

Microarray Data Analysis Tool データ解析の進め方 Vol.3

— 複数比較解析 —

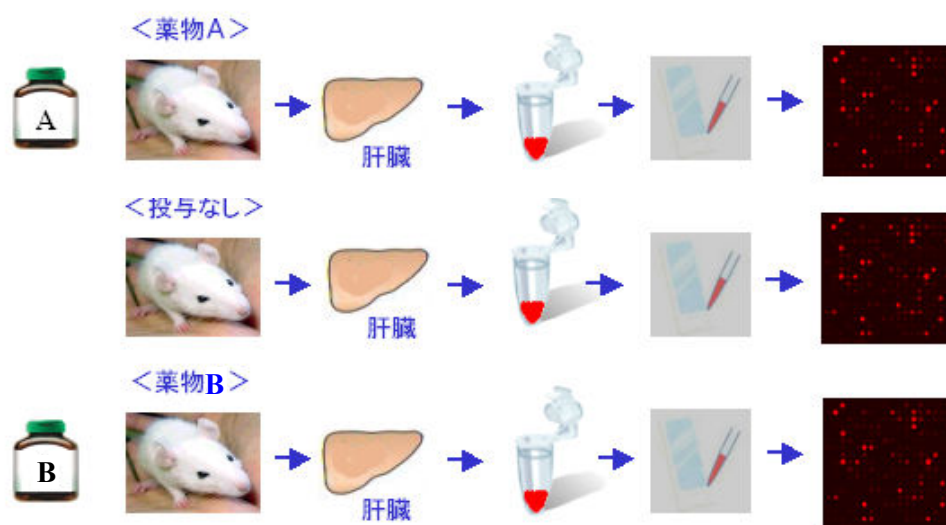
1比較解析に引き続き、複数比較解析についてのソフトウェアの使用方法を、実際の解析例と合わせて説明します。

ここでは

「薬物A投与マウスおよび薬物B投与マウスにおける発現変動遺伝子群の抽出と機能解析」をテーマに、

薬物Aを投与したマウスおよび薬物Bを投与したマウスをテストサンプル、投与していないマウスをコントロールサンプルとした実験例を紹介します。まず、それぞれの遺伝子発現プロファイルをマイクロアレイ実験で測定し、「薬物A vs コントロール」、「薬物B vs コントロール」の1比較実験データを作成します。その後、実験データの中から、コントロールサンプルに対して、薬物A投与マウスおよび薬物B投与マウスで高発現、または低発現している遺伝子を絞り込み、これらの遺伝子がどのような機能をもっているのか調べます。この結果、これらの2種の薬物が生体に対してどのような作用をするのか考察することができます。

<実験概要> サンプル調製 → 組織の抽出 → 蛍光ラベル → ハイブリ → スキャンング



複数比較解析では変動遺伝子群の扱い方が1比較解析と異なります。例として今回の複数比較解析では、変動遺伝子群を次の3つの要素に分けて解析を行います。

- i) 薬剤Aおよび薬剤Bの投与によって発現が共通に変動した遺伝子群
- ii) 薬剤A投与でのみ特異的に発現が変動した遺伝子群
- iii) 薬剤B投与でのみ特異的に発現が変動した遺伝子群

i)、ii)、iii)のどの変動遺伝子群に着目するかによって解析対象の薬剤が変わってきますので、この点が複数比較解析の主な特徴の一つです。

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

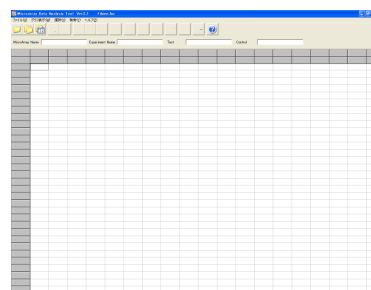
CDに入っている「Microarray Data Analysis Tool」のフォルダを、自分のコンピュータにコピーしてください。

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。

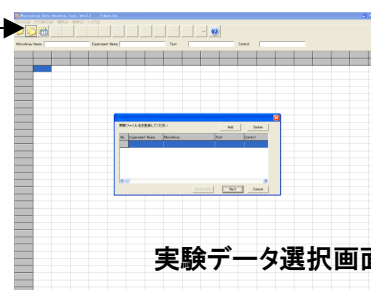


ソフトウェアが起動後は右図のメイン画面が表示されます。



1-3. 実験データ読み込み

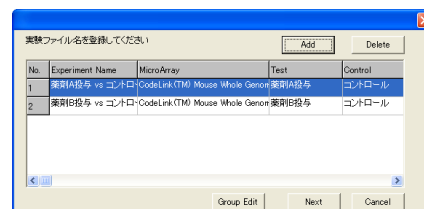
複数比較データ解析アイコンをクリックすると、下記の画面が表示されます。「Add」ボタンをクリックし、対象となる実験データの選択・読み込みを行います。



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、実験データ名・サンプル名等が表示されます。

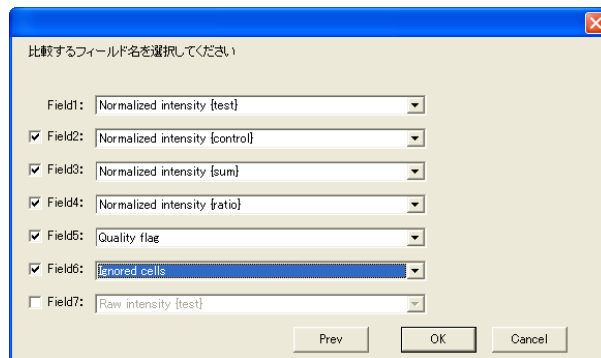
表示の確認が終わったら、「Next」ボタンをクリックして次へ進みます。



「Next」ボタンをクリックすると、比較するフィールド名を選択する画面が現れます。複数比較の場合、選択したフィールド名のみをメイン画面上に表示させます。本機能では最大7つのフィールドを選択・表示することができます。

ここでは出力するフィールドとして、以下の項目を選択します。

- Normalized intensity {test}
- Normalized intensity {control}
- Normalized intensity {sum}
- Normalized intensity {ratio}
- Quality flag
- Ignored cells



「OK」ボタンをクリックするとデータがメイン画面に表示されます。


薬物A vs コントロール 薬物B vs コントロール

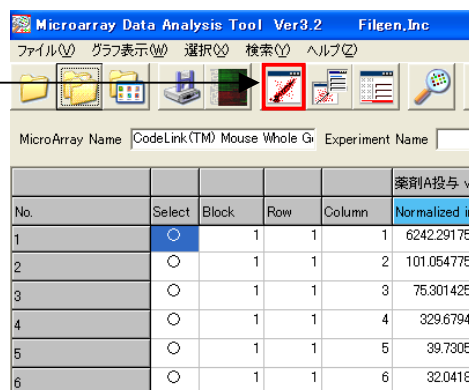
No.	Select	Block	Row	Column	Normalized	Normalized	Normalized	Normalized	Quality flag	Ignored cells	Normalized	Normalized	Normalized	Normalized	Quality flag	Ignored cells	Ignore	FOR
1		1	1	1	6242.2177	971.72189	1241.9383	170.8235	0.01		119.966	119.62567	292.72475	1.03595	0.0	0	1	
2		1	1	2	161.05476	119.02567	220.89045	0.842348	0.0		119.966	119.62567	292.72475	1.03595	0.0	0	1	
3		1	1	3	76.30425	37.6428	112.94478	2.00428	0.0		60.769428	37.6428	98.432076	1.61498	0.0	0	1	
4		1	1	4	329.6794	297.9794	627.6588	1.06393	0.0		300.192428	297.9794	598.171925	1.07427	0.0	0	1	
5		1	1	5	99.7395	62.076975	101.80745	0.54302	0.0		50.016428	62.076975	115.093425	0.854044	0.0	0	1	
6		1	1	6	32.8419	22.584425	54.626225	1.418831	LL		15.0034	22.584425	37.589825	0.664266	LL	0	1	
7		1	1	7	591.82668	655.782125	1248.64883	0.90282	0.0		625.648525	655.782125	1287.427395	0.965485	0.0	0	1	
8		1	1	8	5961.413425	5954.1056	11795.51903	1.021821	0.0		5899.546875	5954.1056	11733.65248	1.011217	0.0	0	1	
9		1	1	9	140.90425	132.912575	273.8167	1.007126	0.0		149.339186	132.912575	282.244425	1.128594	0.0	0	1	
10		1	1	10	12.6668	12.78225	25.38905	0.999174	LL		17.9971	12.78225	30.654425	1.40732	LL	0	1	
11		1	1	11	244.29375	328.4788	570.76225	0.78562	0.0		222.826275	328.4788	549.307125	0.898919	0.0	0	1	
12		1	1	12	690.05075	691.795975	1381.84673	1.023889	0.0		626.026125	691.795975	1317.7921	1.019664	0.0	0	1	
13		1	1	13	595.46545	590.555075	1186.02053	1.00706	0.0		594.9311	590.555075	1185.47998	1.1047	0.0	0	1	
14		1	1	14	25.9572	14.14005	39.21625	1.771693	LL		17.469575	14.14005	31.604425	1.23962	LL	0	1	
15		1	1	15	18.27245	24.70225	43.07875	0.742739	LL		13.123	24.70225	37.826225	0.621295	LL	0	1	
16		1	1	16	41.799205	40.618775	82.4164	1.029022	0.0		30.05646	40.618775	74.176225	0.826132	0.0	0	1	
17		1	1	17	171.96425	11.0448	28.24125	1.566998	LL		14.32425	11.0448	25.377225	1.297898	LL	0	1	
18		1	1	18	6061.1119	6405.49045	12466.60225	0.942237	0.0		6138.91295	6405.49045	12544.4034	0.956383	0.0	0	1	
19		1	1	19	5984.71095	5795.53225	11780.24328	1.039015	0.0		5990.921	5795.53225	11756.45433	0.969714	0.0	0	1	
20		1	1	20	1.918975	12.714275	14.6346	0.742425	LL		7.46268	12.714275	20.181625	0.896759	LL	0	1	
21		1	1	21	89.98985	117.28975	206.27975	0.785643	0.0		144.564	117.28975	261.65975	1.120546	0.0	0	1	
22		1	1	22	62.501025	56.76235	119.263375	0.91792	0.0		55.16129	56.76235	111.92375	0.971283	0.0	0	1	
23		1	1	23	564.42575	596.807125	1161.2323	1.013639	0.0		630.843375	596.807125	1185.29035	1.011132	0.0	0	1	
24		1	1	24	6185.98965	6295.2389	12929.22975	0.996737	0.0		6306.825725	6295.2389	12541.89563	1.016499	0.0	0	1	
25		1	1	25	34.33975	33.081475	67.420225	1.088005	LL		34.296825	33.081475	67.3783	1.036798	LL	0	1	
26		1	1	26	77.564825	90.456875	168.0217	0.871984	0.0		111.8925	90.456875	202.35375	1.236834	0.0	0	1	
27		1	1	27	17.9889	37.327225	55.326125	0.482192	0.0		24.468225	37.327225	61.79675	0.665514	0.0	0	1	
28		1	1	28	6203.267625	6009.278575	12229.5464	1.038111	0.0		6398.19925	6009.278575	12407.4698	1.064719	0.0	0	1	
29		1	1	29	6139.94175	5785.76775	11916.11763	1.057969	0.0		5746.08665	5785.76775	11530.26243	0.96927	0.0	0	1	
30		1	1	30	258.78025	254.897625	480.687875	0.922753	0.0		241.81925	254.897625	496.887625	0.946799	0.0	0	1	
31		1	1	31	41.428875	43.127475	84.55635	0.956873	0.0		17.99975	43.127475	61.11765	0.417325	0.0	0	1	
32		1	1	32	690.059175	636.18725	1316.2443	1.048893	0.0		690.212025	636.18725	12907.49023	0.96482	0.0	0	1	

上図のように、選択したフィールドのデータが実験ごとに表示されます。

2. スキャタープロットの作成

2-1. スキャタープロットの表示


メイン画面のScatterPlot表示アイコンをクリックしてください。あとは、自動でグラフが表示されます。

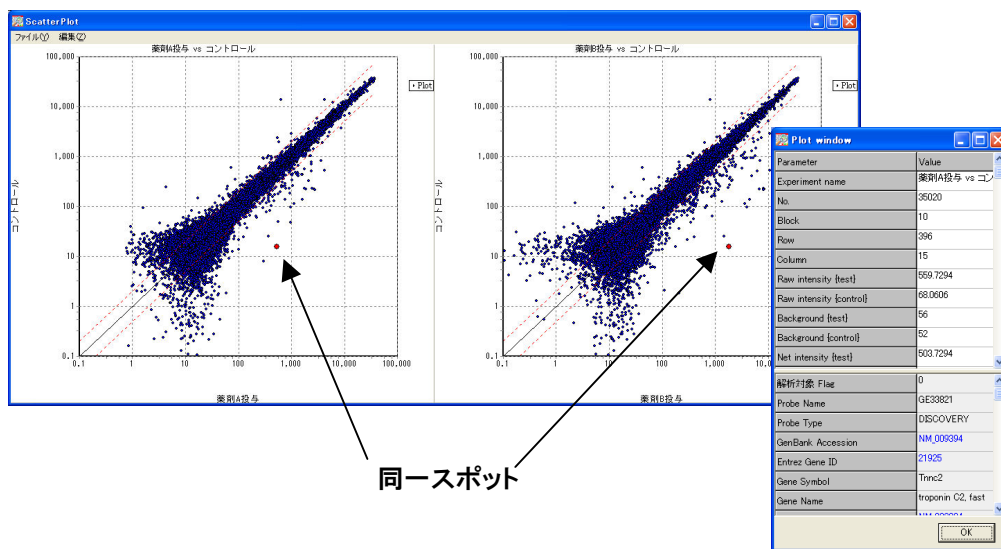


Microarray Name CodeLink(TM) Mouse Whole Gi Experiment Name

No.	Select	Block	Row	Column	薬剤A投与 vs
1	<input checked="" type="radio"/>	1	1	1	6242.29175
2	<input type="radio"/>	1	1	2	101.054775
3	<input type="radio"/>	1	1	3	75.301425
4	<input type="radio"/>	1	1	4	329.6794
5	<input type="radio"/>	1	1	5	39.7305
6	<input type="radio"/>	1	1	6	32.0418

スキャタープロット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変わります。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、スキャタープロット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。




上記までのステップでスキャタープロットを表示させ、2比較における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。また、注目したいプロットの表示を行うことで、薬剤A投与と薬剤B投与による該当遺伝子の発現パターンの比較を視覚的に確認す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。ばらつきの幅が広が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。

3. 発現変動遺伝子の抽出

3-1. Filter Options検索画面の表示

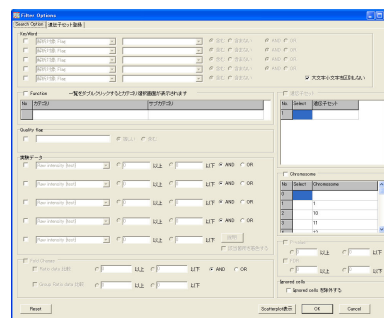
メイン画面のFilter Optionsアイコンをクリックし、Filter Options検索画面を開きます。



No.	Select	Block	Row	Column	実験A投与	Normalized
1	<input checked="" type="radio"/>	1	1	1	6242.29175	
2	<input type="radio"/>	1	1	2	101.054775	
3	<input type="radio"/>	1	1	3	75.301425	
4	<input type="radio"/>	1	1	4	329.6794	
5	<input type="radio"/>	1	1	5	39.7305	
6	<input type="radio"/>	1	1	6	32.0418	

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
実験データ
染色体番号
遺伝子セット



3-2. 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

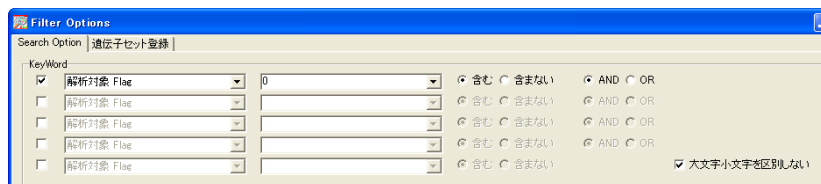
- ①コントロール遺伝子などを除去し、解析対象遺伝子のみにする。

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

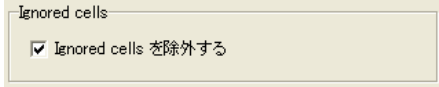
「2」は解析対象外遺伝子(Negative Control,Blank Spot,マーカーコントロールなど)

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を除外することができます。

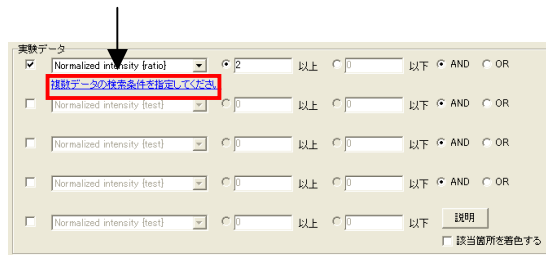


③発現差のあるデータを抽出する。

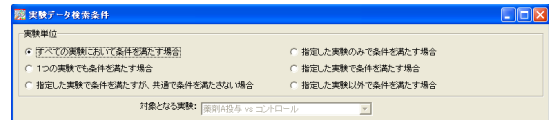
ここでは、実験データ検索を使用します。実験の内容や結果によっても変わってきますが、シグナルの差が任意の倍数以上あるデータを発現差のあるデータとして抽出します。

左のチェックボックスをクリックし、項目「Normalized intensity [ratio]」で条件を設定します。(下図は、「Normalized intensity [ratio]」を2以上にした例です。)

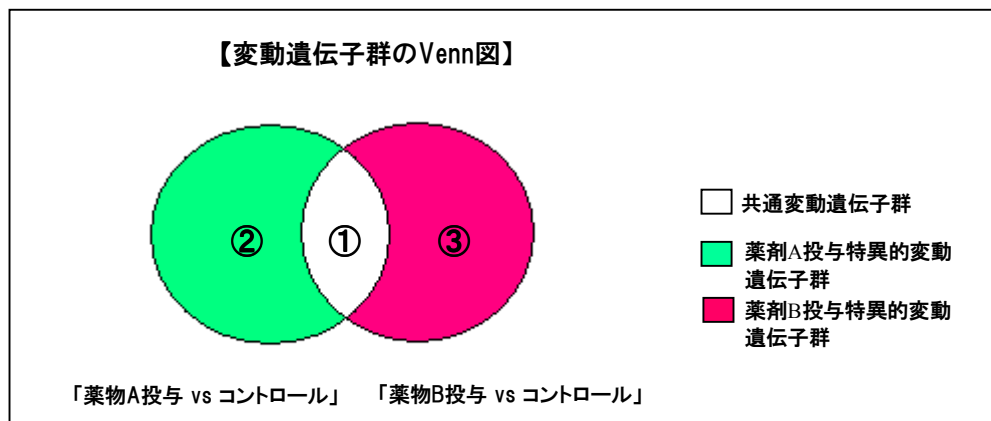
続いて、「複数データの検索条件を指定してください」の項目をクリックします。



指定した検索に対して、どのデータを対象として実行するか選択し、チェックボックスをクリックします。

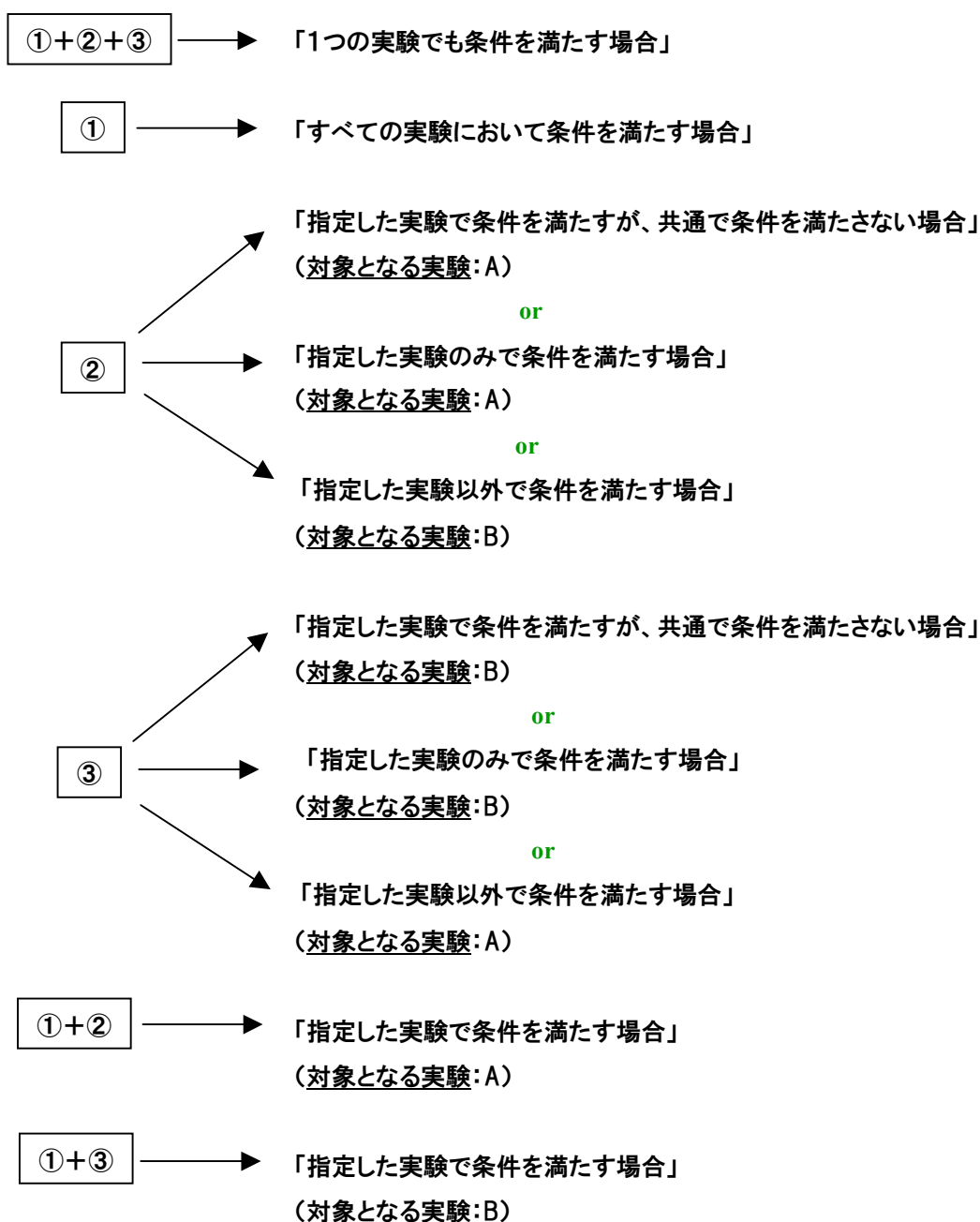


今回の2比較(「薬物A投与 vs コントロール」、「薬物B投与 vs コントロール」)では変動遺伝子群が3つの要素に分けられることを先ほど触れましたが、検索したい遺伝子群と実験条件検索の組み合わせを次ページで図示します。



検索したい遺伝子群

検索条件



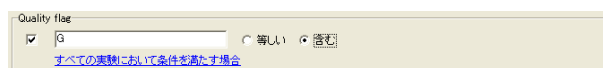
④ノイズレベルの信頼性の低いデータを除外する。(CutOff)

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。

④-1 Quality flagを用いてCutOffを行う。

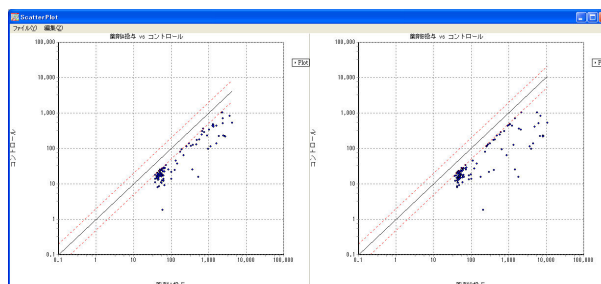
例1. 各実験データ共通に、テストサンプルまたはコントロールサンプルのうち、少なくともどちらかの実験データでQuality flagが「G」の遺伝子を抽出する場合

項目「Quality flag」で「Gを含む」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」を指定します。



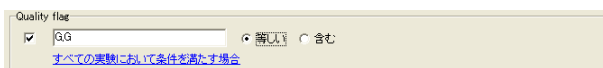
上記のようにCutOffを設定し、さらに③の発現変動遺伝子抽出の設定を同時に行うことで、発現変動を示し、なおかつノイズレベルのデータを除去した遺伝子のみを抽出することができます。

③で「Normalized intensity {ratio}」を2以上、かつ複数データの検索条件の選択を、「すべての実験において条件を満たす場合」に設定した場合のスカッタープロットを以下に示します。



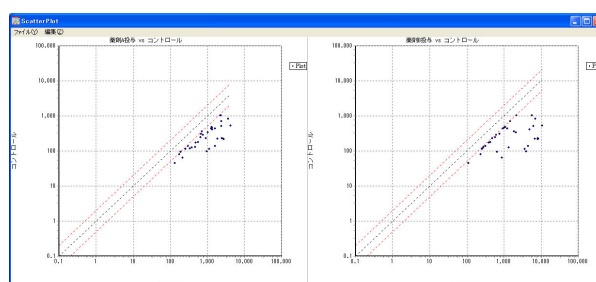
例2. 各実験データ共通に、テストサンプルとコントロールサンプル両方の実験データで、Quality flagが「G」の遺伝子を抽出する場合

項目「Quality flag」で「G,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」を指定します。



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。

③で「Normalized intensity {ratio}」を2以上、かつ複数データの検索条件の選択を、「すべての実験において条件を満たす場合」に設定した場合のスカッタープロットを以下に示します。



④-2 intensityを用いてCutOffを行う。

複数比較解析でintensityを用いてCutOffを行う場合は、まず実験データ共通のCutOff値を決めます。ここでは、共通のCutOff値として、2実験のネガティブコントロールの平均値のうち、値が小さい方（「薬物A投与 vs コントロール」が「27」、「薬物B投与 vs コントロール」が「29」だった場合では「27」の方）を採用した設定を以下に示します。（共通のCutOff値は、便宜検索結果と照らし合わせて、決めてください。）

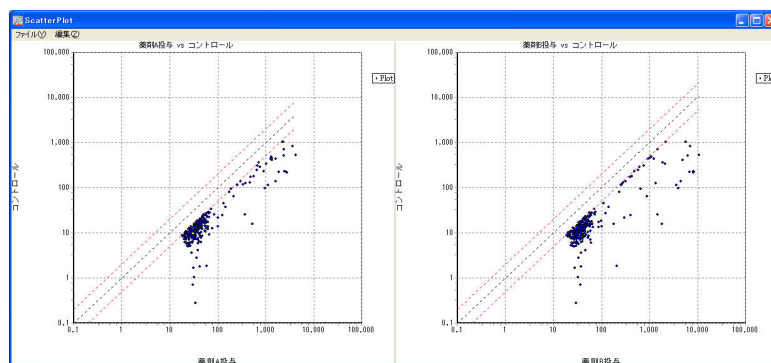
例1. 各実験データ共通に、2倍以上のUp変動を示し、かつNormalized intensity [sum]の値が「27」以上のデータを抽出する場合

実験データ

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

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

スクリーンショットは下図のようになります。



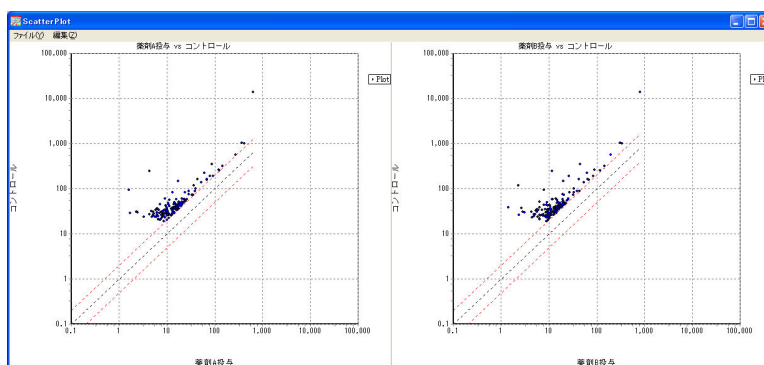
例2. 各実験データ共通に、2倍以上のDown変動を示し、かつNormalized intensity [sum]の値が「27」以上のデータを抽出する場合

実験データ

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

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

スクリーンショットは下図のようになります。



前述の例はCutOffの設定を、2つの実験データ、「薬物A vs コントロール」と「薬物B vs コントロール」に、まとめて設定しています。

そのため、一方の実験データでは条件を満たしていても（例：薬物A vs コントロールでは、Quality flagが「G,G」）、もう一方の実験データで条件を満たしていなければ（例：薬物B vs コントロールでは、Quality flagが「L,L」）データは抽出されません。

実験データごとに条件を設定したい場合は、複数データの検索条件に「指定した実験で条件を満たす場合」を設定し、さらに「対象となる実験：」で実験データを選択してください。

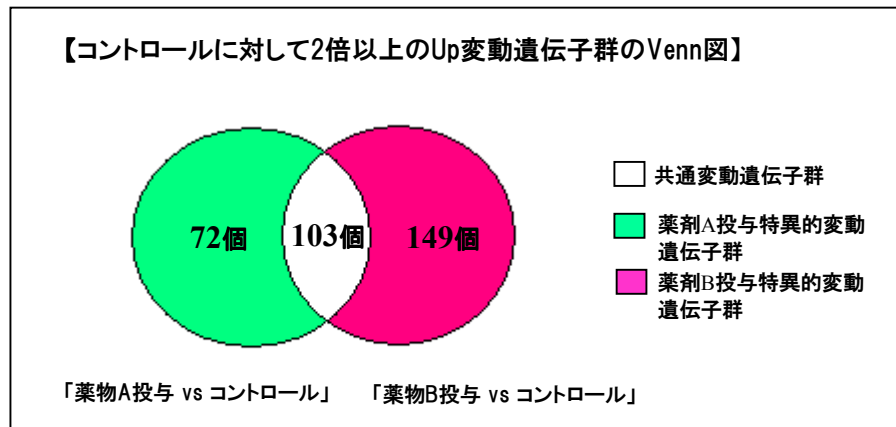
以上で、条件設定は終わりです。最後にFilter Options検索画面右下の「OK」ボタンをクリックするとデータの検索が実行され、下図のようにメイン画面にデータが表示されます。

No.	Select	Block	Row	Column	Normalized v	Normalized v	Normalized v	Quality flag	Quality flag	Normalized v	Normalized v	Normalized v	Quality flag	Quality flag	P-value	FDR	検出条件 Flag	Probe
80	<input type="radio"/>	1	1	103	50.42925	18.17965	2.73977	0	1	53.90675	18.17965	2.96527	0	1	1	1.0		GE148
84	<input type="radio"/>	1	1	107	259.17365	1135847	2.281766	0	0	254.7623	1135847	2.242928	0	0	1	1.0		GE144
225	<input type="radio"/>	1	3	56	1036.784075	3338446	3.102990	0	0	2124.017275	3338446	6.262293	0	0	1	1.0		GE235
617	<input type="radio"/>	1	7	106	58.851275	17.450375	3.372490	0	1	88.527225	17.450375	5.079113	0	1	1	1.0		GE151
717	<input type="radio"/>	1	8	110	1329.5222	485.649475	2.737623	0	0	1107.5285	485.649475	2.289509	0	0	1	1.0		GE423
991	<input type="radio"/>	1	12	2	51.162125	21.68715	2.359099	0	1	43.419275	21.68715	2.002074	0	1	1	1.0		GE137
2118	<input type="radio"/>	1	24	31	2706.68335	222.3105	12.148114	0	0	8362.196475	222.3105	37.569968	0	0	1	1.0		GE156
2801	<input type="radio"/>	1	31	83	64.97975	26.8813	2.417285	0	1	123.10075	26.8813	4.57942	0	1	1	1.0		GE141
4012	<input type="radio"/>	2	46	93	61.887325	12.984575	5.121183	0	1	38.08195	12.984575	3.151211	0	1	1	1.0		GE377
4072	<input type="radio"/>	2	47	55	40.601825	0.001	4.0601825	0	1	40.093625	0.001	4.093625	0	1	1	1.0		GE145
4160	<input type="radio"/>	2	48	59	2238.49395	1033.1294	2.166712	0	0	2302.49535	1033.1294	2.131868	0	0	1	1.0		GE122
4230	<input type="radio"/>	2	49	111	50.6975	24.96725	2.03056	0	1	58.946475	24.96725	2.381072	0	1	1	1.0		GE426
5351	<input type="radio"/>	2	61	73	369.389975	25.4924	14.49302	0	1	1489.850575	25.4924	58.442931	0	1	1	1.0		GE102
5569	<input type="radio"/>	2	63	101	144.3426	37.11455	3.889111	0	1	140.753125	37.11455	3.762397	0	1	1	1.0		GE155
5630	<input type="radio"/>	2	64	62	41.60275	15.793225	2.637037	0	1	182.38025	15.793225	11.568375	0	1	1	1.0		GE142
5650	<input type="radio"/>	2	64	83	57.236575	15.591675	3.44956	0	1	46.4329	15.591675	2.788566	0	1	1	1.0		GE145
5650	<input type="radio"/>	2	66	1	218.666225	63.300075	3.453944	0	0	892.677975	63.300075	14.100316	0	0	1	1.0		GE337
6812	<input type="radio"/>	2	77	25	704.706975	286.384125	2.635292	0	0	618.647575	286.384125	2.160202	0	0	1	1.0		GE145
7046	<input type="radio"/>	2	79	86	42.35655	17.459975	2.425911	0	1	50.12545	17.459975	2.670878	0	1	1	1.0		GE111
7896	<input type="radio"/>	3	90	87	40.7811	14.941025	2.904425	0	1	49.950125	14.941025	3.557441	0	1	1	1.0		GE370
8507	<input type="radio"/>	3	101	108	44.589475	15.547625	2.853153	0	1	64.088575	15.547625	4.122751	0	1	1	1.0		GE383
8683	<input type="radio"/>	3	102	53	62.395375	13.739675	4.541586	0	1	45.622425	13.739675	3.32073	0	1	1	1.0		GE149
9244	<input type="radio"/>	3	105	27	47.108675	18.9461	2.486458	0	1	44.7147	18.9461	2.3601	0	1	1	1.0		GE148
9583	<input type="radio"/>	3	108	86	48.254775	22.681975	2.12745	0	1	46.74625	22.681975	2.060943	0	1	1	1.0		GE321
9903	<input type="radio"/>	3	112	7	50.3249	16.726525	3.00885	0	1	36.88675	16.726525	2.255403	0	1	1	1.0		GE141
9967	<input type="radio"/>	3	113	17	64.784725	25.6538	3.304958	0	1	72.652875	25.6538	2.820351	0	1	1	1.0		GE345
10139	<input type="radio"/>	3	114	86	1977.957475	220.354675	8.522612	0	0	6968.23585	220.354675	31.622819	0	0	1	1.0		GE383
10211	<input type="radio"/>	3	115	58	57.255225	18.075475	3.167564	0	1	97.8229	18.075475	5.411913	0	1	1	1.0		GE394
11225	<input type="radio"/>	4	128	38	44.51395	12.8907	3.45316	0	1	47.0272	12.8907	3.649149	0	1	1	1.0		GE279
11469	<input type="radio"/>	4	130	104	61.9944	17.096675	3.620844	0	1	59.597	17.096675	3.458637	0	1	1	1.0		GE154
11742	<input type="radio"/>	4	132	111	43.822375	19.309625	2.29407	0	1	47.261075	19.309625	2.452163	0	1	1	1.0		GE149
12948	<input type="radio"/>	4	138	104	2309.9934	1017.57665	2.27006	0	0	5691.0179	1017.57665	5.59272	0	0	1	1.0		GE283

例1および2において、「Normalized intensity {ratio}」の複数データの検索条件の選択で、様々な条件を指定し、検索を行った結果、以下の絞り込みができました。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。

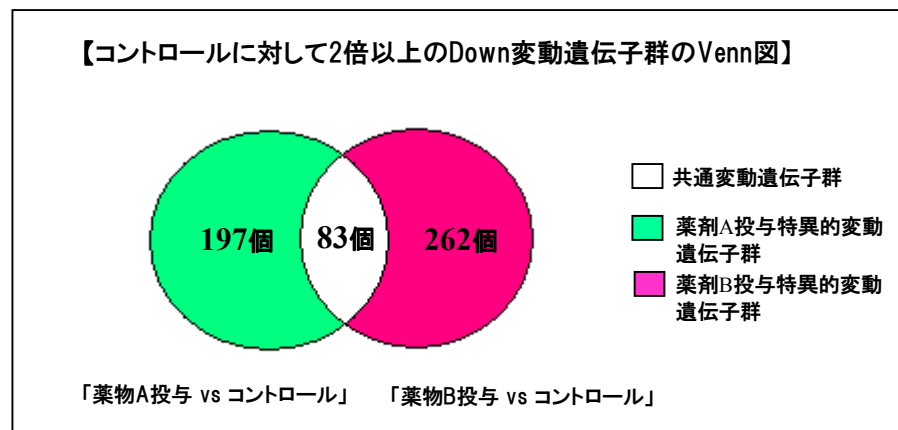
例1:薬剤AおよびBのうちいずれか一方でも、2倍以上のUp変動した遺伝子の総数:324個

- ・共通変動遺伝子群:103個
- ・薬剤A投与特異的変動遺伝子群:72個
- ・薬剤B投与特異的変動遺伝子群:149個



例2:薬剤AおよびBのうちいずれか一方でも、2倍以上のDown変動した遺伝子の総数:542個

- ・共通変動遺伝子群:83個
- ・薬剤A投与特異的変動遺伝子群:197個
- ・薬剤B投与特異的変動遺伝子群:262個



①および②の検索後の変動遺伝子は、AおよびBの薬剤投与によって発現が誘導され、両薬剤の作用を規定するうえで重要な遺伝子だと考えることができます。

この後、各変動遺伝子群において、Gene Ontologyによる生物学的機能の検索と、パスウェイの検索を行い、薬剤Aおよび薬剤Bの作用に関連する遺伝子の機能を調べていくこととなります。

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。