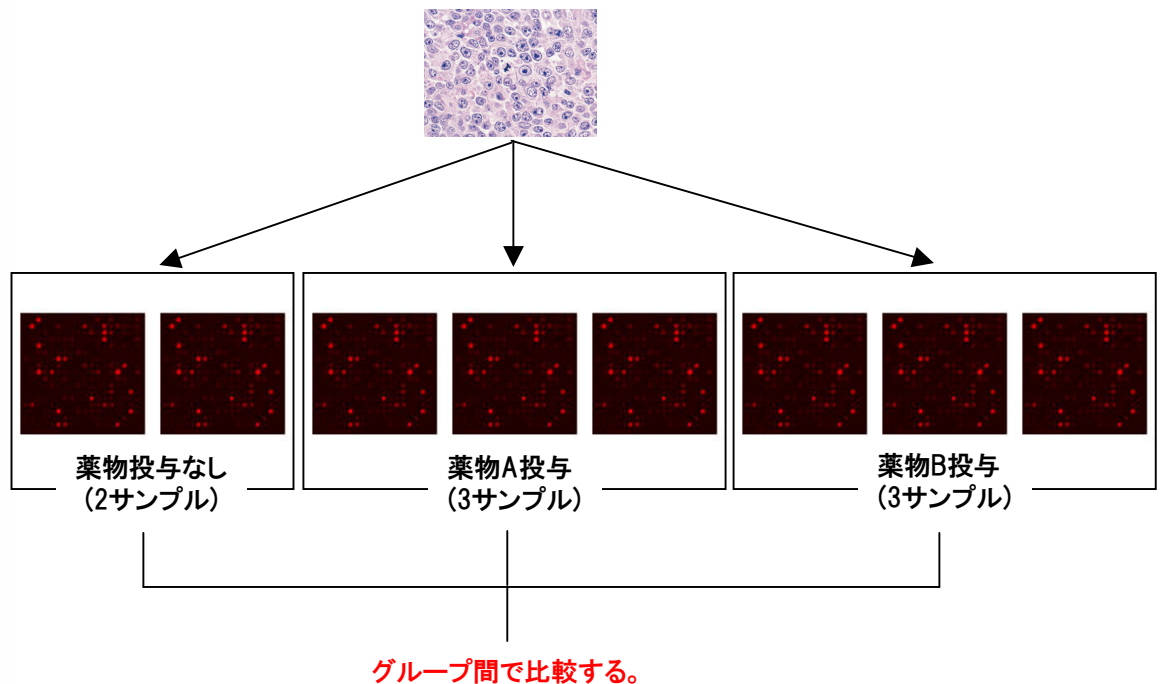


## Microarray Data Analysis Tool データ解析の進め方 Vol.6 – 3グループ以上の比較解析 –

Vol.5と同様に、繰り返し実験を含む多数のサンプルを、今度は3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

「薬物A投与-薬物B投与-薬物投与なしヒト細胞における、グループ間の遺伝子発現プロファイルの探索」です。



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
- ⑤クラスタリング解析で、発現量に有意差があるグループの同定

Vol.5と同様に、Signal比較解析を用いて、全実験データの正規化およびグループ分けを行い、統計検定によって、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。そして、グループが3つ以上ある場合は、様々な組み合わせのグループ間で有意差をもつデータが混在しているため、クラスタリング解析によって、グループの組み合わせごとにデータをまとめる処理が必要になります。

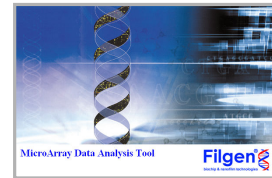
## 1. ソフトウェアの起動と実験データの読み込み

### 1-1. ソフトウェアの準備

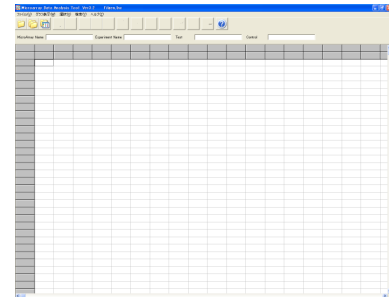
CDに入っている「Microarray Data Analysis Tool」のフォルダを、自分のコンピュータにコピーしてください。

### 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。

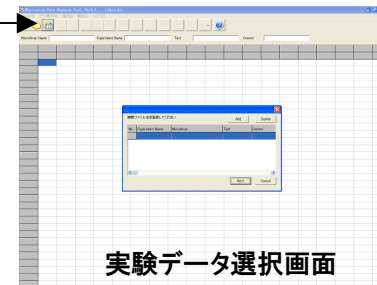


ソフトウェアが起動後は右図のメイン画面が表示されます。



### 1-3. 実験データ読み込み

 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を選択・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。



実験データ選択画面

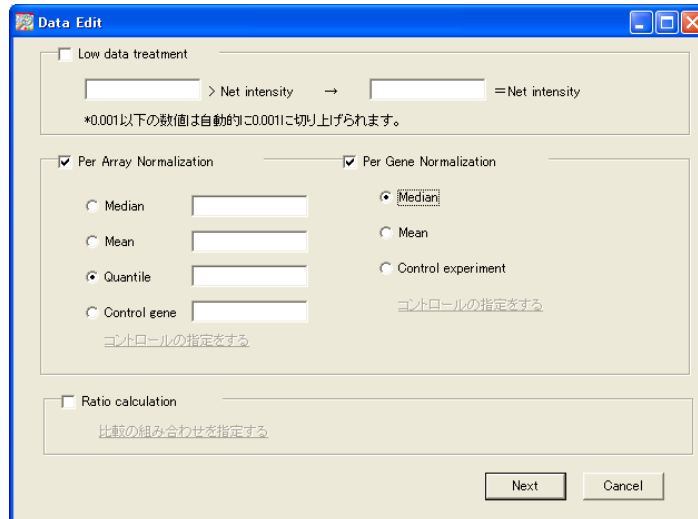
実験データを読み込むと、右図のように選択したファイルの使用マイクロアレイ名、サンプル名等が表示されます。ここでは「Normal-1～2」、「DrugA-1～3」、「DrugB-1～3」の合計8つの実験データを読み込みます。表示の確認が終わったら、「Next」ボタンをクリックして次へ進みます。



## 2. 実験データの正規化とグループ分け

### 2-1. 正規化手法の指定

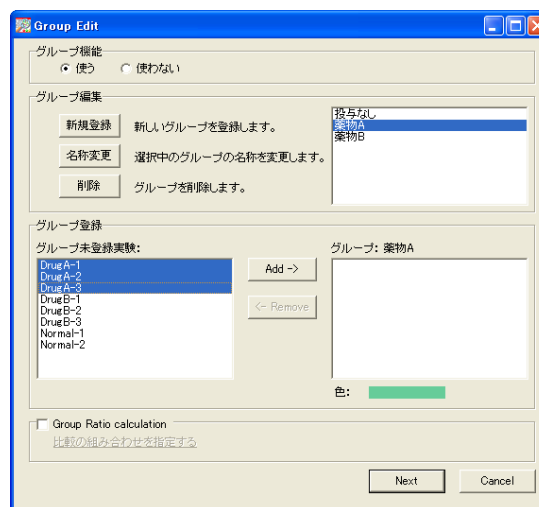
「Next」ボタンをクリックすると、「Data Edit」(下図)の画面が現れます。この画面では正規化の手法や、比較の組み合わせの選択を行います。各項目の詳細については、本ソフトウェアのユーザーマニュアル、またはVol.4を参照してください。



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。

### 2-2. グループ化

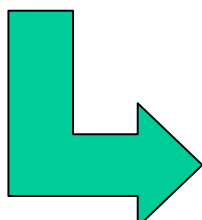
Data Edit画面から進むと、「Group Edit」画面が現れます。ここで実験データをグループ分けすることで、グループ単位の検索や統計検定が実行可能になります。



ここでは3つのグループを作成し、それぞれのサンプルの設定を行いました。

グループ分けの設定を行ったら、次に「Group Ratio calculation」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。

Group Ratio calculation:  
[比較の組み合わせを指定する](#)



クリック！

Select	テストグループ名		コントロールグループ名
	投与なし	VS	薬物A
	投与なし	VS	薬物B
<input type="radio"/>	薬物A	VS	投与なし
<input type="radio"/>	薬物A	VS	薬物B
<input type="radio"/>	薬物B	VS	投与なし
<input type="radio"/>	薬物B	VS	薬物A

All 全てにチェックを入れます。  
Delete 全てのチェックを外します。

OK Cancel

「Ratio calculation」と同様、「Group Ratio calculation」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、「薬物Aグループ vs 投与なしグループ」「薬物Bグループ vs 投与なしグループ」「薬物Bグループ vs 薬物Aグループ」の3つの組み合わせで、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、本解析では何も設定せず次に進みます。

Group Edit

Radio dataグループ機能  
 使う  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 選択中のグループの名称を変更します。  
削除 グループを削除します。

Ratioグループ登録  
グループ未登録Ratio data Add -> <- Remove Ratioグループ:  
色:

Group 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 選択中のグループの名称を変更します。  
削除 グループを削除します。

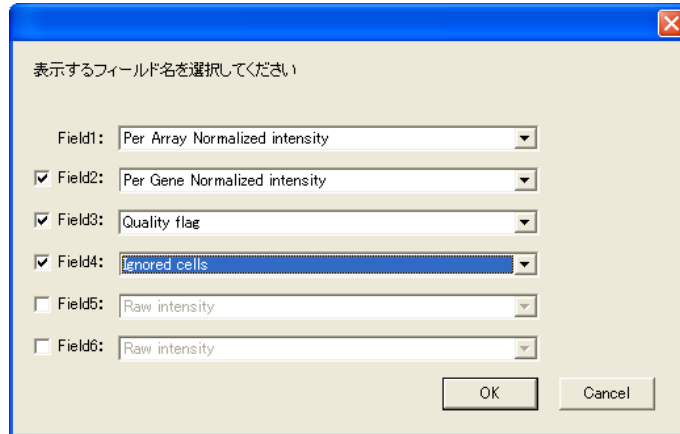
Group Ratioグループ登録  
グループ未登録Group Ratio data Add -> <- Remove Group Ratioグループ:  
薬物A vs 投与なし  
薬物B vs 投与なし  
薬物B vs 薬物A  
色:

Next Cancel

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



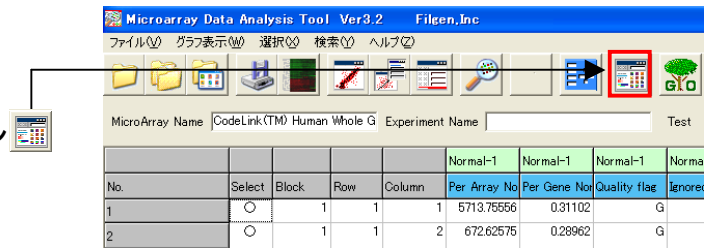
「OK」ボタンをクリックすると、選択したフィールドのデータが、サンプルごとにメイン画面に表示されます。またグループ設定を行っている場合は、実験データはグループごとに分類されて表示されます。

No.	Select	Block	Row	Column	Normal-1	Normal-1	Normal-1	Normal-1	Normal-2	Normal-2	Normal-2	Normal-2	DrugA-1	DrugA-1	DrugA-1	DrugA-1	DrugA-2
					Per Array No	Per Gene No	Quality Flag	Ignored cells	Per Array No	Per Gene No	Quality Flag	Ignored cells	Per Array No	Per Gene No	Quality Flag	Ignored cells	Per Array No
1	<input type="radio"/>	1	1	1	5713.76566	0.31102	G	0	4210.61766	-0.12939	G	0	4174.02589	-0.14198	G	0	3790.64857
2	<input type="radio"/>	1	1	2	672.62576	0.28962	G	0	4245.7347	-0.37417	G	0	489.63401	-0.18677	G	0	3603.10556
3	<input type="radio"/>	1	1	3	15.2968	0.40226	L	0	5.54599	-1.06076	L	0	16.69598	0.62517	L	0	6.63329
4	<input type="radio"/>	1	1	4	47.65778	1.3332	L	0	14.97378	-0.39607	L	0	45.61433	1.26988	L	0	19.49886
5	<input type="radio"/>	1	1	5	15.06966	0.10056	L	0	13.49159	-0.1433	L	0	18.83531	0.30992	L	0	13.73259
6	<input type="radio"/>	1	1	6	81.4231	-0.84128	L	0	25.29106	0.79212	L	0	5.8273	-1.32389	L	0	16.39359
7	<input type="radio"/>	1	1	7	13.62247	-0.35653	L	0	17.71341	0.02223	L	0	24.79561	0.50689	L	0	24.11455
8	<input type="radio"/>	1	1	8	5236.92876	0.17858	G	0	4698.05755	0.00845	G	0	4125.77121	-0.17919	G	0	3744.62746
9	<input type="radio"/>	1	1	9	79.02209	-0.27546	G	0	86.16308	-0.15065	G	0	105.13051	0.13639	G	0	7724635
10	<input type="radio"/>	1	1	10	460.44966	0.53312	G	0	325.1052	0.03098	G	0	343.43194	0.1101	G	0	29711149
11	<input type="radio"/>	1	1	11	271.4624	-0.18641	L	0	34.67745	0.16684	L	0	33.6798	0.12472	L	0	25.33151
12	<input type="radio"/>	1	1	12	92.76447	0.78869	G	0	63.17817	0.23459	G	0	38.98708	-0.46558	L	0	52.26985
13	<input type="radio"/>	1	1	13	44.5576	0.88933	G	0	6.17514	-1.96179	L	0	15.7871	-0.60759	L	0	30.29588
14	<input type="radio"/>	1	1	14	18.98934	-1.26901	L	0	42.05812	-0.12081	L	0	47.15946	0.04435	L	0	44.30423
15	<input type="radio"/>	1	1	15	4877.6637	0.08986	G	0	4029.38105	-0.19578	G	0	3678.10646	-0.31737	G	0	3431.7331
16	<input type="radio"/>	1	1	16	26.2973	0.6706	L	0	7.9514	-1.26822	L	0	16.23037	-0.12561	L	0	19.18339
17	<input type="radio"/>	1	1	17	22.21656	0.13043	L	0	15.67626	-0.37262	L	0	7.80394	-1.37883	L	0	40.42781
18	<input type="radio"/>	1	1	18	31.31004	0.72815	L	0	25.25459	0.41807	L	0	18.19743	-0.05474	L	0	29.10845
19	<input type="radio"/>	1	1	19	19.18884	0.32308	L	0	17.30359	0.17388	L	0	30.00612	0.96807	L	0	15.30271
20	<input type="radio"/>	1	1	20	9.23262	-0.9268	L	0	17.859	0.05504	L	0	25.33902	0.92975	L	0	17.24449
21	<input type="radio"/>	1	1	21	539.71351	0.45111	G	0	395.66096	-0.23408	G	0	415.83245	0.07492	G	0	339.03517
22	<input type="radio"/>	1	1	22	4913.35681	0.17727	G	0	3601.21429	-0.27095	G	0	3451.89304	-0.33305	G	0	3209.63926
23	<input type="radio"/>	1	1	23	23.81506	-0.53199	L	0	33.46474	-0.04122	L	0	49.63396	0.49628	L	0	17.37657
24	<input type="radio"/>	1	1	24	19.97192	0.79736	L	0	12.95356	0.17277	L	0	24.95655	1.11884	L	0	7.30724
25	<input type="radio"/>	1	1	25	22.26336	0.60247	L	0	11.10405	-0.40092	L	0	11.207	-0.3876	L	0	26.89872
26	<input type="radio"/>	1	1	26	11.53801	-0.05927	L	0	12.50651	0.05694	L	0	5.39956	-1.1574	L	0	13.34685
27	<input type="radio"/>	1	1	27	1078.77184	0.27314	G	0	778.08935	-0.20754	G	0	883.85659	-0.01497	G	0	843.23769
28	<input type="radio"/>	1	1	28	18.41688	-0.02876	L	0	25.27251	0.42778	L	0	0.001	-14.1975	L	0	19.25215
29	<input type="radio"/>	1	1	29	5747.08448	0.23638	G	0	4608.67595	-0.0821	G	0	4568.27966	-0.006	G	0	4086.04155
30	<input type="radio"/>	1	1	30	911.96578	-0.18903	G	0	1322.41716	0.34709	G	0	1031.89296	-0.01079	G	0	988.23463
31	<input type="radio"/>	1	1	31	18.57532	0.13388	L	0	18.41586	0.12144	L	0	20.71027	0.29084	L	0	7.22089
32	<input type="radio"/>	1	1	32	29.59654	0.83021	L	0	16.67624	0.00839	L	0	32.28915	0.96173	L	0	0.001

### 3. 統計検定の実行

#### 3-1. 検定手法の選択

メイン画面の統計検定アイコンをクリックしてください。



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 & Hochberg)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3グループ以上の比較解析になるので、使用する検定手法として「One-Way ANOVA」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)。

No.	薬物A vs 投	薬物A vs 投	薬物B vs 投	薬物B vs 投	薬物B vs 薬	薬物B vs 薬	One-Way ANOVA	One-Way ANOVA
	ratio	log2ratio	ratio	log2ratio	ratio	log2ratio	P-value	FDR
1	0.84599	-0.24129	0.96789	-0.04708	1.1441	0.1942	0.327935740747098	
2	0.90124	-0.15002	1.0308	0.04376	1.14376	0.1937	0.687245021539172	
3	0.94811	-0.07687	1.63525	0.70951	1.72474	0.7863	0.549534083188592	
4	0.89679	-0.15716	0.65196	-0.61715	0.72699	-0.4599	0.594436457322577	
5	0.78438	-0.35038	1.20459	0.26854	1.53573	0.6189	0.602342750907332	
6	0.49282	-1.02086	0.90155	-0.14951	1.82937	0.8713	0.319390576299947	
7	1.29071	0.36816	0.99464	-0.00776	0.77061	-0.3759	0.71924642487852	
8	0.85572	-0.2248	0.92102	-0.11869	1.07632	0.106	0.330879181271094	
9	1.28091	0.35717	1.38811	0.47312	1.08369	0.1159	0.407011990167122	
10	0.83238	-0.26469	0.79108	-0.35646	0.93838	-0.0917	0.202829814556516	

## 4. 発現変動遺伝子の抽出

### 4-1. Filter Options検索画面の表示

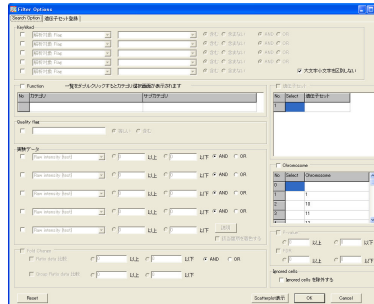
メイン画面のFilter Optionsアイコンをクリックし、Filter Options検索画面を開きます。



No.	Select	Block	Row	Column	Sample A
1	<input type="radio"/>	1	1	1	14858.9777
2	<input type="radio"/>	1	1	2	122.9206
3	<input type="radio"/>	1	1	3	38.1178
4	<input type="radio"/>	1	1	4	32.2077
5	<input type="radio"/>	1	1	5	882.8602
6	<input type="radio"/>	1	1	6	148.1519

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できます。

keyword  
機能カテゴリー  
実験データ  
染色体番号  
遺伝子セット



### 4-2. 検索条件を設定

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

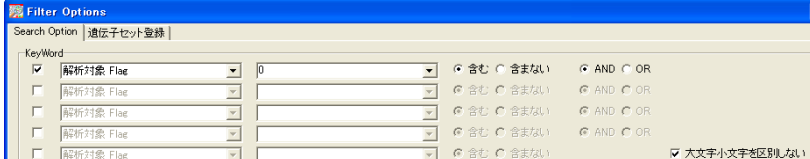
##### ①コントロール遺伝子などを除去し、解析対象遺伝子のみにする。

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できました。



## ②グループ間で発現差のあるデータを抽出する。

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。

### ②-1 Fold Changeで検索する。

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。

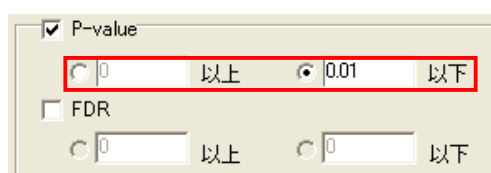
左のチェックボックスをクリックし、検索条件を入力します。(下図は「Group Ratio data 比較」を2以上にした例です。)

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、3つのGroup Ratioを計算しているため、複数データの検索条件の指定を求められます。



### ②-2 P-value(FDR)で検索する。

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、メイン画面上的「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ため、ここでは「P-value」が「0.01」以下と設定します。



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8つ)のうち1つでも解析除外スポットがある遺伝子は、検索時にデータが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。

統計検定の検定手法に、One-Way ANOVAを選択した場合は、「比較するグループのうち、どれか2つのグループ間の発現量に差がある」遺伝子でP-valueが小さくなります。したがって、P-valueで検索を行うだけでは、抽出された遺伝子がどのグループ間で発現量に差があるか分かりません。そこでGroup Ratioで検索を行い、グループ間のFold Changeを指標にして、発現量に差があるグループを決定する方法もありますが、検索条件の設定を何度も変更して検索を実行しないといけません。そのため本解析では、ここではP-valueのみで検索を行い、その後に抽出されたデータに対してクラスタリング解析を実行し、発現量に差があるグループの組み合わせごとに遺伝子を分類します。



### ③ノイズレベルの信頼性の低いデータを除外する。(CutOff)

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。

#### ③-1 Quality flagを用いてCutOffを行う。

##### 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

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  等しい  含む  
[1つの実験でも条件を満たす場合](#)

##### 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。

Quality flag  
 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  
 G  等しい  含む  
[1つのグループ\(内すべての実験\)でも条件を満たす場合](#)

**\*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。

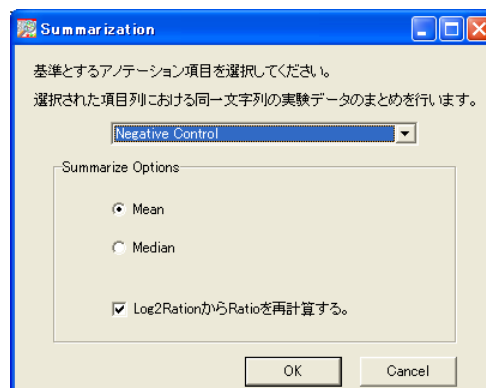
8サンプル全ての実験データで、Quality flagが「G」のデータのみを解析に使用すれば、データの信頼性は上がりますが、場合によっては条件が厳しすぎて、抽出される遺伝子が少なくなってしまうことがあります。例1～3を参考に、必要に応じて抽出条件の変更を行ってください。

### ③-2 intensityを用いてCutOffを行う。


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異なるため、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。

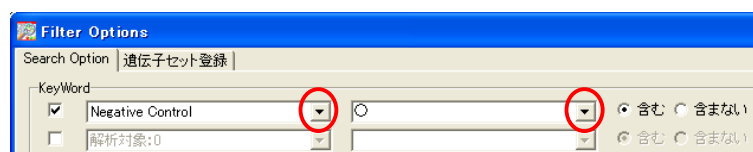
まずSummarizationを行い、データ内のネガティブコントロールスポット蛍光強度の平均値を求めます。

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、P-value(or FDR)を計算してください。



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アイコンをクリックし、Key Word検索で左側の入力欄には「Negative Control」、右側の入力欄には「○」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。



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。

	Normal-1	Normal-1	Normal-1	Normal-1	Normal-2	Normal-2	Normal-2	Normal-2	Drug A-1	Drug A-1	Drug A-
No.	Per Array No	Per Gene No	Quality flag	Ignored cells	Per Array No	Per Gene No	Quality flag	Ignored cells	Per Array No	Per Gene No	Quality
216	11.19368	-1.33358	L L L L L L L L	0	11.73284	-1.31418	L L L L L L L L	0	11.92741	-1.4785	L L L L L
Mean value											
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「11」だったので、このように設定を行いました。

実験データ

Per Array Normalized intensity  11 以上  0 以下  AND  OR  
[すべての実験において条件を満たす場合](#)

Per Array Normalized intensity  0 以上  0 以下  AND  OR

Per Array Normalized intensity  0 以上  0 以下  AND  OR

Per Array Normalized intensity  0 以上  0 以下  AND  OR

Per Array Normalized intensity  0 以上  0 以下

該当箇所を着色する

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。

No.	Select	Block	Row	Column	Normal-1 Per Array No	Normal-1 Per Gene No	Normal-1 Quality Flag	Normal-1 Ignored cells	Normal-2 Per Array No	Normal-2 Per Gene No	Normal-2 Quality Flag	Normal-2 Ignored cells	Drug-A-1 Per Array No	Drug-A-1 Per Gene No	Drug-A-1 Quality Flag	Drug-A-1 Ignored cells	Drug-A-2 Per Array No	Drug-A-2 Per Gene No
127	<input checked="" type="radio"/>	1	2	20	2176628419	0.38226	G	0	2166344824	0.37343	G	0	1622410441	-0.0437	G	0	149702798	
140	<input type="radio"/>	1	2	34	1061824754	0.40215	G	0	1271240027	0.66184	G	0	849649621	0.08089	G	0	879219013	
160	<input type="radio"/>	1	2	44	18363309	-0.15219	G	0	1491218	-0.45263	G	0	17825319	-0.19509	G	0	20290806	
174	<input type="radio"/>	1	2	70	38200066	-0.28512	G	0	32026163	-0.39706	G	0	38689606	-0.0681	G	0	3626412	
209	<input type="radio"/>	1	2	108	31419161	0.70432	G	0	26245761	0.44476	G	0	18200854	-0.08332	G	0	16859665	
573	<input type="radio"/>	1	6	34	8457086	-0.47442	G	0	3060319	-0.54374	G	0	12075019	0.03938	G	0	11535269	
1164	<input type="radio"/>	1	11	81	29450522	0.0453	G	0	29948766	0.0695	G	0	34403165	0.20955	G	0	27629745	
1186	<input type="radio"/>	1	11	103	119975191	0.74016	G	0	98786146	0.45981	G	0	77331474	0.10656	G	0	69413546	
1294	<input type="radio"/>	1	12	106	10117492	0.76276	G	0	86216832	0.51512	G	0	66474017	-0.00376	G	0	61952507	
1343	<input type="radio"/>	1	13	45	15406694	-0.49646	G	0	13878893	-0.64744	G	0	19714344	-0.14077	G	0	18712903	
1876	<input type="radio"/>	1	18	31	36237704	0.17921	G	0	40235149	0.33017	G	0	3576793	0.16119	G	0	33919631	
2004	<input type="radio"/>	1	19	48	47258635	0.38278	G	0	47821783	0.39987	G	0	32653951	-0.15054	G	0	31994614	
2216	<input type="radio"/>	1	21	44	2520872	0.26889	G	0	23238878	0.15461	G	0	20441733	-0.0382	G	0	19727031	
2691	<input type="radio"/>	1	25	84	33316339	-0.06609	G	0	36133927	0.05103	G	0	31996669	-0.12427	G	0	32597677	
2833	<input type="radio"/>	1	27	8	17919333	-0.57573	G	0	16037416	-0.7368	G	0	24102533	-0.14806	G	0	21689718	
3821	<input type="radio"/>	1	33	38	102963859	0.88496	G	0	84288846	0.59623	G	0	5547121	-0.00737	G	0	5603934	
3840	<input type="radio"/>	1	36	26	18578004	-0.39499	G	0	16563806	-0.56067	G	0	23684921	-0.04462	G	0	19523743	
3916	<input type="radio"/>	1	36	106	2767153	-0.45108	G	0	27171199	-0.4774	G	0	32128625	-0.23662	G	0	3589176	
4249	<input type="radio"/>	1	39	110	29050536	-0.36964	G	0	2843016	-0.40078	G	0	33735369	-0.15394	G	0	34725236	
5033	<input type="radio"/>	2	47	70	33939414	0.62204	G	0	28438015	0.36994	G	0	19592096	-0.17065	G	0	19958804	
5062	<input type="radio"/>	2	47	99	6627324	0.03296	G	0	58039274	-0.02163	G	0	64478925	0.13027	G	0	64137992	
5214	<input type="radio"/>	2	49	27	7477075	-0.05499	G	0	3059084	0.05297	G	0	1023806	0.0394	G	0	9921326	
5360	<input type="radio"/>	2	50	64	6589776	-0.73622	G	0	3334741	-0.39731	G	0	10492848	-0.06511	G	0	12428787	
5545	<input type="radio"/>	2	52	29	3772301	-0.68071	G	0	31854911	-0.92438	G	0	45040097	-0.42514	G	0	46215359	
5825	<input type="radio"/>	2	54	96	16490884	0.26547	G	0	15076071	0.13606	G	0	12286504	-0.15912	G	0	1220718	
5860	<input type="radio"/>	2	55	10	6372985	-0.31348	G	0	6063857	-0.38545	G	0	7589859	-0.06138	G	0	8249609	
5841	<input type="radio"/>	2	55	106	30892023	-0.30289	G	0	25634573	-0.57203	G	0	37123079	-0.03781	G	0	38416955	
6096	<input type="radio"/>	2	57	38	6380103	-0.44863	G	0	6549155	-0.4109	G	0	8458706	-0.04178	G	0	8038059	
6106	<input type="radio"/>	2	57	48	101973362	-0.04582	G	0	965927845	-0.12402	G	0	104023962	-0.0171	G	0	996389619	
6154	<input type="radio"/>	2	57	98	361333903	-0.34857	G	0	365992324	-0.32712	G	0	502796069	0.13108	G	0	469885755	
6451	<input type="radio"/>	2	60	60	23014669	0.07276	G	0	23097222	0.07793	G	0	2401924	0.1344	G	0	22476184	
6908	<input type="radio"/>	2	62	88	17498347	-0.37535	G	0	15413784	-0.55835	G	0	21893306	-0.06208	G	0	19635804	

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: P-valueが0.01以下
- ノイズデータ除去: すべての実験データで、Quality flagが「G」


この結果、499個の遺伝子が抽出されました。結論として、この499遺伝子は「薬物投与なしグループ」「薬物A投与グループ」「薬物B投与グループ」のどれか2グループ間で、発現量に差があ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。

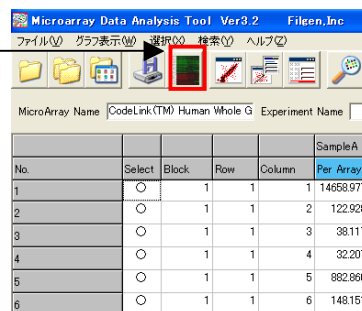
次に抽出した遺伝子を、発現量に差があるグループの組み合わせごとに分類するため、クラスタリング解析を行います。

## 5. 抽出した遺伝子のクラスタリング解析

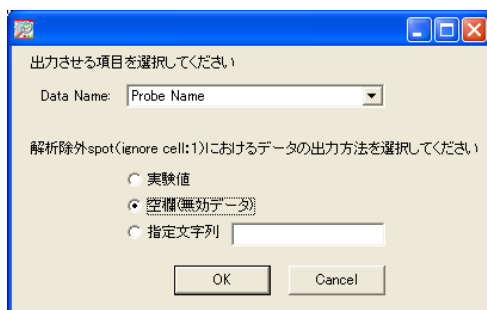
絞り込んだ遺伝子について、発現量に差があるグループの組み合わせごとに分類するため、クラスタリング解析を行います。

- ① Filter Optionsでデータの絞り込みを行います。  
ここでは4. 発現変動遺伝子の抽出のときに指定した条件をそのまま使用します。  
なお、Signal比較解析でクラスタリング解析を行う場合は、必ずPer Gene Normalizationを行ってください。

- ② クラスタリング用データ出力アイコンをクリックし、出力設定ウインドウを開きます。  
本解析では、「Data Name」に「Probe Name」を、「解析除外spotにおけるデータの出力方法」として、「空欄(無効データ)」を選択しました。  
その後「OK」ボタンをクリックして、ファイルを出力します。



No.	Select	Block	Row	Column	Per Array N
1	<input type="radio"/>	1	1	1	14658.9777
2	<input type="radio"/>	1	1	2	122.9286
3	<input type="radio"/>	1	1	3	38.1178
4	<input type="radio"/>	1	1	4	32.2077
5	<input type="radio"/>	1	1	5	882.8602
6	<input type="radio"/>	1	1	6	148.1519



出力させる項目を選択してください

Data Name:

解析除外spot(ignore cell:1)におけるデータの出力方法を選択してください

実験値

空欄(無効データ)

指定文字列

OK Cancel

解析除外spotにおけるデータの出力方法

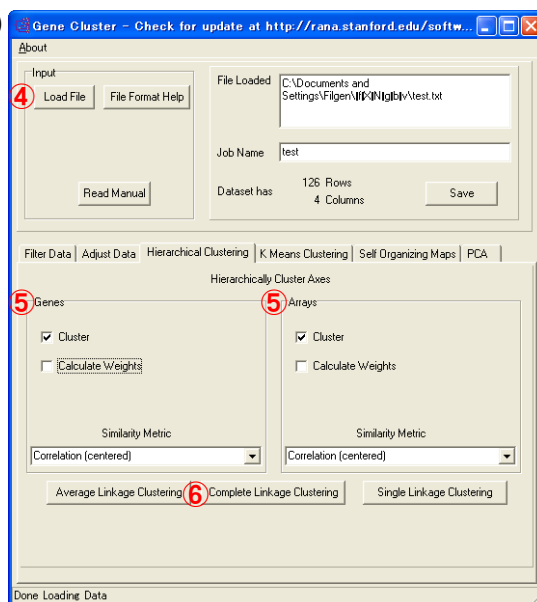
- ・実験値: 実験データをそのまま出力します。
- ・空欄(無効データ): データなしとして出力します。
- ・指定文字列: 入力した文字列(または数値)として、出力されます。

- ③ Eisen Lab (<http://rana.lbl.gov/EisenSoftware.htm>)より、「Cluster」と「TreeView」をダウンロードし、インストールを行います。

- ④ インストールしたClusterを起動し、ウインドウ左上の「Load File」ボタンをクリックし、②で出力したファイルを選択します。

- ⑤ 「Hierarchical Clustering」タブをクリックし、解析の設定を行います。  
本解析では、遺伝子とサンプルの両方でクラスタリングを行うので、「Genes」と「Arrays」の両方の「Cluster」で、チェックボックスにチェックをいれます。  
次にデータの類似性の尺度である「similarity metric」から、「Correlation (centered)」を「Genes」と「Arrays」の両方で選択します。

- ⑥ 最後にウインドウ下側で、3つのクラスタリング手法の中から、「Complete Linkage Clustering」ボタンをクリックし、クラスタリングを実行します。計算が終了すると、2~3個の計算結果ファイルが出力されます。



Gene Cluster - Check for update at <http://rana.stanford.edu/softw...>

Input

File Loaded: C:\Documents and Settings\Filgen\My Recent\test.txt

Job Name: test

Dataset has: 126 Rows, 4 Columns

Filter Data | Adjust Data | Hierarchical Clustering | K Means Clustering | Self Organizing Maps | PCA

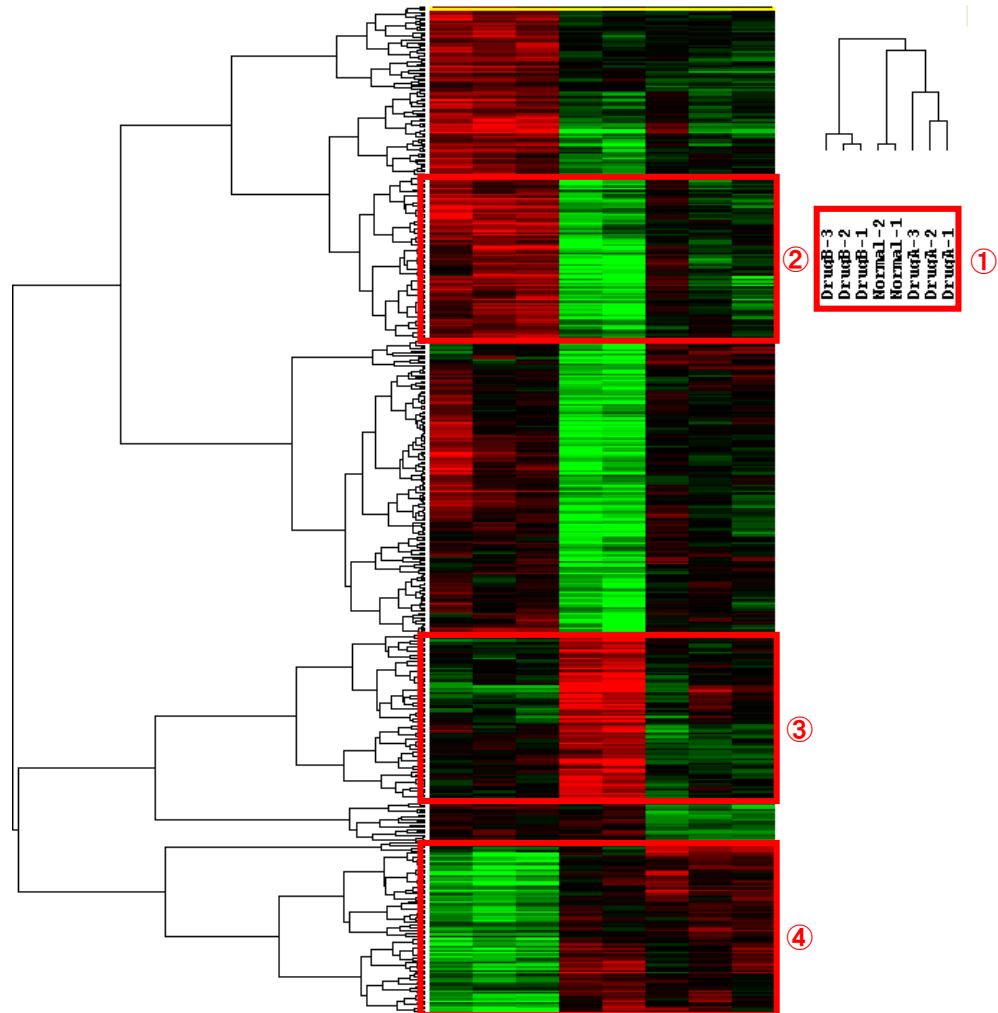
Hierarchically Cluster Axes

Genes:  Cluster,  Calculate Weights, Similarity Metric: Correlation (centered)

Arrays:  Cluster,  Calculate Weights, Similarity Metric: Correlation (centered)

Average Linkage Clustering |  Complete Linkage Clustering | Single Linkage Clustering

- ⑦ TreeViewを起動し、ウインドウ上部のメニューバーから、「File」→「Load」と選択し、  
 ⑥で出力された計算結果ファイルのうち、拡張子が「.cdt」となっているものを  
 読み込みます。
- ⑧ クラスタリング結果が図で表示されるので、オプション設定などで自分の見やすいように、  
 図を加工します。下図は、本解析結果の表示例です。



図の列は各サンプルを、行は各遺伝子を表しており、さらに緑色の領域は、その遺伝子が各サンプルにおいて発現量が高いことを、赤色の領域は発現量が低いことを示しています。

まず①のサンプルの並び方より、薬物処理の種類に基づいて、各サンプルの発現プロファイルが近いものであることが分かります。

次に、サンプルごとの発現量の違いに基づき、特徴的なクラスターを抽出しました。例えば、②のクラスターは薬物Bを投与したサンプルで発現量が低く、投与なしのサンプルでは発現量が高い遺伝子を多く含み、③は投与なしのサンプルで発現量が低い遺伝子を、④では薬物Bを投与したサンプルで発現量が高く、投与なし及び薬物A投与のサンプルで発現量が低い遺伝子を多く含んでいることが分かります。

なお、各クラスター領域をマウスで選択しますと、図の右側に該当遺伝子のProbe Nameが表示されるので、そこから各遺伝子の詳細情報を検索してください。

このように、クラスタリング解析を行うことによって、P-valueで抽出した、いずれかのグループで発現量に差がある遺伝子を、差のあるグループごとに分類することができます。

なお、本クラスタリング解析で用いた計算手法などの詳細は、「Cluster and TreeView」のマニュアルをご参照ください。