

大規模データの解析における問題点 (2009年8月号掲載)

山田陸裕・上田泰己 (理化学研究所発生再生科学総合研究センター)



最近、DNA マイクロアレイを用いた解析結果を示す論文が多くみられます。自分は統計学的な考え方に不慣れなので、“ \times 倍の発現変動があった”という言葉が鵜呑みにしてしまいます。他人が解析した結果を解釈するうえで、どのようなところに気をつけてデータをみればよいのでしょうか。



平均値の変動は実験データのばらつきと比較して評価する必要があります。というのも、いつかは自説を支持してくれるデータがでてくれることを信じて何百回も同じ実験をくり返せば、そのうち1回くらいは望みの結果の得られることがあるかもしれませんが、それは偶然にすぎず、意味のある結果とは認められないからです。DNA マイクロアレイ実験ではいちどに数万個の遺伝子の発現量を測定するため、不特定多数の遺伝子が \times 倍の発現変動を示します。このような平均値の変動が実験データのばらつきに対してどのくらい意味のあるものなのかを評価するため、統計的な手法は必要不可欠です。このような点をふまえて、適切な統計手法を用いて実験結果を評価しているか、その際に有意性はどのくらいあるのか、有意判定された遺伝子についてどのような検証を行なっているのか、評価や検証の信頼性と主張とのあいだに解離がないか、これらに気をつけてデータをみていく必要があると思います。



DNA マイクロアレイを用いて2つのサンプルのあいだでの遺伝子発現量の比較を行なうと、発現量の低い遺伝子はしばしば陽性扱いにならないことがあると思います。ただ、実験の内容からその遺伝子の発現変動が期待される場合、有意水準を意図的にゆるめて陽性とするのは、そののちの検証実験でこれを正しいと示せば問題ないのでしょうか。また、この行為はデータを正確に解析したことになるのでしょうか。



実験結果によってその解析方法を意図的に変更することは望ましくありません。ただ、質問の状況では、陽性となることが期待される特定の(発現量の低い)候補遺伝子があり、その候補遺伝子が真に陽性かあるいは陰性かどうかを確認することが実験目的となっているように推察します。その場合は、DNA マイクロアレイ法のような包括的な実験手法ではなく、定量的PCR法など、対象をしばったより精度の高い実験手法を用いて、通常の統計検定法とその有意水準により判定するとよいでしょう。なお、DNA マイクロアレイを用いた実験では、実験内容から陽性となることが期待されている遺伝子のほかに、未知の陽性遺伝子を発見するスクリーニングを目的とする場合が多いと思います。有意水準をゆるめると陽性判定となる遺伝子の数が増えますが、同時に、偽陽性の割合も増えますので、のちの検証実験の負担が増します。記事で紹介したFDRは陽性判定のなかに含まれる偽陽性の割合を表わしますので、有意水準を調整する際のリスクを勘案するための有効な指標となります。



有意水準の設定の仕方とFDRとのあいだにトレードオフの関係があることは直感的にもよくわかりました。ただ、実際に実験を行なったとき、有意水準を厳しくしたにもかかわらずFDRが満足に小さくならないような場合は、実験の設計からやり直すのが最良の解決策なのでしょうか。例としてあげられた、DNA マイクロアレイ実験のサンプル数を増やすこと(これにも、金銭的な問題などかわると思いますが)での解決は現実的ではないことが多いのでしょうか。

A 場合によります。陽性対照となる遺伝子の発現データが平均値が数倍も変動するなどその発現変動が大きく実験間のばらつきが小さい場合には、DNA マイクロアレイの枚数を増やすことで改善がみられることがあります。一方で、平均値の変動が小さい場合や実験間のばらつきが大きい場合には、現在の実験スキームでは変化をとらえきいていない可能性があります。このような場合には、時系列をとるなど実験の設計からやりなおすのが最終的には近道である場合があります。いずれにせよ、陽性対照となる遺伝子の発現を定量的PCR法で確認しながら実験スキームの設計を進めることが鍵になります。もし、そのような遺伝子がない場合には、2回以上のDNA マイクロアレイ実験（1回目は陽性対照探しを目的とする実験、2回目は包括的な解析を目的とする実験）を組むことが必要となる場合があります。

Q 網羅的な解析を目的としてDNA マイクロアレイ実験を行なう場合は、同じ解析を3回以上行なわなくてはならない、と耳にしたことがあります。実験に大きく依存するとは思いますが、とりあえず行なうとしたら、1条件あたり何回（何枚）くらい行なえばよいと考えられますか。

A もちろん多ければ多いほど望ましいわけですが、1条件あたり最低2枚はないとばらつきの評価ができません。一方、たとえば、筆者らの研究室では、発現量が概日変動する遺伝子をスクリーニングする目的で、4時間ごとに1枚ずつ計12枚（48時間）のDNA マイクロアレイを使用して実験・解析を行なったこともあります。ただし、この場合は2日間（48時間）の実験になりますので、実質的には各時点で2枚ずつのDNA マイクロアレイ実験を行なったのと同等です。

Q p 値を0.01とした場合、10,000回の検定を行なえば偶然だけでは得られないデータが100回“も”得られるということ、はたして（一方向的に）多いといい切つてよいものなのでしょうか。偽陽性（である確率が、とても高い）が多くはなるでしょうし、なにかの出荷製品の検査であれば大問題ですが、10,000回行なったうちの100回というのは、母集団の大きさをみればたいしたことないように思えてしまいます。

A 比較は、母集団に対してではなく、そこに含まれる真に陽性の遺伝子に対して行なうべきだと思います。というのも、たとえば、偽陽性のものが100個あるのに対して真に陽性のものが1個しかない場合、真に陽性のものを1個得るための検定に必要な労力は膨大なものになります。一方で、偽陽性のものが100個あるのに対して真に陽性のものが900個ある場合は、真に陽性のものを1個得るために必要な労力は前者に比べて格段に少なくなります。このため、記事でも示したように、陽性判定に含まれる偽陽性の推定割合を示す指標であるFDRを用いるのが重要です。

Q Bonferroni補正について、偽陽性の出現頻度を高めに設定したり、片側検定にしたりすることは、DNA マイクロアレイ実験のように検定数の多い場合は有意水準の厳しさを緩和することにはならないのでしょうか。また、この行為は統計学上、まちがいではないのでしょうか。

A 統計的な指標はわれわれがデータを比較するための尺度をあたえてくれますが、有意水準をどこにとるかはその尺度を使うわれわれ自身が決める必要があります。かつては、 p 値で0.05あるいは0.01を有意水準として用いることが一般的でしたが、記事で例としたDNA マイクロアレイ解析のように、多重検定が必要となる場合には一般的な有意水準（閾値）があるわけではありませんので、実験者あるいは解析者が自分自身で十分に納得することのできるレベルに設定することが重要です（記事で述べた、より客観的に解釈しやすいFDRという指標も、閾値をどこに設けるかという意味では同様の問題をかかえています）。十分に納得することができるレベルというのは、やみくもに自説に都合のよいように調整するのは異なります。この閾値をどのように設定するかについては主観的な要素がどうしても入ってしまうものですので、論文で主張したい説を支持する重要なデータについては、定量的PCR法などの別法で検証することが欠かせません。