

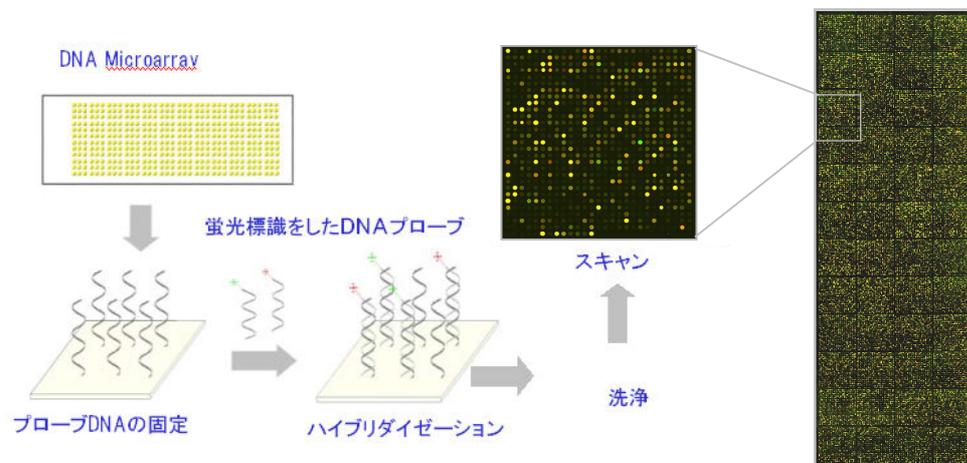
Microarray Data Analysis Tool

データ解析の進め方



まず、はじめに

マイクロアレイ(下図参照)とは数千から数万といった膨大な遺伝子情報をもつDNAプローブが貼り付けられた基板のことをいいます。実験対象となる細胞から抽出されたmRNAを蛍光標識をし、基板のDNAプローブに対して、ハイブリダイゼーションを行います。その後、スキャナーで蛍光強度を読み取り、数値化を行います。これにより、一度の実験で膨大な遺伝子の発現情報を得ることが出来ます。これまでのような個々の遺伝子を対象とした解析から網羅的な解析、ネットワーク的な解析が行えるようになりました。

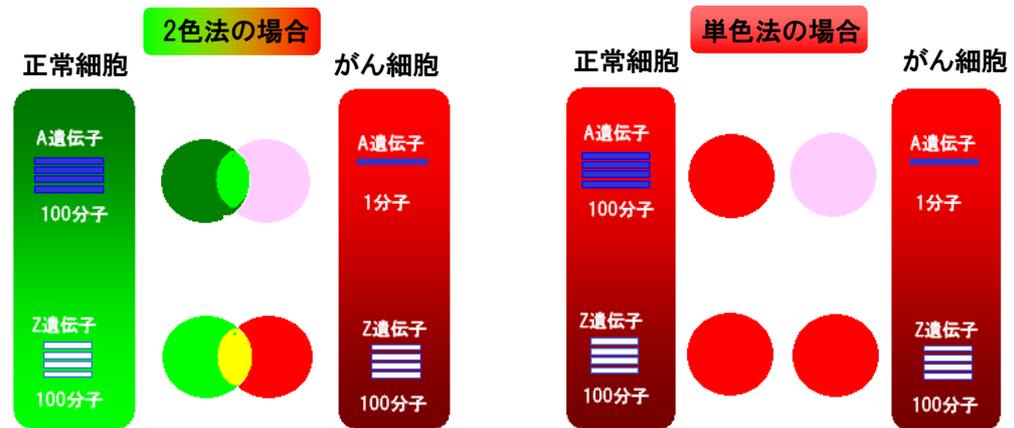


マイクロアレイの解析とは.....

実験後、スキャナーで読み取られた画像データは上図のように丸いスポットが規則正しく配置されたイメージとなります。この、丸いスポットの1つ1つがDNAプローブに対応し、これらの蛍光強度を数値化します。上図の場合黄色のスポットが多く、ところどころ緑と赤のスポットがあります。これは、異なるサンプルを異なる蛍光色素(Cy3,Cy5)でラベリングし、1枚のアレイ上で競合的ハイブリダーゼーションを行う2色法の場合です。黄色は赤と緑が混ざった状態となります。つまり、赤と緑の比率が発現差の比率となります。2色法の他に、異なるサンプルを同じ蛍光色素でラベリングし、1アレイで1サンプルのハイブリダーゼーションを行い、アレイ間のデータを比較する単色法があります。

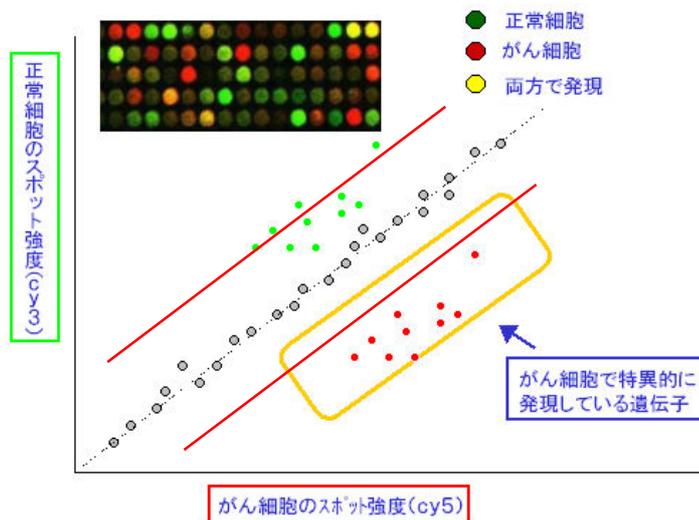
下記が2色法と単色法の概要となります。

2色法の場合は同一スポットに対して異なる蛍光色素でラベリングしたmRNAがハイブリダイズします。そのため、赤と緑という2つのシグナルデータを一つのデータとして表示するため緑のシグナル/赤のシグナルから計算したRatio値(比率)を用います。一方、単色法では別々のアレイに対して、同じ蛍光色素でラベリングしたmRNAを使用するため、Ratio値(比率)だけでなく、シグナル値の比較も可能となります。



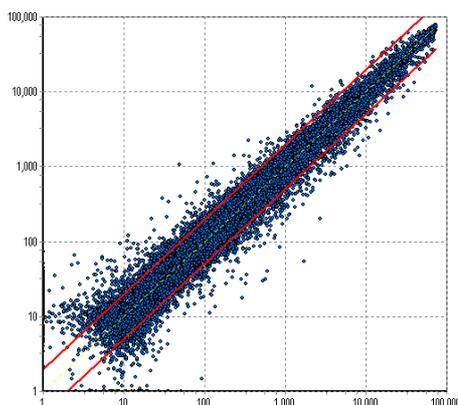
ここまでの話で、マイクロアレイの解析は異なるサンプル間におけるスポットの蛍光強度を比較するということがお分かりになったと思います。しかしながら、実際のアレイのデータは数千から数万という膨大なデータ量であるため、個々のデータを一つ一つみることは非常に大変な作業となります。

そこで、下図のようにスキャッタープロットと呼ばれ散布図がマイクロアレイの解析においてよく使われています。このグラフでは、正常細胞側のスポットの蛍光強度をY軸に、がん細胞側のスポットの蛍光強度をX軸に分布します。各サンプルの発現レベルが同じであれば、プロットは中心45度のラインに乗ってきます。一方、発現レベルの差があるほど、中心から外れてきます。例えば赤いプロットはがん細胞で特異的に高い発現を示したDNAプローブとなります。このように、スキャッタープロットはサンプル間の発現状態の全体像をを視覚的に捉えることができます。



それでは、どのくらいの発現差があれば
有意な差のあるデータといえるのでしょうか？

下図はおよそ3万のデータがプロットされたスカッタープロットです。先にも述べたように通常、比較サンプル間によるシグナルの比率をみます。スカッタープロットにおける中心45度のラインがRatio1のラインになります。そして、そのラインを真ん中とし、両方に平行移動した赤の2本のラインが2倍以上発現(Ratio2、Ratio0.5)のラインになります。



アレイの品質にもよりますが、同じサンプルを用いてデータを比較した場合、理論上は45度のラインのRatio1のラインにプロットが収束されます。しかしながら、マイクロアレイのデータの中にはアレイの品質、サンプル調整、蛍光標識、ハイブリダイゼーション、洗浄など様々な要因のノイズが含まれ、実際には1.5倍前後のライン付近にプロットが収束されます。そこで、2倍以上の差があれば有意なデータであるとして、データ抽出における閾値の基準に利用されています。この方法には統計的な裏付けはありませんが、マイクロアレイのようにスクリーニング的な要素が強い手法においては、非常によく使用されています。また、この方法で抽出されたデータは、後でリアルタイムPCRなどの定量解析法で確認試験を行うことを推奨します。逆に、マイクロアレイだけでデータをまとめようとする場合は、統計的な解析を求められるケースがおおいので、再現性実験が必要となります。

先の説明では発現差という部分でのノイズの影響を説明したのですが、シグナルの大きさでのノイズの影響はどうでしょうか？

上図のスカッタープロットにおいてシグナルが低いほどバラツキの幅が大きくなっていることがわかります。例えば、シグナルが1と10でも10倍の発現差があります。同じように1000と10000でも10倍の発現差があります。前者は、ノイズレベルのシグナルであり、シグナルが大きい後者の方がデータの信頼度が高くなります。大量のデータから発現差だけを基準にデータを抽出すると、上記のようなノイズ的なデータを多く含め抽出してしまいますので、基準を設け、ノイズデータを除去する必要があります。



それでは、何を基準にしてノイズの除去をすればいいのでしょうか？

使用するアレイやプラットフォームの違いによっても異なりますが、Negative Controlのシグナルを基準にする方法やバックグラウンドを基準にする方法が多く利用されています。

しかしながら、基準を決めたとしても、すべてのノイズが除去できるとは限りません。逆に除去した中に正しいデータが含まれている可能性もあります。先にも述べましたがノイズであるかどうか、再現性実験を行うことが一番よい方法であることは確かです。逆に、1比較のデータしかない場合は、基準を高めにし、より厳しい条件でデータを抽出することを推奨します。。

次はサンプル間のデータの比較(ノーマライゼーション)について説明します。

従来はbeta-actinやGAPDH等のHousekeeping遺伝子をコントロールとし、それらがサンプル間で一定レベルであるという前提でサンプル間を補正する方法がよく使用されてきました。しかしながら、最近のように全遺伝子を網羅したアレイの補正においては、比較サンプル間での遺伝子の総発現量はほぼ同じであるというグローバルノーマライゼーションが採用されています。刺激によって一部の遺伝子の変動があっても、全体としての発現レベルは変わらないという解釈になります。2色法の場合では異なる蛍光色素の影響を補正するためにloess法が利用されています。

以上、マイクロアレイの概要、解析を進める上でのポイントを紹介してきました。

次はいよいよソフトウェアの使用について説明します。本ソフトウェアは弊社受託サービス専用となっているため、データを読み込むだけですぐに解析に進めることができます。操作も非常に簡単です。ぜひ、トライしてください。



Microarray Data Analysis Tool

1 比較解析の進め方

ここからは、ソフトウェアの使用法と実際の解析例を合わせて説明します。

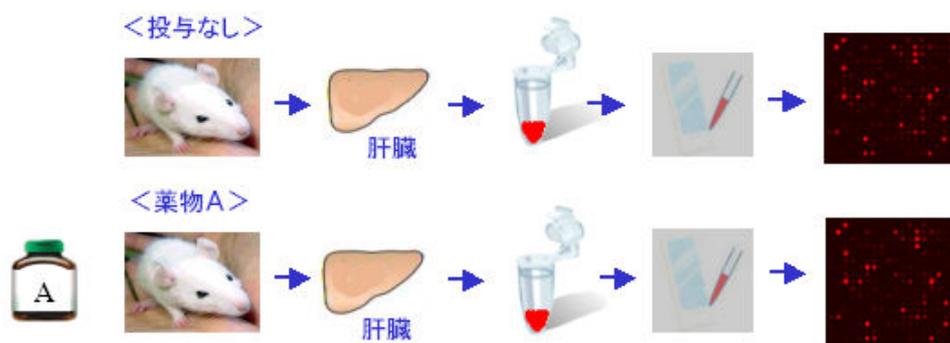
解析テーマは

「薬物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」です。

薬物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て、両サンプルの遺伝子発現プロファイルをマイクロアレイ実験で比較します。その後、比較データの中から、薬物投与マウスのみで高発現、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>

サンプル調整 → 組織の抽出 → 蛍光ラベル → ハイブリ → スキャニング



上記のテーマについて解析を行うには、具体的に下記のようなステップで解析を進める必要があります。

- ①薬剤投与によって発現が変動する(2倍Up、Downなど)遺伝子群の抽出
- ②発現変動遺伝子群について、Gene Ontologyによる生物学的機能の検索
- ③発現変動遺伝子群について、GenMAPPによるパスウェイの検索

まず①の工程では、2倍以上の変動を有意な差として、薬物投与マウスで発現が増加または減少している遺伝子を抽出します。さらに、ノイズ的なシグナルの除去も忘れてはいけません。

次に②、③の工程で、①で抽出した遺伝子がどのような機能またはパスウェイと関連があるかを調べます。ここで用いられるGene Ontology(GO)とは生物学的機能を簡単な記述で表したもので、抽象度によって分類された階層構造をもっています。またGenMAPPは、細胞の分子間相互作用情報(パスウェイ)を図で表したものです。現在までの研究で機能が明らかになった遺伝子には、このGOやパスウェイがアノテーション情報として付けられるので、発現変動遺伝子がどのような機能やパスウェイと関連付けられているかを調べれば、薬物投与によってどのような生物学的作用が起こっているのかを推測することができます。

それでは、ソフトウェアを立ち上げて進めていきましょう！

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

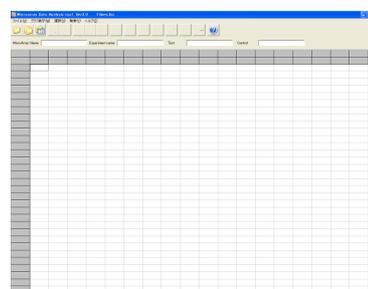
CDに入っている「Microarray Data Analysis Tool」のフォルダを、自分のコンピュータにコピーしてください。

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。

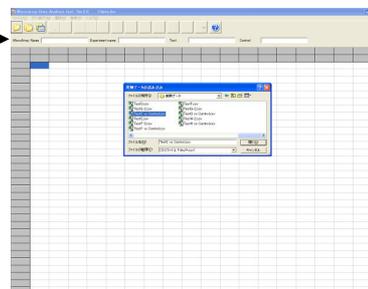


ソフトウェアが起動後は右図のメイン画面が表示されます。



1-3. 実験データ読み込み

1比較データ解析アイコンをクリックし、「実験データ」フォルダを選択し、フォルダ内から希望の実験データファイル(〇〇 vs 〇〇.csv)を読み込ませてください。



実験データファイルを読み込み後、右図のようなデータが表示された画面が現れます。

左の側が実験数値データで、各プローブに該当するアノテーション情報が右側に表示されます。青の文字はダブルクリックすると該当データベースへハイパーリンクします。

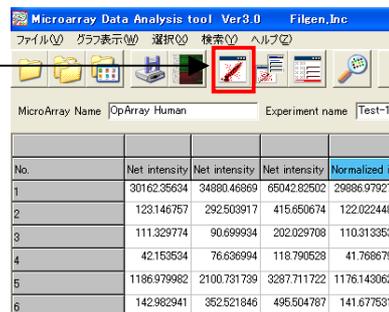
なお、ここで表示されているデータはExcelで作成された解析データと同じデータとなります。

Probe ID	Gene Name	Expression Value	Annotation
123456	Gene A	1.2	Gene A (clickable)
789012	Gene B	3.5	Gene B (clickable)
345678	Gene C	0.8	Gene C (clickable)
901234	Gene D	2.1	Gene D (clickable)
567890	Gene E	4.3	Gene E (clickable)

2. スキャッタープロットの作成

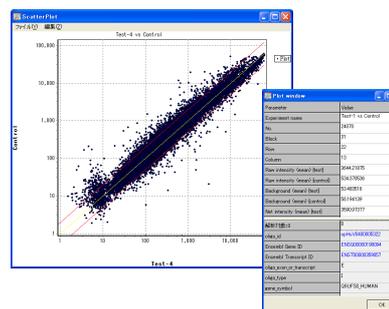
2-1. スキャッタープロットの表示

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。



No.	Net intensity	Net intensity	Net intensity	Normalized
1	30162.35634	34980.46969	65042.82502	29886.97927
2	123.146767	292.503917	415.650674	122.022448
3	111.329774	90.699934	202.029708	110.313353
4	42.153534	76.636994	118.790528	41.768679
5	1186.979982	2100.731739	3287.711722	1176.143062
6	142.982941	852.521846	495.504787	141.677531

スキャッタープロットの上の個別のプロットをクリックするとプロットの色が青から赤に変わります。また、メイン画面上の対応するプロンプ(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。
さらに、メイン画面のPlotwindowアイコンをクリックすると、スキャッタープロットの各プロットの個別のデータを表示するPlotwindow画面が現れます。



上記までのステップでスキャッタープロットを表示させ、比較サンプル間における遺伝子発現の全体像を見ることができました。スキャッ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。

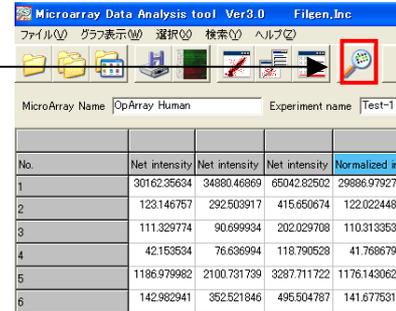
次は、変動遺伝子の抽出となります。



3. 発現変動遺伝子の抽出

3-1. Filter Options検索画面の表示

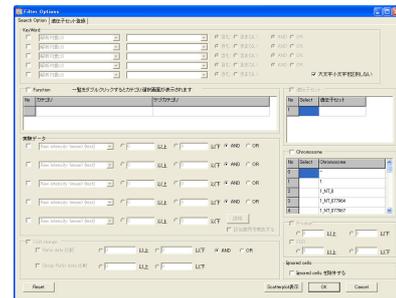
メイン画面のFilter Optionsアイコンをクリックし、Filter Options検索画面を開きます。



No.	Net intensity	Net intensity	Net intensity	Normalized r
1	30162.35634	34980.46969	65042.82502	29886.97927
2	123.146767	292.503917	415.650674	122.022448
3	111.329774	90.699934	202.029708	110.313353
4	42.153534	76.636994	118.790528	41.768679
5	1186.979982	2100.731739	3287.711722	1176.143062
6	142.982941	852.521846	495.504787	141.677531

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



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。

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、3」という数字で識別をしています。

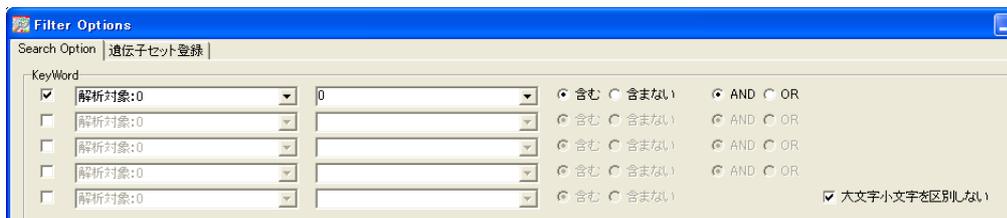
「0」は解析対象遺伝子

「2」は解析対象外遺伝子(NegativeControl,Blank,QCコントロールなど)

「3」はデータベースの更新により対応プローブの情報が削除された遺伝子

***アレイの種類により解析対象の条件設定が必要ない場合もあります。**

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:0」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。



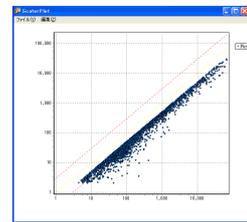
③発現差のあるデータを抽出する。

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、ここではシグナルの差が2倍以上あるデータを差のあるデータとして抽出します。

コントロールに対して2倍以上のUp変動データを抽出する場合

左のチェックボックスをクリックし、項目「Normalized intensity (mean) [ratio]」で「2」以上を設定をします。

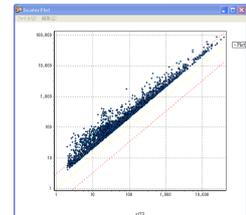
スキッタープロットは下図のようになります。



コントロールに対して2倍以上のDown変動データを抽出する場合

左のチェックボックスをクリックし、項目「Normalized intensity (mean) [ratio]」で「0.5」以下を設定をします。

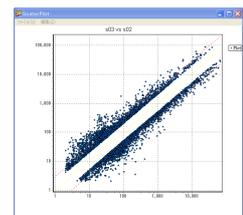
スキッタープロットは下図のようになります。



コントロールに対して2倍以上のUpとDown変動データを一度に抽出する場合

左のチェックボックスをクリックし、項目「Normalized intensity (mean) [ratio]」で「2」以上を設定をします。同様に、項目「Normalized intensity (mean) [ratio]」で「0.5」以下を設定をします。この時、AND ORの選択でORを選択します。

スキッタープロットは下図のようになります。



④ノイズレベルの信頼性の低いデータを除外する。

ここでは、③と同様、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、エクセルで作成された解析データのCutoffの条件設定を参考に設定をします。

例1

Normalized intensity (mean) がnegative-controlのNet intensity(sum)の平均値(=300)であれば、左のチェックボックスをクリックし、項目「Normalized intensity (mean) {sum}」で「300」以上を設定をします。

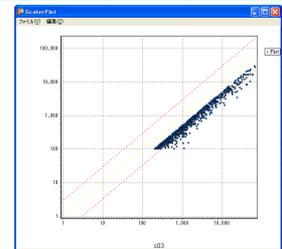
例2

Normalized intensity (mean) がBackground (mean) $\times 2$ (=100) であれば、左のチェックボックスをクリックし、項目「Normalized intensity (mean) {test}」で「100」以上、同様に「Normalized intensity (mean) {control}」で「100」以上と設定をします。

下図では例1の条件設定における、ノイズ除去の設定例です。

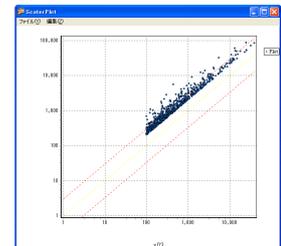
コントロールに対して2倍以上のUp変動データを抽出する場合

スキッタープロットは下図のようになります。



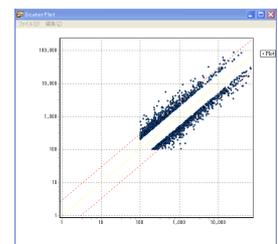
コントロールに対して2倍以上のDown変動データを抽出する場合

スキッタープロットは下図のようになります。



コントロールに対して2倍以上のUpとDown変動データを一度に抽出する場合

スキッタープロットは下図のようになります。



ここでは、Filter Optionsなどにより抽出されたデータ群を遺伝子セットとして登録する方法を説明します。一度登録すれば、後日同じデータを簡単に選択することができます。また、同じ規格のアレイであれば、別の実験データにおいて登録した遺伝子セットの結果も抽出できます。

4. 遺伝子セットの登録方法

①ここでは、先と同様に2倍のUp変動したデータを抽出します。抽出結果がメイン画面に表示されたのを確認後、Filter Options検索画面において遺伝子セット登録を選択します。



②次に、新規登録のボタンを押します。入力画面が表示されますので、検索名に名前を入力します。ここでは「2Up」と入力します。また、変更ボタンを押すと、色の変更ができます。ここでの色は、登録した遺伝子セットのプロット色となります。ここでは「青」にします。最後にOKをクリックします。



③登録後は右図のように登録情報が表示されます。



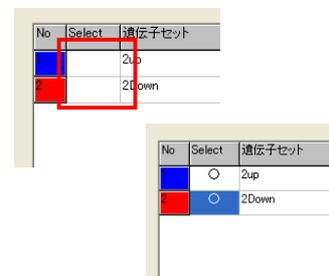
④ 次にFilter Options検索画面に戻り今度は2倍のDown変動の条件設定で抽出を行います。抽出結果がメイン画面に表示されたのを確認後、先と同様に遺伝子セット登録を選択します。ここでは「2Down」と入力します。また、色の登録は「赤」にします。最後にOKボタンをクリックします。



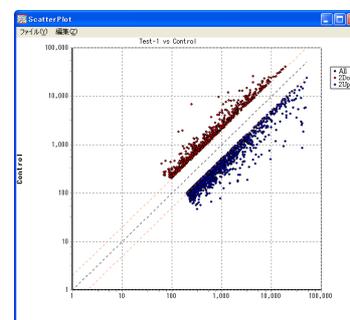
⑤ 登録後は右図のように登録情報が追加されます。



⑥ [select]の列の空欄をダブルクリックすると、○が表示されます。この状態で、今、2つの遺伝子セットが選択されたこととなります。ここで、OKボタンをクリックすると、2つの遺伝子セットのみのデータがメイン画面に表示されます。



⑦ ⑥の状態でスクアタープロットを表示させると右図のように、先ほど設定した色で表示させることができます。



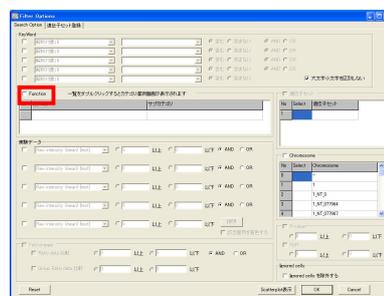
5. カテゴリー検索

先の説明では変動しているデータに着目して抽出を行いました。次は、自分の興味ある機能カテゴリーの検索について説明します。

Function検索

①最初にFunctionのチェックボックスをクリックして、チェックを入れてください。

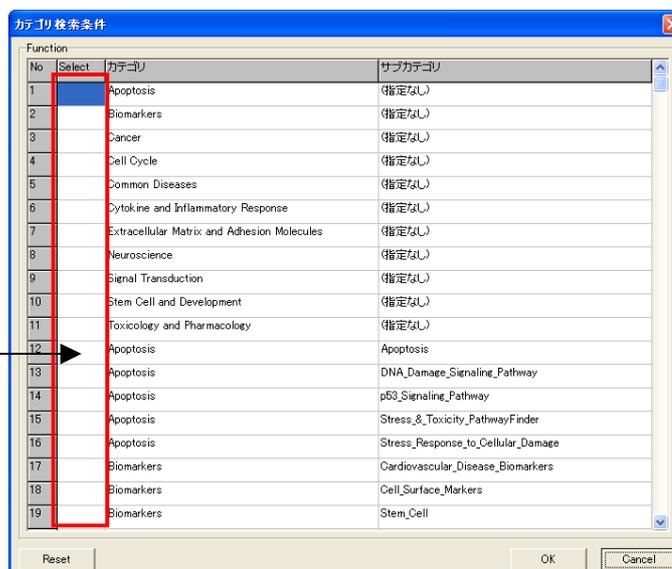
*Human, Mouse, Ratアレイに対応しています。



②検索時にはカテゴリー項目のSelectボックスをダブルクリックしてください。



③下図のカテゴリー検索条件画面が表示されます。分類には大まかなカテゴリー分類と、各カテゴリーをさらに分類したサブカテゴリーがあります。つまり大カテゴリーにおける該当遺伝子数は、その全てのサブカテゴリーの該当遺伝子数の合計と同じになります。



カテゴリーの選択は[select]の列の空欄をダブルクリックします。○が表示されますので、該当カテゴリーが選択されたこととなります。

6. 抽出した遺伝子の機能解析

絞り込んだデータについて、どのような機能をもった遺伝子があるのかを調べるために Gene Ontology 解析を行います。

① Filter Optionsでデータの絞り込みを行います。

ここでは下記の条件設定を行います。

・解析対象遺伝子のみにする。

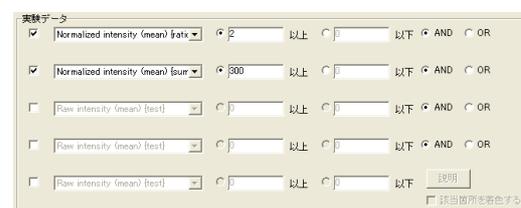
・Ignore cellsを除外する。

・発現差のあるデータ(2倍以上)を抽出する。

・ノイズレベルの信頼性の低いデータを除外する。
(Normalized intensity (mean) {sum} が300以上)

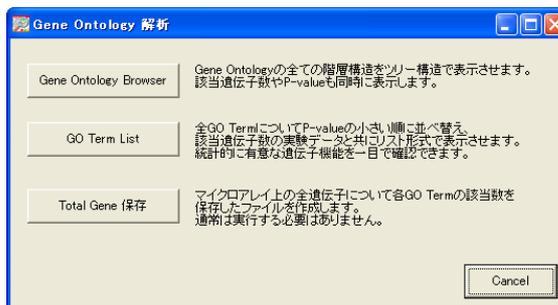


以上で、条件設定は終わりです。最後に Filter Options 検索画面右下の「OK」ボタンを押すとデータの抽出が実行され、右図のようにメイン画面にデータが表示されます。ここでは328件抽出されました。



② GO解析アイコン をクリックし、データベース選択ウインドウを開きます。ここで解析に Gene Ontologyを使用するか、GO Slimを使用するかの選択ができます。本解析では Gene Ontologyを使用するので、そちらをクリックしてください。

③ 次にデータの出力形式を選択します。ここで選択できる項目は、以下の機能をもっています。



・Gene Ontology Browser・・全てのGOカテゴリーを階層構造で表示させます。

・GO Term List統計学的に有意なGOカテゴリーのみを、有意な順にリスト形式で表示させます。

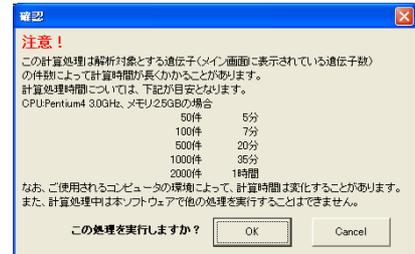
・TotalGene保存..... マイクロアレイ上の全遺伝子についての、全GOカテゴリー該当遺伝子数を記録したファイルを作成します。ここで作成したファイルは納品の時点ですでに作成・添付されているので、この機能は実行する必要がありません。

Gene Ontology Browser

本機能は全てのGOカテゴリーを階層構造で表示し、同時に該当遺伝子数や統計学的有意確率も表示します。この階層構造は生物学的機能を抽象度によって分類したものであり、階層を下にたどっていくごとに、より具体的なGOカテゴリーが出現します。この性質を利用して、自分の興味ある機能について上位階層から見ていき、下位のより細かいGOカテゴリーについて有意なものを調べることができます。

①まず、Gene Ontology解析ウインドウで「Gene Ontology Browser」のアイコンをクリックします。右図のような警告画面が表示されますので、計算処理時間について、確認の上OKボタンをクリックしてください。

その後計算が始まります。計算が終了すると以下のような画面が表示されます。



Gene Ontology Browser初期画面

Gene Ontology Browserでは最初は3つのGOカテゴリーしか表示されません。しかし各カテゴリー名の前にある「+」をクリックしていくたびに、下位のGOカテゴリーが表示されていきます。

それでは例として「molecular function」の下位に属する「catalytic activity」というカテゴリーについて調べてみましょう。「molecular function」の前の「+」をクリックすると以下のような画面が表示されます。

molecular function階層画面

ここで「catalytic activity」の項目(赤線で囲んだ領域)を見てみますと、Z-scoreが0より大きく、なおかつP-valueが0.05以下となっており、このGOカテゴリーが統計学的に有意だということを示しています。ここでいう統計学的有意差とは、発現が変動したものを抽出したデータと、マイクロアレイ上の全データという2つのグループの遺伝子に対して、各GOカテゴリーの該当遺伝子数を数え、その数に偏りがあるかを検定したものです。つまりこの検定で有意差があれば、発現が変動した遺伝子群において有意に機能するGOカテゴリーとして解釈ができます。

さらに「catalytic activity」より下の階層を見てみますと、様々な触媒活性についてのGOカテゴリーがあることが分かります。

+	catalysis of free radical formation	0(0)	0(0)	0	1
+	cyclase activity	1(0)	25(0)	1.343	0.25622
+	deaminase activity	1(0)	19(5)	1.688	0.2036
+	demethylase activity	0(0)	5(0)	0	1
+	glycogen debranching enzyme activity	0(0)	1(0)	0	1
+	glyoxalase III activity	0(0)	0(0)	0	1
+	helicase activity	4(4)	133(121)	2.021	0.00997
+	hydrolase activity	37(18)	2273(900)	2.167	0.02333
+	isomerase activity	1(1)	179(105)	-0.736	0.72615
+	ligase activity	7(6)	339(246)	1.594	0.11661
+	lipic acid synthase activity	0(0)	1(1)	0	1
+	lyase activity	4(2)	164(92)	1.559	0.12309
	Mo-molybdopterin cofactor sulfurase activity	0(0)	1(1)	0	1
	Mo-molybdopterin synthase activity	0(0)	0(0)	0	1
	N-acetylneuraminic acid phosphate synthase activity	0(0)	1(1)	0	1
	N-ethylmaleimide chlorohydrolase activity	0(0)	0(0)	0	1
+	oxidoreductase activity	13(6)	922(567)	0.759	0.42166
	peptidoglycan synthetase activity	0(0)	0(0)	0	1
	phytoene synthase activity	0(0)	0(0)	0	1
	polyketide synthase activity	0(0)	0(0)	0	1
	quinolinate synthetase A activity	0(0)	0(0)	0	1
	recombinase activity	0(0)	2(2)	0	1

catalytic activityの下位階層画面

様々な分子間結合についてのGOカテゴリーがありますが、そのなかでも特にP-valueが小さいものに「hydrolase activity」(赤線で囲んだ領域)があります。「catalytic activity」のみの方がP-valueは小さいのですが、こちらの方が機能としてより具体的で分かりやすくなっています。

各GOカテゴリーは「+」が付いている場合、より下位の階層が存在するということなので、もちろんこの「hydrolase activity」にも下位階層が存在します。

The screenshot shows the GeneOntology browser interface. The search term is 'catalytic activity'. The left pane shows a tree view where 'catalytic activity' is expanded to 'hydrolase activity'. The right pane shows a table of results for 'hydrolase activity' with columns for Term, Changed Genes, Total Genes, Z score, and P-value. The term 'hydrolase activity, acting on ester bonds' is highlighted with a red box, showing 15(2) changed genes, 521(18) total genes, a Z score of 3.723, and a P-value of 0.00111.

Term	Changed Genes	Total Genes	Z score	P-value	
+	molecular_function	157(0)	13724(0)	0	1
+	antioxidant activity	0(0)	19(12)	0	1
+	auxiliary transport protein activity	0(0)	23(0)	0	1
+	binding	110(7)	9855(442)	-0.258	0.6557
+	catalytic activity	76(5)	5360(711)	1.882	0.02102
+	catalysis of free radical formation	0(0)	0(0)	0	1
+	cyclase activity	1(0)	25(0)	1.343	0.25622
+	deaminase activity	1(0)	19(5)	1.688	0.2036
+	demethylase activity	0(0)	5(0)	0	1
+	glycogen debranching enzyme activity	0(0)	1(0)	0	1
+	glyoxalase III activity	0(0)	0(0)	0	1
+	helicase activity	4(4)	133(121)	2.021	0.00997
+	hydrolase activity	37(18)	2273(900)	2.167	0.02333
	beta-alanyl-dopamine hydrolase activity	0(0)	0(0)	0	1
	beta-alanyl-histamine hydrolase activity	0(0)	0(0)	0	1
+	deacetylase activity	0(0)	15(1)	0	1
	DNA demethylase activity	0(0)	1(1)	0	1
	GPI-anchor transamidase activity	0(0)	4(4)	0	1
+	hydrolase activity%, acting on acid anhydrides	11(0)	701(0)	1.058	0.27258
+	hydrolase activity%, acting on acid carbon-carbon bonds	2(0)	0(0)	0	1
+	hydrolase activity%, acting on acid carbon-phosphorus bo...	0(0)	0(0)	0	1
+	hydrolase activity%, acting on acid halide bonds	0(0)	0(0)	0	1
+	hydrolase activity%, acting on acid phosphorus-nitrogen b...	0(0)	0(0)	0	1
+	hydrolase activity%, acting on acid sulfur-nitrogen bonds	0(0)	0(0)	0	1
+	hydrolase activity%, acting on acid sulfur-sulfur bonds	0(0)	0(0)	0	1
+	hydrolase activity%, acting on carbon-nitrogen (but not p...	2(0)	107(16)	0.705	0.35
+	hydrolase activity%, acting on carbon-sulfur bonds	0(0)	0(0)	0	1
+	hydrolase activity%, acting on ester bonds	15(2)	521(18)	3.723	0.00111
+	hydrolase activity%, acting on ether bonds	0(0)	0(0)	0	1
+	hydrolase activity%, acting on glycosyl bonds	0(0)	121(48)	0	1
+	peptidase activity	8(2)	657(244)	0.178	0.84973
+	serine hydrolase activity	0(0)	237(0)	0	1
+	integrase activity	0(0)	5(5)	0	1
+	isomerase activity	1(1)	179(105)	-0.736	0.72615

hydrolase activityの下位階層画面

「hydrolase activity」の下位に属するGOカテゴリーのなかで、Z-scoreが0より大きく、なおかつP-valueが0.05以下のものは「hydrolase activity, acting on ester bonds」(赤線で囲んだ領域)のみでした。また、ここでGOカテゴリー名をダブルクリックすると、このGOカテゴリーをもつ遺伝子の実験データが右側領域に表示されます。

このように自分の興味のあるカテゴリーについて階層構造をたどっていき、そのなかでZ-scoreが大きく、かつP-valueのより小さいGOカテゴリーを選択していくことによって、より具体的で統計学的に有意なGOカテゴリーを見つけることができます。

またBrowser上部のKeyword欄に、自分の調べたいカテゴリー名を入力し検索を行うことで、目的のGOカテゴリーの解析データを、すぐに調べることができます。

②表示されたGO Term Listは左上のファイル(Z)で保存(Y)することができます。表示されている情報がそのまま保存されます。

Term	Changed Genes
hydrolase activity/, acting on ester bonds	15(2)
regulation of cyclin-dependent protein kinase activity	3(1)

③右図は出力項目選択画面で「遺伝子IDのみ」を選択し、保存したデータをエクセルで開いたものです。GO Termの横に該当データが表示されます。

④このリストから希望のデータを抽出したい場合、エクセルのオートフィルタ機能を使用します。右図の様にデータタイトル行を選択し、データ(D)のフィルタ(F)でオートフィルタ(F)を選択します。

⑤選択後は 各列のタイトル行にオートフィルタの機能が追加されます。

⑥⑤で追加したオートフィルタの機能を用いて、リスト内から「Z-score>0 かつ P-value<0.01」を満たすGOカテゴリーのみを抽出します。

まずP-valueの列の▼をクリックしてください。すると右図のように列内の値が表示されます。そして、ここで(オプション...)をクリックすると、オートフィルタ オプションが表示されます。

抽出条件の指定:
P-value: [0.01] [昇順]

AND OR

オートフィルタ オプションで右図のように、数値と条件を入力し、OKボタンをクリックします。Z-scoreの方も、条件を変えて同様の操作をします。

前述の条件「Z-score>0 かつ P-value<0.01」で抽出を行うと、右図のような結果になりました。

Term	Changed Genes	Total Genes	Z-score	P-value	Gene ID
hydrolase activity/, acting on ester bonds	15(2)	521(18)	3.723	0.00118	ENSMUSG000000
regulation of cyclin-dependent protein kinase activity	3(1)	16(4)	6.648	0.00118	ENSMUSG000000
amino acid biosynthetic process	4(2)	400(7)	5.29	0.00146	ENSMUSG000000
negative regulation of kinase activity	4(0)	45(0)	4.908	0.00219	ENSMUSG000000
negative regulation of protein kinase activity	4(1)	45(18)	4.908	0.00219	ENSMUSG000000
negative regulation of transferase activity	4(0)	46(1)	4.838	0.00236	ENSMUSG000000
mitochondrion	21(2)	389(834)	3.288	0.00331	ENSMUSG000000
nucleotide binding	38(34)	2207(1899)	2.768	0.0044	ENSMUSG000000
negative regulation of catalytic activity	4(0)	61(3)	3.886	0.00608	ENSMUSG000000
response to endogenous stimulus	1(0)	346(0)	3.079	0.0069	ENSMUSG000000
amino biosynthetic process	4(0)	84(0)	3.861	0.00714	ENSMUSG000000
negative regulation of cyclin-dependent protein kinase activity	2(2)	100(0)	5.629	0.00768	ENSMUSG000000
synaptic vesicle exocytosis	2(1)	10(5)	5.629	0.00768	ENSMUSG000000
heterotrimeric G-protein complex	3(3)	34(34)	4.141	0.00893	ENSMUSG000000
cytoplasmic part	52(0)	3181(1)	2.41	0.00897	ENSMUSG000000

以上、抽出した発現変動遺伝子に対して、Gene Ontology解析を行うことによって、発現変動遺伝子がどのような機能をもっているのかを調べることができました。本解析は薬物投与によって発現量が増加した遺伝子に対する機能解析なので、この薬物はマウスに投与すると⑥の図のGOカテゴリーにある作用を発揮するものと考えられます。

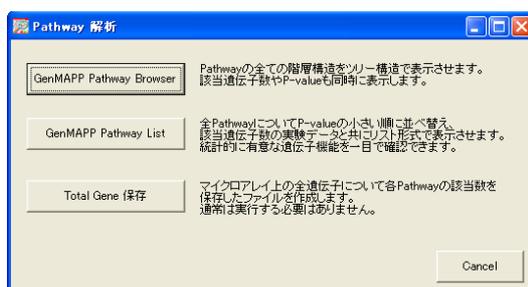
7. 抽出した遺伝子のパスウェイ解析

絞り込んだデータについて、どのようなパスウェイと関連のある遺伝子が多いのかを調べるためにPathway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6. 抽出した遺伝子の機能解析のときに指定した条件をそのまま使用します。

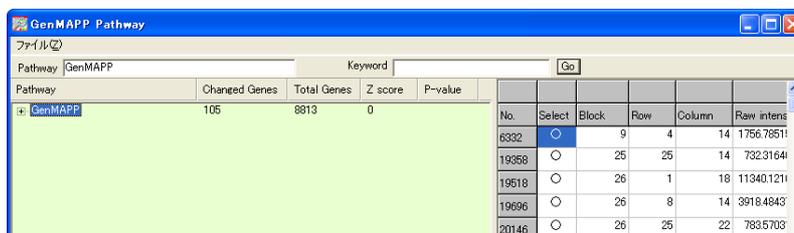
- ② Pathway解析アイコン  をクリックし、Pathway解析ウインドウを開きます。Gene Ontology解析と同様、データの出力形式を選択することができます。またここで選択できる項目は、Gene Ontology解析とほぼ同じ機能もっています。



GenMAPP Pathway Browser

本機能は全てのパスウェイを階層構造で表示し、同時に該当遺伝子数や統計学的有意確率も表示します。この階層構造はパスウェイをカテゴリーごとに分類したものであり、一番下の階層に正式なパスウェイ名のデータが表示されます。

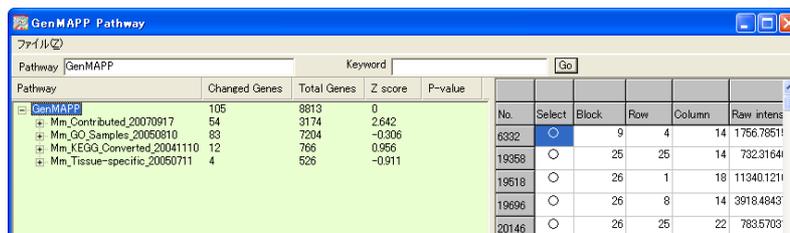
- ①まず、Pathway解析ウインドウで「GenMAPP Pathway Browser」のアイコンをクリックします。その後計算が始まり、計算が終了すると以下のような画面が表示されます。



Pathway	Changed Genes	Total Genes	Z score	P-value
GenMAPP	105	8813	0	

No.	Select	Block	Row	Column	Raw intens
6332	<input checked="" type="radio"/>	9	4	14	1756.78511
19358	<input type="radio"/>	25	25	14	732.31641
19518	<input type="radio"/>	26	1	18	11340.1211
19696	<input type="radio"/>	26	8	14	3918.4843
20146	<input type="radio"/>	26	25	22	783.5703

GenMAPP Pathway Browserでは最初は1つのカテゴリーしか表示されません。カテゴリー名のある「+」をクリックしていくたびに、下位のカテゴリーが表示されていきます。まず最初に、「GenMAPP」の前の「+」をクリックすると以下のような画面が表示されます。



Pathway	Changed Genes	Total Genes	Z score	P-value
GenMAPP	105	8813	0	
Mm_Contributed_20070917	54	3174	2.642	
Mm_GO_Samples_20050810	83	7204	-0.306	
Mm_KEGG_Converted_20041110	12	766	0.956	
Mm_Tissue-specific_20050711	4	526	-0.911	

No.	Select	Block	Row	Column	Raw intens
6332	<input checked="" type="radio"/>	9	4	14	1756.78511
19358	<input type="radio"/>	25	25	14	732.31641
19518	<input type="radio"/>	26	1	18	11340.1211
19696	<input type="radio"/>	26	8	14	3918.4843
20146	<input type="radio"/>	26	25	22	783.5703

ここで表示される4つのカテゴリーは、パスウェイの情報源に基づくカテゴリーです。それぞれ以下のような特徴があります。

Contributed: 既知の分子間相互作用情報をパスウェイごとに分類し、図で表したものです。GenMAPPのパスウェイデータというと、この「Contributed」を指すことがほとんどです。

GO samples: Gene Ontologyでの分類をより簡潔に表したものです。分類の数は多いのですが、この図には「Contributed」のような相互作用情報までは記されていません。

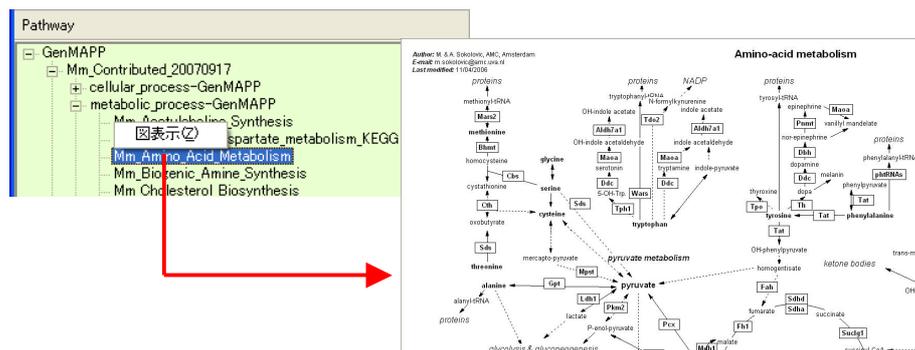
KEGG converted: KEGG (Kyoto Encyclopedia of Genes and Genomes) によるパスウェイデータを独自に編集したものです。「Contributed」と同様、分子間相互作用を表したパスウェイ図があります。

Tissue-specific: 特定の臓器で発現している遺伝子ごとに分類したものです。パスウェイ図は「GO samples」と同様、分子間の相互作用までは記されていません。

それでは、ここで「Mm_Contributed」の階層をたどっていきましょう。このカテゴリーの下には、さらに4つのカテゴリーがあったので、この中の「metabolic_process」を選択しました。

Pathway	Changed Genes	Total Genes	Z score	P-value
GenMAPP	105	8613	0	
Mm_Contributed_20070917	54	3174	2.642	
cellular_process-GenMAPP	30	1950	1.411	
metabolic_process-GenMAPP	12	529	2.282	
Mm_Acetylcholine_Synthesis	0	7	0	1
Mm_Amino_Acid_Metabolism	5	91	3.783	0.00539
Mm_Bioenergetic_Amine_Synthesis	0	14	0	1
Mm_Cholesterol_Biosynthesis	0	14	0	1
Mm_Eicosanoid_Synthesis	0	21	0	1
Mm_Electron_Transport_Chain	1	67	0.227	0.55445
Mm_Fatty_Acid_Beta_Oxidation_1_BIGCoAT	1	26	1.246	0.27403
Mm_Fatty_Acid_Beta_Oxidation_2_BIGCoAT	0	5	0	1
Mm_Fatty_Acid_Beta_Oxidation_3_BIGCoAT	0	7	0	1
Mm_Fatty_Acid_Beta_Oxidation_Meta_BIGCoAT	1	31	1.044	0.31591
Mm_Fatty_Acid_Omega_Oxidation_BIGCoAT	0	6	0	1
Mm_Fatty_Acid_Synthesis_BIGCoAT	0	21	0	1
Mm_Glucocorticoid_Mineralocorticoid_Metabolism	0	13	0	1
Mm_Glutathione_Metabolism_KEGG	0	16	0	1
Mm_Glycogen_Metabolism	0	21	0	1
Mm_Glycolysis_and_Gluconeogenesis	1	38	0.818	0.37055
Mm_Heme_Biosynthesis	0	9	0	1
Mm_Krebs-TCA_Cycle	2	28	2.902	0.04823

すると上図のように、代謝反応に関わるパスウェイが出てきました。そのなかでも特にZ-scoreが大きく、かつP-valueが小さいものに「Mm_Amino_Acid_Metabolism」(赤線で囲んだ領域)があります。なお、Z-scoreやP-valueの意味は、Gene Ontology解析のときと同じです。



パスウェイ図を見たい場合は、該当パスウェイ名を選択し、右クリックメニューの図表示(Z)を選択します。するとブラウザが起動し、パスウェイ図が表示されます。このように、分類されたカテゴリーから自分の興味のあるパスウェイを探し出し、そのパスウェイに対する解析データやパスウェイ図を調べることができます。

