

Microarray Data Analysis Tool データ解析の進め方 Vol.4

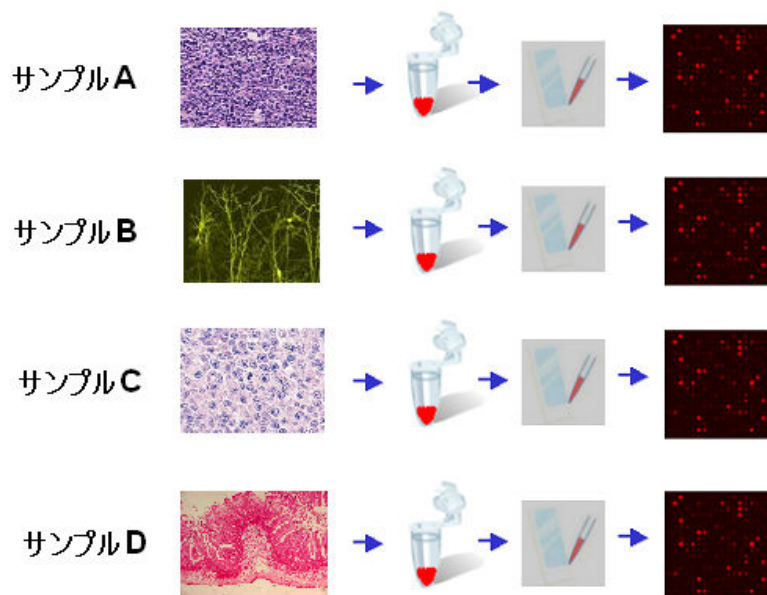
— Signal比較解析 —

複数比較解析に引き続き、Signal比較解析についてのソフトウェアの使用方法を、実際の解析例と合わせて説明します。

ここでは

「4種類のヒト細胞株における、遺伝子発現プロファイルの探索」をテーマに、

ヒト細胞を用いた単色実験の解析例を紹介します。まず、4種類のヒト細胞について、マイクロアレイの単色実験を行い、それぞれの遺伝子発現データを得ます。このデータを用いて、本ソフトウェアでデータの正規化や比較計算を行うことで、4種類のサンプルに対して任意の組み合わせで、データの比較・検索を行うことができます。



上記のテーマについて解析を行うには、本ソフトウェア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任意の組み合わせでの、サンプルの比較データの計算
- ③指定条件でのデータ検索による、遺伝子の抽出
- ④抽出した遺伝子群を、クラスタリングで分類

Signal比較解析では、様々な条件で得たマイクロアレイ実験の生データを読み込むため、それらのデータを同じ条件で比較できるように、まず最初に正規化を行います。そして正規化を行ったデータに対して、2サンプルごとの発現比(Ratio)の計算を行います。この発現比計算では、自分で比較したい組み合わせを自由に指定することができます。そして、これらの計算したデータに基づいて遺伝子の検索を行うことで、特定のサンプルで変動している遺伝子の抽出を行います。最後に、抽出した遺伝子のデータを用いて、クラスタリング解析を行い、どのような発現プロファイルをもつ遺伝子が抽出されたかを調べます。

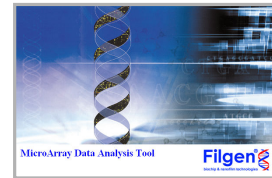
1. ソフトウェアの起動と実験データの読み込み

1-1. ソフトウェアの準備

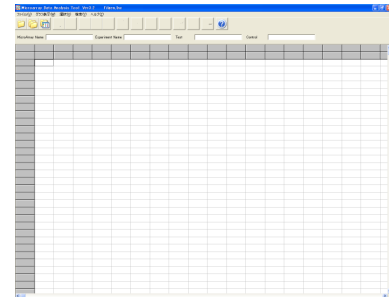
CDに入っている「Microarray Data Analysis Tool」のフォルダを、自分のコンピュータにコピーしてください。

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。

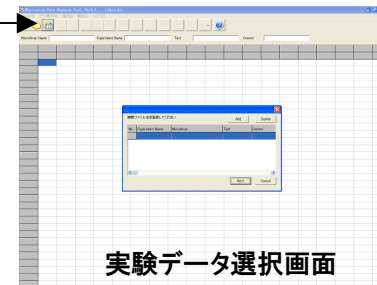


ソフトウェアが起動後は右図のメイン画面が表示されます。



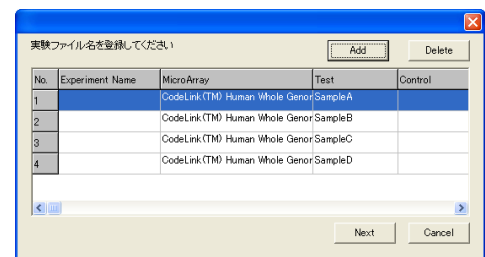
1-3. 実験データ読み込み

 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

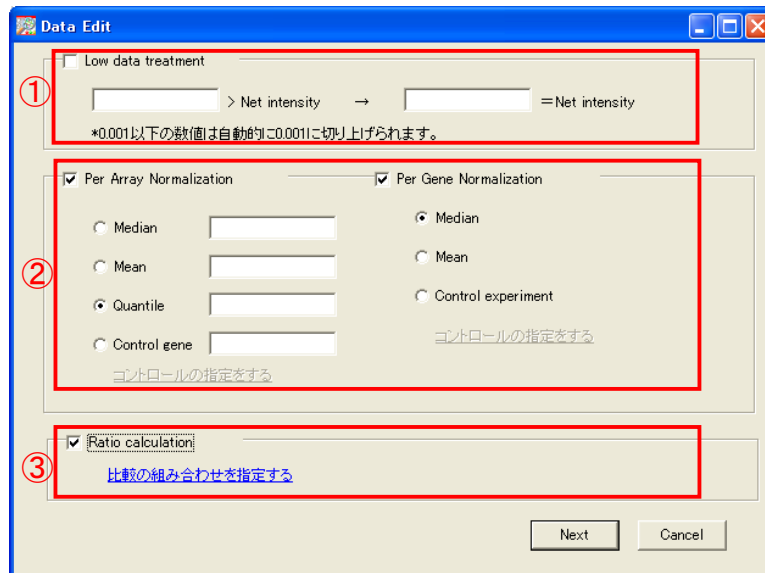
実験データを読み込むと、右図のように選択したファイルの使用マイクロアレイ名、サンプル名等が表示されます。ここでは、「SampleA～D」の4つの実験データを読み込みます。表示の確認が終わったら、「Next」ボタンをクリックして次へ進みます。



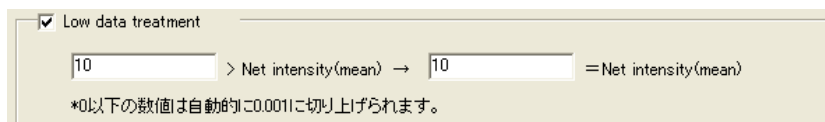
2. 実験データの正規化と比較データの作成

2-1. 正規化と比較の組み合わせの指定

「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。



①の項目では、弱いシグナル強度のスポットに対して、一定の値に数値を引き上げる処理の設定を行います。例えば、



のように設定しますと、正規化前のシグナル強度が「10」未満のスポットは、全て「10」に変換されます。ただし、何も設定を行わなかった場合でも、シグナル強度が「0.001」以下のスポットは、自動的に「0.001」に変換されます。本解析では、この項目は設定せずに、操作を進めていきます。

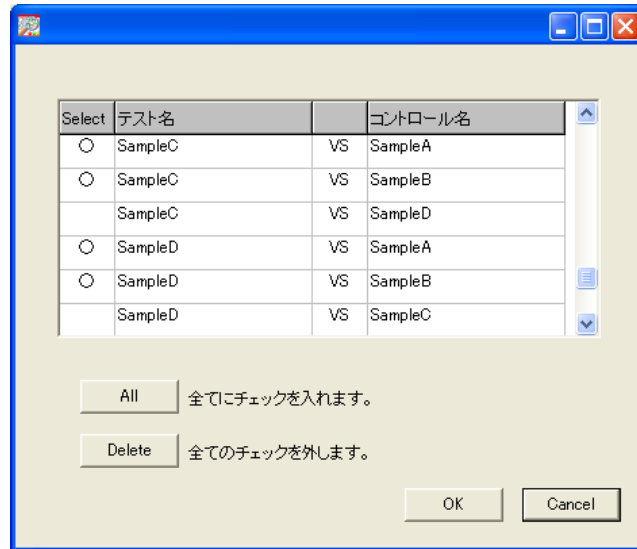
次に②の項目で正規化の設定を行います。正規化には「Per Array Normalization」と「Per Gene Normalization」の2種類がありますが、本解析では両方の設定を行います。

Per Array Normalizationでは選択できる手法が4種類ありますが、弊社マイクロアレイ受託解析サービスでは、多くの場合「Median」、または「Quantile」で補正を行っています。Medianは読み込んだ全マイクロアレイデータの中央値を、一定の値に揃えることで正規化を行います。単色のマイクロアレイデータの補正に、最もよく使用されているのですが、実験日や実験環境の違いによって、マイクロアレイ間に偏りができる場合があります。Quantileはそのような偏りを補正することができますが、発現プロファイルが極端に違うデータ同士に対して使用すると、それらの有意な違いも補正してしまうことがあります。

Per Gene Normalizationでは、Per Array Normalization後の発現量に対して、各遺伝子ごとに他の実験データと比較しやすいように標準化を行います。クラスタリングや発現量の統計検定を行うためには、この処理が必要となります。

本解析ではPer Array NormalizationにQuantileを選択し、Per Gene NormalizationにMedianを選択して、その後の解析を行いました。

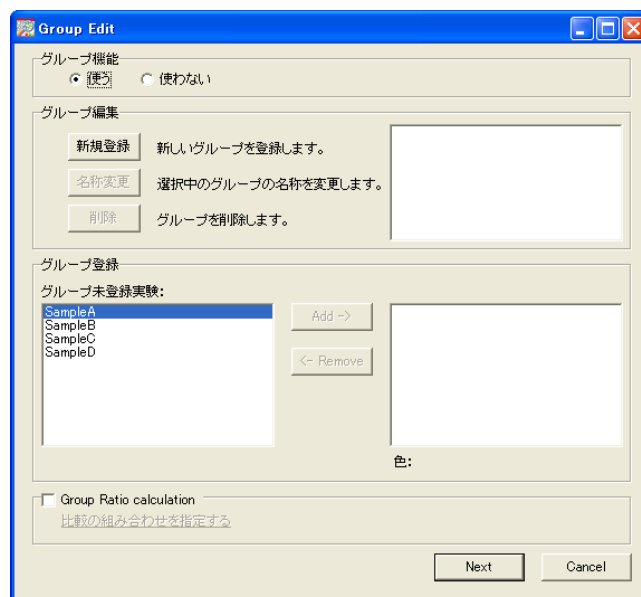
最後に③の項目で、比較するサンプルの組み合わせを指定します。チェックボックスにチェックを入れた後、「比較の組み合わせを指定する」という青色の文章をクリックすると下のような画面が表示されます。



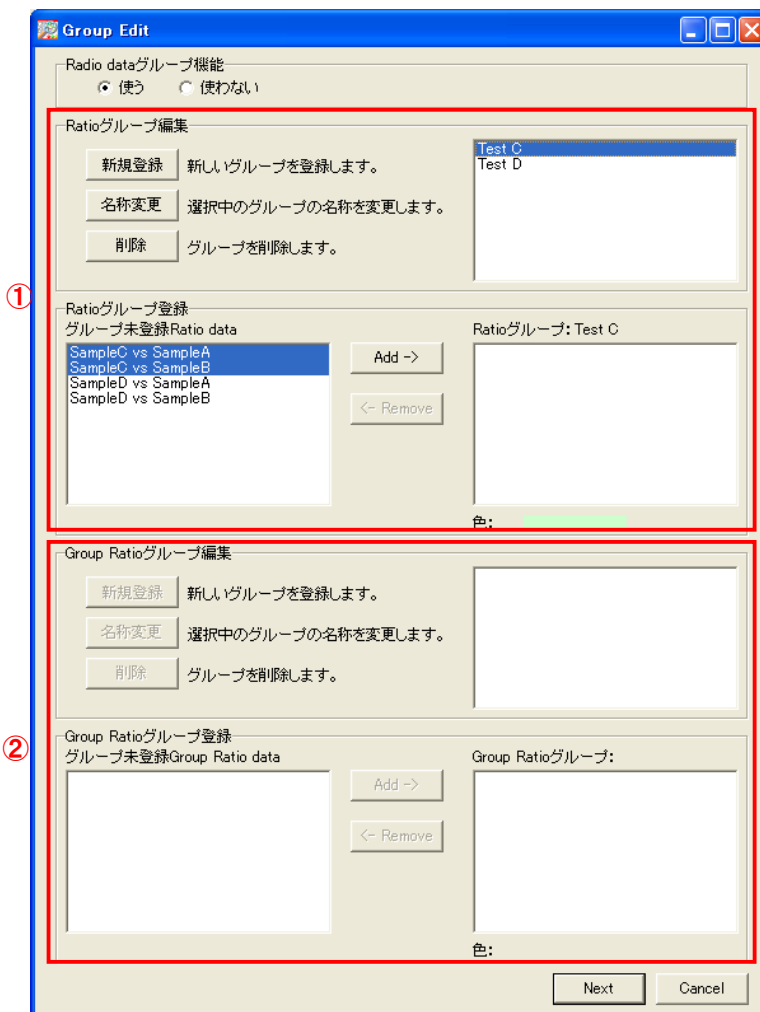
この画面では読み込んだ全てのサンプルデータの組み合わせ(テスト、コントロールの入れ替え含む)が表示されます。各組み合わせの「Select」列をダブルクリックすると、「○」が入力されるので、このようにして比較したい組み合わせを選択してください。なお、本解析では「SampleA」、「SampleB」をコントロールとし、「SampleC」、「SampleD」それぞれに対する発現比を計算したいので、選択する組み合わせは上図のようになります。選択終了後、「OK」ボタンをクリックするとData Editの画面に戻るので、「Next」ボタンをクリックして先に進んでください。

2-2. 様々なデータのグループ化

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が可能になります。ただし本解析では、4種類の異なる細胞株をサンプルに使用していますので、いずれも同一グループにまとめることができません。そのため、本解析では何も設定せず、「Next」ボタンをクリックし次に進みます。



Group Editから進むと、2つ目のGroup Edit画面が現れます。前の画面では、実験データのグループ化を行ったのに対し、この画面ではRatioデータ(Data Editで指定した比較の組み合わせや、前のGroup Editで計算したGroup Ratio)のグループ化を行います。



①の領域では、Data Editで指定した組み合わせのRatioデータのグループ化を行います。「グループ未登録Ratio data」の欄に、Data Editで指定した発現比計算を行う組み合わせが表示されるので、必要に合わせてグループ分けを行ってください。本解析では、計算した4つの組み合わせを、テストサンプルを基準にして2つのグループに分けます。すなわち

TestC グループ — SampleC vs SampleA
 SampleC vs SampleB

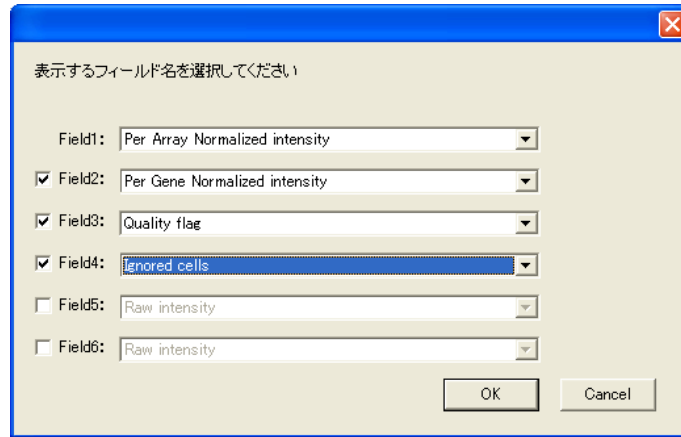
TestD グループ — SampleD vs SampleA
 SampleD vs SampleB

のように、SampleC基準のグループとSampleD基準のグループに分けます。

②の領域では、1つ目のGroup Editで指定した、**実験データグループ間発現比 (Group Ratio) のグループ分け**を行います。ただし本解析では、実験データのグループ分けを行っておらず、グループ間発現比の計算を指定していませんので、「グループ未登録 Group Ratio data」の欄は空白になっています。よって、今回は①のみ設定を行い、「Next」ボタンをクリックして次に進みます。

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。最大6項目まで選択でき、ここで選択された項目のみがメイン画面に表示されます。選択できる項目は以下のとおりです。

- | | |
|---------------------------------|------------------------------------|
| Raw intensity: | スポットの蛍光強度 |
| Background: | スポット周囲の蛍光強度 |
| Net intensity: | Raw intensity – Background |
| Per Array Normalized intensity: | Per Array Normalization後のintensity |
| Per Gene Normalized intensity: | Per Gene Normalization後のintensity |
| Quality flag: | 各スポットにつけられたフラグ |
| Ignored cells: | 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、メイン画面に選択したフィールドのデータが、サンプルごとに表示されます。またRatioデータは、指定した組み合わせのデータが、右側にまとめて表示されます。

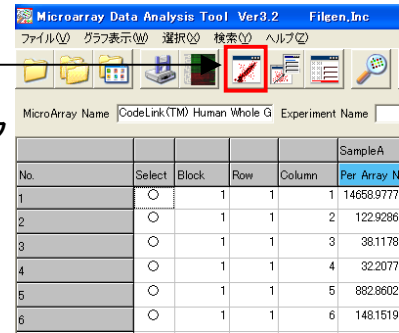
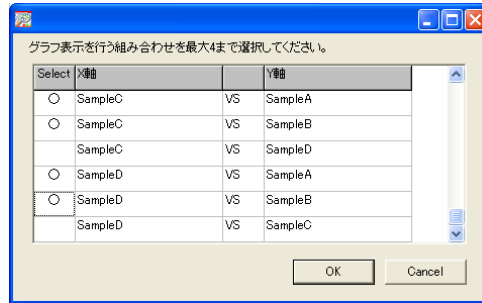
実験データ
Ratioデータ

No.	SampleC				SampleD				SampleC vs		SampleC vs		SampleC vs		SampleD vs		SampleD vs		SampleD vs	
	Per Array No	Per Gene No	Quality flag	Ignored cells	Per Array No	Per Gene No	Quality flag	Ignored cells	ratio	betaRatio	ratio	betaRatio	ratio	betaRatio	ratio	betaRatio	ratio	betaRatio		
1	898145606	-0.05025	G	0	892212025	-0.05518	G	0	0.89191	-0.10598	0.92757	-0.10848	0.89541	-0.11768	0.92122	-0.11138				
2	59490787	-0.18322	G	0	72752997	0.10702	G	0	0.99618	-0.15492	0.86394	-0.21117	1.09841	0.13541	1.05641	0.07917				
3	1933655	0.34899	L	0	1105598	-0.46113	L	0	1938355	14.24295	0.91713	-0.1248	11055975	13.42263	0.82311	-0.93482				
4	0.001	0	L	0	0.001	0	L	0	4E-5	-14.83763	1	0	4E-5	-14.83763	1	0				
5	813315	-1.76592	L	0	46.95	0.77332	L	0	0.20966	-2.25387	0.90372	-0.99932	1.21031	0.27537	2.90779	1.53992				
6	34.78375	-0.12826	L	0	33.83338	-0.1674	L	0	0.84321	-0.24604	0.79937	-0.32307	0.82064	-0.28518	0.77798	-0.3622				
7	25.80105	0.09347	L	0	45.04545	0.89407	L	0	1.14347	0.19341	2.0994	1.06998	1.99172	0.99402	3.65679	1.87058				
8	7863.6004	-0.12107	G	0	8614.42785	0.01049	G	0	0.92626	-0.11051	0.78321	-0.26522	1.0147	0.02105	0.85799	-0.22097				
9	246.63405	-0.88008	G	0	730.19723	1.17984	G	0	0.70329	-0.49759	0.83207	-0.26522	2.097	1.06833	2.46347	1.30069				
10	398.48735	-0.1846	G	0	499.89002	0.17914	G	0	0.90587	-0.14263	0.95536	-0.22539	1.16564	0.22112	1.10065	0.12836				
11	46.30083	0.26672	L	0	58.16225	0.95677	L	0	46300.825	15.49975	1.50964	0.5942	581.6225	15.8278	1.89638	0.92225				
12	60.07125	0.19425	L	0	44.9364	-0.22454	L	0	0.9453	-0.08116	1.20827	3.99487	0.70713	-0.49995	9.03848	3.17608				
13	67.78932	0.36371	L	0	38.06717	-0.4738	L	0	0.98986	-0.01646	1.76949	0.82333	0.95518	-0.84987	0.93566	-0.00910				
14	18.6506	0.07765	L	0	7.2595	-1.28363	L	0	1.11707	0.15972	0.42056	-1.24962	0.4348	-1.20156	0.1637	-2.6109				
15	7255.2241	-0.15463	G	0	8066.4021	-0.00172	G	0	0.89729	-0.15535	0.78206	-0.35466	0.97761	-0.00345	0.86949	-0.20175				
16	211.48795	-0.31171	G	0	722.34262	1.4504	G	0	0.96872	-0.04584	0.69962	-0.53612	3.3387	1.72627	2.35642	1.23594				
17	24.22803	0.02217	L	0	8.28459	-1.82093	L	0	1.02145	0.04468	0.64108	-0.64142	0.92639	-1.50361	0.21921	-2.18962				
18	16.76882	-0.82436	L	0	28.30003	-0.04717	L	0	0.54098	-0.88635	0.40349	-1.11269	0.91889	-0.12675	0.78301	-0.3629				
19	23.95645	0.01137	L	0	23.61880	-0.01146	L	0	2.89565	14.55047	0.59043	-0.74074	2.861875	14.52765	0.5904	-0.76356				
20	23.35697	-0.02845	L	0	27.72258	0.21881	L	0	0.96169	-0.05535	1.24568	0.31692	1.14148	0.90091	1.47856	0.56419				
21	471.96362	-0.06977	G	0	538.92768	0.12267	G	0	0.96144	-0.05673	0.9575	-0.06265	1.09174	0.12663	1.08727	0.12071				
22	7148.85475	-0.07033	G	0	7745.49135	0.04494	G	0	0.96354	-0.02394	0.82346	-0.28023	1.05535	0.91133	0.89195	-0.16496				
23	140.72373	0.14142	G	0	437.99465	1.77947	G	0	1.89787	0.91661	1.22964	0.29223	5.87528	2.55466	3.82718	1.59628				
24	23.23005	0.38197	L	0	11.34185	-0.66237	L	0	0.63745	-0.6496	1.86992	0.90298	0.31123	-1.68394	0.91297	-0.13136				
25	76.95395	1.92103	L	0	7.62453	-1.41372	L	0	6.16719	2.62461	2.73235	1.45014	0.61126	-0.71014	0.27082	-1.88461				
26	15.3966	-0.36772	L	0	12.6239	-0.65349	L	0	0.44885	-1.15581	0.63266	-0.66065	0.36781	-1.44258	0.51897	-0.94627				
27	5410.65417	-0.12385	G	0	1170.65313	-2.34134	G	0	0.88328	-0.25449	0.70302	-0.57709	0.18137	-2.46238	0.14503	-2.78558				
28	5.0655	-1.47036	L	0	12.3265	-0.18481	L	0	0.32216	-1.63416	0.16471	-2.60303	0.78534	-0.34861	0.40151	-1.31648				
29	8833.65025	-0.09392	G	0	8233.70367	-0.16234	G	0	0.88142	-0.1821	0.83108	-0.26684	0.84059	-0.25082	0.79259	-0.33536				
30	3288.52245	0.18414	G	0	1245.90003	-1.18054	G	0	1.31519	0.39527	0.80136	-0.31947	0.51071	-0.09941	0.11118	-1.68416				
31	18.25807	0.34488	L	0	14.009793	-0.03817	L	0	2.92895	1.38986	1.24596	0.31725	1.9434	0.95691	0.96206	-0.0558				
32	40.31703	0.29784	L	0	25.27623	-0.37677	L	0	1.70038	0.76596	0.8121	-0.30026	1.06903	0.09225	0.50914	-0.97987				

3. スキャタープロットの作成

3-1. 表示する組み合わせの選択

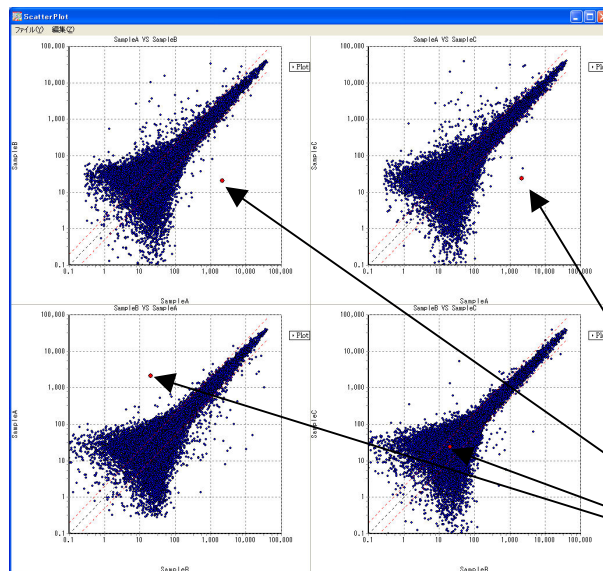
メイン画面のScatterPlot表示アイコンをクリックしてください。



すると上図のように、サンプルデータの組み合わせ一覧が表示されます。ここでグラフを作成したい組み合わせを、最大4つまで指定してください。「OK」ボタンをクリックすると、グラフが表示されます。なお、ここでプロットされるデータは、メイン画面上の表示の有無に関わらず、Per Array Normalized intensityの値が用いられます。

Scatter Plotの各プロットをクリックすると、選択されたプロットが赤いプロットになります。そして、メイン画面上の対応する遺伝子(行)が青色で選択され、行の先頭にスクロール表示されます。

また、1つのグラフのプロットをクリックすると、他のグラフの対応するプロットも赤くなります。さらに、メイン画面のPlot windowアイコンをクリックすると、Scatter Plot中の各プロット個別のデータを表示するPlot window画面が現れます。



Parameter	Value
Experiment name	SampleA vs Sampl
No.	15651
Block	4
Row	145
Column	94
Raw intensity SampleA	4921.7114
Raw intensity SampleB	1324.5784
Background SampleA	140
Background SampleB	1177.5784
Net intensity SampleA	4781.7114
解析対象 Flag	0
Probe Name	GE61199
Probe Type	DISCOVERY
GenBank Accession	NM_009598
Entrez Gene ID	4050
Gene Symbol	LTB
Gene Name	lymphotoxin beta (

Plot window画面

同スポット


上記までのステップで、お好みの組み合わせでグラフを作成し、サンプル間の遺伝子発現の全体像を見ることができました。


スキャタープロットが中心ライン付近に収束していれば、比較サンプル間の発現差は小さいと解釈が出来ます。ばらつきの幅が広ければサンプル間における発現差が大きいと解釈できます。

次は、変動遺伝子の抽出となります。

4. 発現変動遺伝子の抽出

4-1. Filter Options検索画面の表示

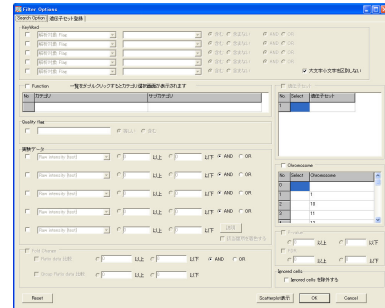
メイン画面のFilter Optionsアイコンをクリックし、Filter Options検索画面を開きます。



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スポット上にごみ等があり、シグナルに影響するデータを除外する。
- ③発現差のあるデータを抽出する。
- ④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

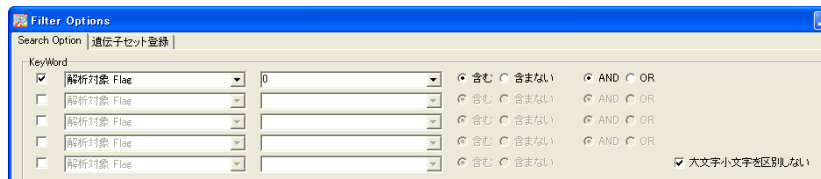
①コントロール遺伝子などを除去し、解析対象遺伝子のみにする。

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。

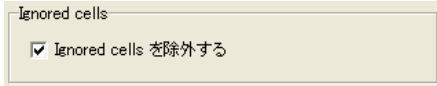


②スポット上にごみ等があり、シグナルに影響するデータを除外する。

ここでは、Ignore cellの設定を使用します。

事前にアレイに搭載されているデータの中でごみなどの影響を受け、正確なシグナルを算出できていない

可能性があるデータはメイン画面で表示されているIgnored cellsの項目が「1」になっています。上図、チェックボックスをクリックすることで、Ignored cellsを除外することが出来ます。

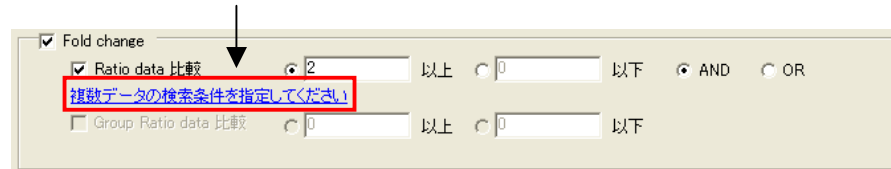


③発現差のあるデータを抽出する。

4ページで指定したサンプルごと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。本解析ではグループ間発現比の計算を行っていませんので、「Group 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「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」の項目をクリックします。

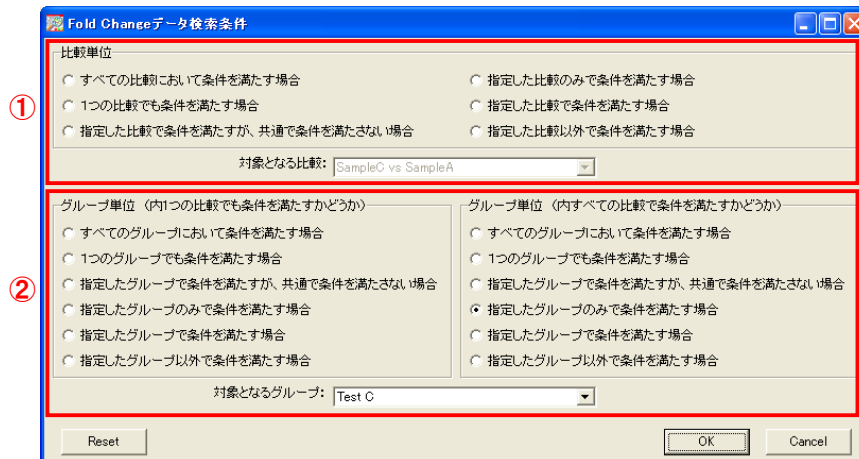


「Fold Changeデータ検索条件」画面が開き、指定した検索条件に対して、どのRatioデータを対象として実行するか選択することができます。

Ratioデータは、①:比較(1組の発現比)単位、②:グループ(5ページで指定した、発現比データを格納したグループ)単位で行うことができ、それぞれ下図の同一番号の領域で、条件を指定します。

例えば、①の領域で「指定した比較で条件を満たす場合」にチェックを入れ、「対象となる比較」に「SampleC vs SampleA」を選択した場合、SampleCのSampleAに対する発現比が2以上の遺伝子が検索されます。

また②の右側の領域で、「指定したグループのみで条件を満たす場合」にチェックを入れ、「対象となるグループ」に「Test C」を選択した場合、Test Cグループに格納されている全てのRatioデータ(本解析ではSampleC vs SampleAとSampleC vs SampleB)の発現比が2以上であり、さらにその他のグループに格納されているRatioデータの発現比が2未満の遺伝子が検索されます。本解析では、このグループ検索を用いた検索条件の設定(下図)を使用します。



④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。

④-1 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](#)


上記のようにCutOffを設定し、さらに③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。

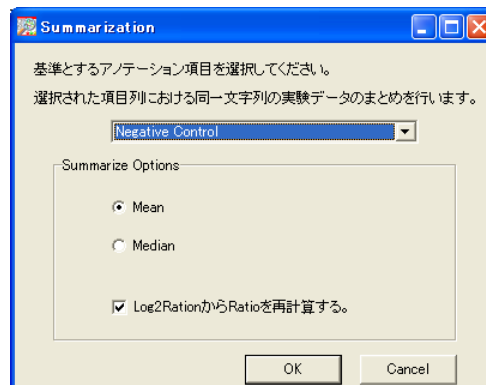
4サンプル全ての実験データで、Quality flagが「G」のデータのみを解析に使用すれば、データの信頼性は上がりますが、場合によっては条件が厳しすぎて、抽出される遺伝子が少なくなってしまうことがあります。例1と2を参考に、必要に応じて抽出条件の変更を行ってください。

④-2 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。

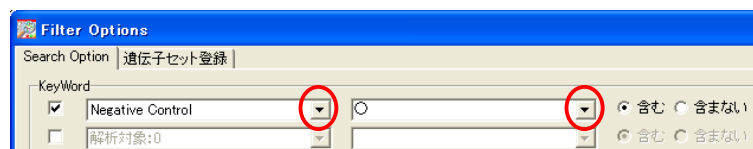
まずSummarizationを行い、データ内のネガティブコントロールスポット蛍光強度の平均値を求めます。

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。



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「0」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、4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。

	SampleA	SampleA	SampleB	SampleB	SampleC	SampleC	SampleD	SampleD			Flag	
No.	Per Array Normalized intensity	Ignored cells	Per Array Normalized intensity	Ignored cells	Per Array Normalized intensity	Ignored cells	Per Array Normalized intensity	Ignored cells	P-value	FDR	解析対象:0	Negative Control
90	57.64255	0	53.66399	0	61.42175	0	63.8643	0	1	1	2	○
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4サンプルでの最小値が「54」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、「④ノイズレベルの信頼性の低いデータを除外する」で、④-2 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。

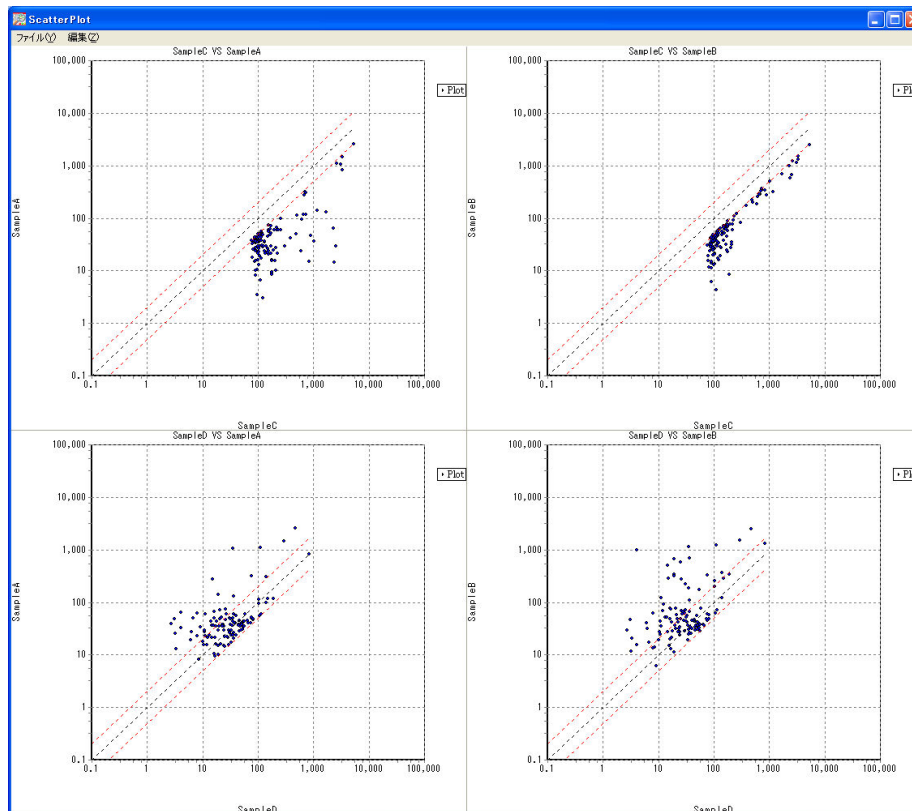
最後にFilter Options検索画面右下の「OK」ボタンをクリックすると、データの抽出が実行され、下図のようにメイン画面にデータが表示されます。なおノイズデータの除去には、④-1例1の方法を使用しています。

No.	Select	Block	Row	Column	SampleA	SampleA	SampleA	SampleA	SampleB	SampleB	SampleB	SampleB	SampleC	SampleC	SampleC	SampleC	SampleD	SampleD
					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.	Per Gene No.
1381	<input type="radio"/>	1	13	84	35.15558	0.32178	L	0	21.0907	-0.4179	L	0	103.70165	1.8231	G	0	17.0807	
1402	<input type="radio"/>	1	13	106	45.29285	-0.28944	L	0	38.03878	-0.54175	L	0	149.49475	1.4328	G	0	65.4563	
2277	<input type="radio"/>	1	21	105	34.92348	-0.02141	L	0	35.96782	0.02109	L	0	99.62965	1.49097	G	0	19.9582	
2369	<input type="radio"/>	1	22	80	10.66526	-0.76527	L	0	19.92287	0.13658	L	0	96.9538	2.41947	G	0	16.32345	
2377	<input type="radio"/>	1	22	98	9.2343	-0.81162	L	0	23.18135	0.51627	L	0	178.17745	3.46663	G	0	0.001	
3823	<input type="radio"/>	1	33	40	15.08228	-0.94631	L	0	33.67027	0.2125	L	0	86.53697	1.57434	G	0	24.44738	
3853	<input type="radio"/>	1	36	39	15.47808	-0.90395	L	0	28.10665	-0.04326	L	0	90.65738	1.64625	G	0	29.81795	
4119	<input type="radio"/>	1	38	88	47.38445	-0.45366	L	0	48.85177	-0.49398	L	0	106.69178	0.71731	G	0	80.92493	
4356	<input type="radio"/>	1	40	107	60.44512	-0.63303	L	0	72.32897	-0.37408	L	0	153.51083	0.71161	G	0	115.15	
5154	<input type="radio"/>	2	48	79	23.6465	-0.68428	L	0	52.3484	0.46224	L	0	156.56105	2.04275	G	0	13.05655	
5254	<input type="radio"/>	2	49	70	65.41	-0.12953	L	0	77.6892	0.11885	L	0	217.4358	1.63047	G	0	16.17738	
5511	<input type="radio"/>	2	51	107	56.46135	0.15699	L	0	44.67815	-0.17871	L	0	113.88462	1.71723	G	0	39.92247	
5833	<input type="radio"/>	2	54	105	30.29402	-0.51289	L	0	33.16698	-0.38217	L	0	117.06448	1.43731	G	0	63.28623	
6043	<input type="radio"/>	2	56	97	3.03835	-1.64927	L	0	16.01223	0.74949	L	0	123.15813	3.69276	G	0	0.001	
6242	<input type="radio"/>	2	58	74	23.62502	-0.29281	L	0	19.36998	-0.5793	L	0	131.47538	2.1836	G	0	34.25735	
6607	<input type="radio"/>	2	61	107	24.2304	0.56287	L	0	8.75205	-0.92669	L	0	188.9373	3.52589	G	0	0.001	
7119	<input type="radio"/>	2	66	74	22.7356	-0.23992	L	0	34.81688	0.27491	L	0	101.00252	1.81146	G	0	11.03825	
7653	<input type="radio"/>	2	71	71	15.67225	-0.45103	L	0	27.17625	0.3431	L	0	203.7661	3.2496	G	0	10.37698	
8180	<input type="radio"/>	2	76	51	6.71267	0.28704	L	0	4.29047	-0.38971	L	0	109.05038	4.30901	G	0	0.001	
8198	<input type="radio"/>	2	76	69	59.9516	-0.00926	L	0	60.72407	0.00921	L	0	214.37108	1.82898	G	0	26.99815	
8543	<input type="radio"/>	2	79	89	20.51185	-1.32824	L	0	66.7892	0.37427	L	0	156.07837	1.5996	G	0	36.25	
8756	<input type="radio"/>	2	81	92	12.97612	-0.36242	L	0	20.3876	0.28941	L	0	102.59318	2.62059	G	0	3.2869	
9620	<input type="radio"/>	3	90	9	49.87103	-0.34821	L	0	44.5692	-0.51036	L	0	101.14447	0.67194	G	0	77.09828	
9650	<input type="radio"/>	3	90	39	49.9989	0.22214	L	0	30.071	-0.4705	L	0	111.19345	1.37623	G	0	56.72043	
9805	<input type="radio"/>	3	91	87	27.2905	-0.37135	L	0	43.32085	0.29607	L	0	90.6379	1.36013	G	0	10.81928	
10210	<input type="radio"/>	3	95	56	18.90383	-0.66762	L	0	41.15118	0.45466	L	0	107.21912	1.83621	G	0	6.04805	
10389	<input type="radio"/>	3	97	19	22.26345	-0.72745	L	0	48.75923	0.40355	L	0	141.95817	1.94266	G	0	24.96455	
10945	<input type="radio"/>	3	102	35	3.52473	-2.44475	L	0	34.85468	0.86102	L	0	96.4324	2.32919	G	0	0.001	
11005	<input type="radio"/>	3	103	43	14.95507	-3.58384	L	0	339.11753	0.91944	G	0	831.9551	2.21416	G	0	19.4754	
11181	<input type="radio"/>	3	111	56	51.14437	-0.40399	L	0	58.72047	-0.2047	L	0	168.6805	1.31765	G	0	76.6244	
12116	<input type="radio"/>	3	113	1	28.9691	-0.36117	L	0	41.41545	0.15449	L	0	105.32362	1.50107	G	0	33.00415	
12135	<input type="radio"/>	3	113	61	33.57515	0.64812	L	0	61.32693	-1.80368	L	0	90.61355	2.08045	G	0	9.27417	

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。

変動遺伝子抽出: SampleC vs SampleAとSampleC vs SampleBでは発現比が2倍以上
かつ、SampleD vs SampleAとSampleD vs SampleBでは2倍以下
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、126個の遺伝子が抽出されました。
また、この126個のデータに対してScatter Plotを作成すると下図のようになります。




SampleAとSampleBに対して、SampleCでは全て発現比が2倍以上(上段2つ)、またSampleDでは全て2倍以下(下段2つ)になっていることを、視覚的に表現できました。

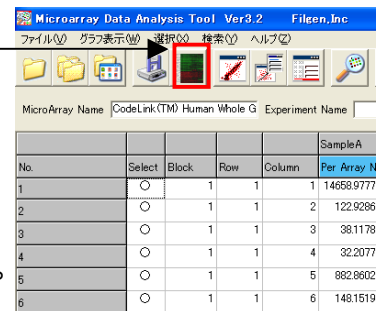
結論として、この126遺伝子はサンプルCの細胞で、特に高発現している遺伝子だといえます。また、この後Gene Ontology解析やPathway解析を行うことで、これらの遺伝子がどのような機能やパスウェイと関連があるかを調べることができます。以降の手順は、1比較解析と同じ流れになりますので、Vol.2の「4.遺伝子セットの登録方法」以降を参照してください。代わりに、本マニュアルではこれら126遺伝子を使用した、クラスタリング解析の手順を紹介します。

5. 抽出した遺伝子のクラスタリング解析

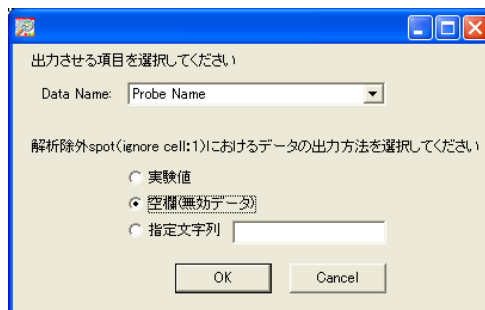
絞り込んだ遺伝子について、どのような発現プロファイルをもつものがあるのかを調べるために、クラスタリング解析を行います。

- 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

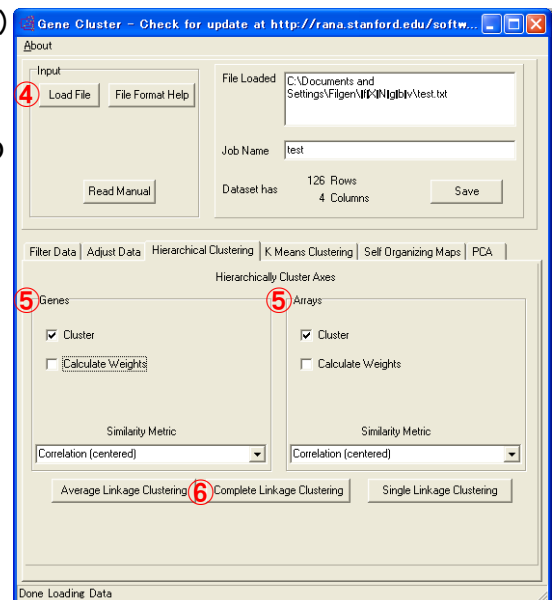
- ・実験値: 実験データをそのまま出力します。
- ・空欄(無効データ): データなしとして出力します。
- ・指定文字列: 入力した文字列(または数値)として、出力されます。

- 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

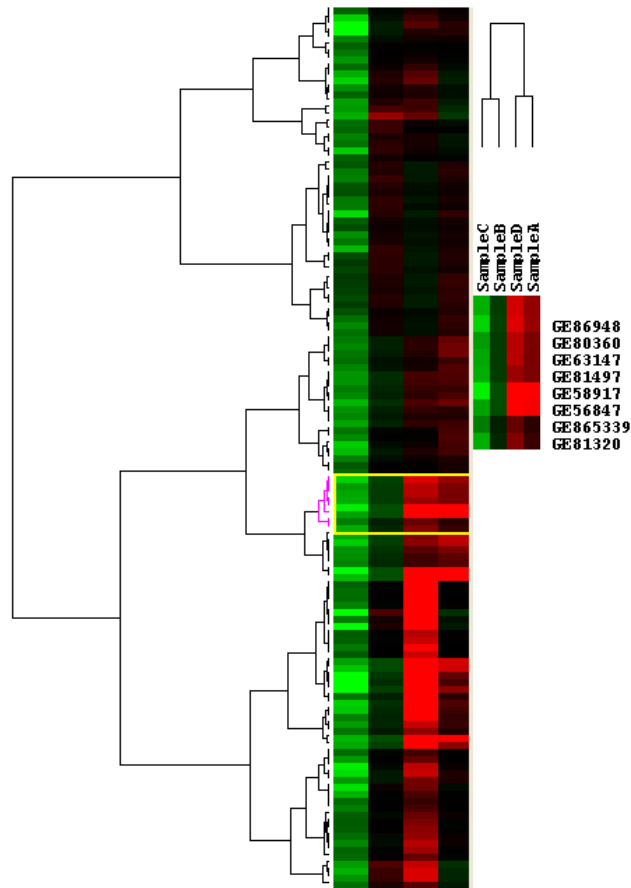
Hierarchically Cluster Axes

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。

クラスタリング解析結果の図から、4. 発現変動遺伝子の抽出で抽出された126遺伝子は、大きく2つのクラスターに分かれることが確認できました。

このように、クラスタリング解析を行うことによって、指定の条件で抽出された遺伝子群やサンプルを、発現プロファイルにしたがって分類を行うことができます。
 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。