出の前に行って ください。

最後にFilter Options検索画面右下の「OK」ボタンをクリックすると、データの抽出が実行され、メイン画面に抽出されたデータが表示されます。

7711W 5 D 👸	>7¥∓W ¥		**** /			E:		8		0	検索結果	l: 43 (4					
MicroArray Na	me CodeLink(TM) Huma	n Whole G	Experiment	Name			Test [Control						
		[1		腫瘍サンブル	腫瘍サンブル	腫瘍サンブル	腫瘍サンブル	腫瘍サンブル	腫瘍サンブル	腫瘍サンブル	腫瘍サンブル	腫瘍サンブル	腫瘍サンブル	腫瘍サンブル	腫瘍サンブル	正常サンプ
No.	Select	Block	Row	Column	Per Array No	Per Gene Nor	Quality flag	Ignored cells	Per Array No	Per Gene Nor	Quality flag	Ignored cells	Per Array No	Per Gene No	Quality flag	Ignored cells	Per Array N
843	0		1 1	7 110	128.57327	0.49504	G	0	124.53997	0.44906	G	(125.41695	0.45919	G	0	39.0290
388	0		1 3	2 9	139.58617	0.32457	G	0	234.14433	1.07081	G	(172.33038	0.62859	G	0	81.4962
517	0	3	2 5	2 1	148.72007	0.3998	G	0	164.32782	0.54377	G	(184.08897	0.7076	G	0	67.9583
617	0		2 6	2 5	119.09652	0.25884	G	0	184.32188	0.88894	G	(152.76202	0.618	G	0	56.1874
7164	0		2 6	7 8	132.40803	0.32163	G	0	179.72337	0.76242	G	(158.43288	0.58052	G	0	79.487
2136	0		3 11	3 21	74.90452	0.3506	L	0	102.24778	0.79955	G	(72.562	0.30477	L	0	44.92
2409	0		3 11	5 74	81.44463	0.48161	L	0	71.90395	0.30187	G	(99.28627	0.76739	G	0	27.929
2592	0		3 11	7 38	65.68058	0.50678	L	0	78.932	0.77193	G	(59.93608	0.37474	L	0	25.003
4044	0		4 13	1 11	113.9543	0.80006	G	0	86.77975	0.40704	G	(86.0731	0.39524	G	0	44.820
6692	0	1	1 15	5 46	61.73762	0.41221	L	0	74.64865	0.68617	G	(96.95542	1.06338	G	0	21.410
6723	0	1	1 15	5 77	98.97498	0.47877	L	0	103.35262	0.54121	G	(94.05145	0.40516	G	0	47.995
7967	0	1	1 16	7 9	143.91552	0.40267	G	0	180.21417	0.72716	G	(129.2032	0.24709	G	0	57.2
17968	0		1 16	7 10	138.50893	0.6193	G	0	112.97778	0.32536	G	(150.355	0.73769	G	0	58.636
19802	0		5 17	5 12	120.49913	0.93545	G	0	123.9393	0.97606	G	0	80.34667	0.35074	G	0	45.666
0450	0		5 18	1 21	112.54332	0.71331	G	0	143.3388	1.06226	G	(89.2333	0.37848	G	0	24.777
20000	0		5 18	9 7	79.86813	0.42252	L	0	173.54995	1.54218	G	(120.54987	1.01646	G	0	36.623
24416	0	-	5 22	7 19	147.334	0.86542	G	0	97.52383	0.27016	G	(141.94395	0.81165	G	0	64.215
24410	0	1	5 22	8 84	89.01737	0.45834	L	0	103.3014	0.67304	G	(80.60592	0.31514	G	0	48.971
04799	0		5 23	0 3	59.75398	0.66202	L	0	69.24678	0.87473	G	(45.71645	0.2757	L	0	28,480
24732	0		5 24	1 6	81.3992	0.39326	G	0	97,94042	0.66015	G	(95,2101	0.61936	G	0	31,250
2923	0		5 24	4 56	79.69163	0.50137	L	0	80.81488	0.52157	G		85.63712	0.60518	G	0	32,902
0290	0		7 25	7 94	74 20445	0.82395	1	° n	5214155	0.31497	0	r	92 57455	114306	6	0	1740
:1120	ő		7 26	8 71	63,31,387	0.59304		0	52 56365	0.32458	6		61 20697	054421	1	0	31 383
20093	ő		3 30	2 54	51 1113	0.50661		0	61 10112	0.76416	6		7312642	1.02336		0	11.422
62538			2 20	0 40	67 06075	0.00001	L 1	0	7016700	0.50410	6		04.4175	0.00717		0	37.922
13181			, 30	0 49	100.04005	0.32040	L	U	110 20500	0.03933	G		100.06540	0.80717	6	0	01.620
3583		-	0 01	2 8	102.04030	0.3933	G	U	100.44057	0.02401	G		120.90042	0.03870	6	0	41.414
34352	0		s 31	9 8	128.13925	0.35379	G		182,44957	0.86357	G		120.92205	0.32861	G		14.622
35616	0		33	U 77	62.9164	0.30403	L	0	92.84683	1.11516	G		80.39012	0.90733	G	0	32.8
5898	0		3 33	3 33	53.36292	0.48114	L	0	85.86387	1.16735	G	(59.33462	0.63417	L	0	10.900
8498	0		9 35	7 69	62.83662	0.36931	L	0	62.95495	0.37203	G	6	70.1126	0.52738	L	0	27.081
0974	0	1	38	0 112	83.24098	0.44917	L	0	76.27437	0.32307	G	(103.2305	0.75967	G	0	45.667
11339	0	1	38	4 36	82.89187	0.24586	L	0	120.768	0.78879	G	(104.4418	0.57925	G	0	35.968

本解析では、遺伝子の抽出に以下のような条件を設定しました。

コントロール遺伝子などの除去:解析対象 Flagが「0」 変動遺伝子抽出:腫瘍グループ vs 正常グループで発現比が2倍以上 かつ、P-valueが0.01以下 ノイズデータ除去:少なくとも1つの実験データで、Quality flagが「G」

この結果、43個の遺伝子が抽出されました。結論として、この43遺伝子は腫瘍グループの細胞で、発現が増加している遺伝子だといえます。また、この後Gene Ontology解析や Pathway解析を行うことで、これらの遺伝子がどのような機能やパスウェイと関連がある かを調べることができます。

なお、腫瘍グループで発現が減少している遺伝子を抽出する場合は、その他の設定は変更せずに、Group Ratioの設定だけを「0.5以下」などにすることで行えます。