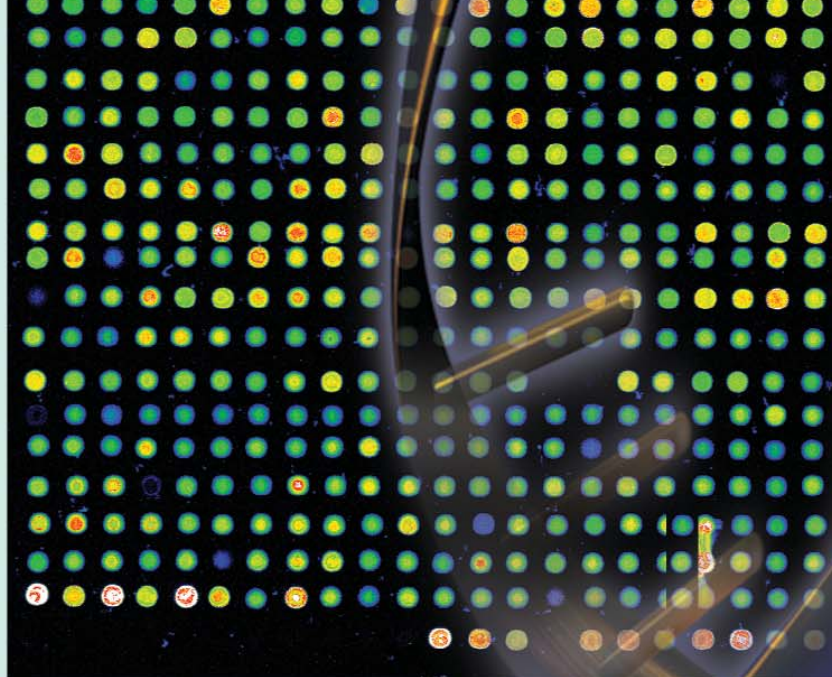


A troubleshooting guide:
Microarray experts share
their advice on handling
key aspects of validation

Microarray Confirmation and Quality Control

&A:

Optimize your microarray assays



You **rely on IDT** for top quality oligos, so **trust IDT** with your microarrays. We offer an assortment of products to support researchers who print their own microarrays.

Epoxide Spotting Buffer (ESB)

- Optimized for spotting amine-modified or I-Linker™ modified oligos on epoxy surface slides
- Designed to reduce evaporation and deliver ~85 micron spots with uniform hybridization signals
- Suitable for both short and overnight microarray print runs

SpotQC

Detector oligos have improved thermodynamics allowing greater hybridization uniformity over a wider variety of sequence compositions.

- Cy3™-labeled detector oligo developed for direct hybridization fluorescence detection.
- Biotin-labeled detector oligo developed for secondary fluorescent, chemiluminescent, or colorimetric hybridization detection assays.

SpotQC Hybridization Buffer

Optimized hybridization solution specifically formulated for use with IDT's SpotQC oligos. It allows for the dilution of the SpotQC detector oligo up to 20fold while maintaining the superior detection of a broader range of oligo-probe lengths and GC compositions with our proprietary detector oligo.

Learn more at www.idtdna.com/microarrays

Innovation and Precision in Nucleic Acid Synthesis

www.idtdna.com

800.328.2661

IDT
INTEGRATED DNA
TECHNOLOGIES, INC.

Corporate Headquarters
Integrated DNA Technologies, Inc.
1710 Commercial Park
Coralville, IA 52241 USA

West Coast Operations
Integrated DNA Technologies, Inc.
6868 Nancy Ridge Drive
San Diego, CA 92121 USA





Table of contents

Letter from the Editor	5
Index of Experts	5
Q1: What do you consider when choosing and applying validation techniques to confirm your results?	6
Q2: What normalization techniques do you use? How do you confirm normalized data?	9
Q3: How do you determine the number of replicates needed to achieve sufficient sensitivity?	10
Q4: What methods do you use to identify differentially expressed genes in replicate experiments?	12
Q5: What techniques do you use for replacing missing data or identifying poor quality spots?	13
Q6: How do you compare array data across different platforms?	14
List of Resources	18

Enabling Better Decisions



Quality & Experience:

- First GLP Compliant Microarray Laboratory
- CLIA-registered for clinical trials
- Ph.D. Staff with over 30 years experience in microarrays

Comprehensive Services:

- RNA Expression: Genome & Exon
- Genotyping: High Density & Focused
- Isolation: Expertise in RNA, DNA and Blood Protocols

Innovation:

- Experienced in FDA submissions
- Leader in standardization efforts
- Novel analysis tools



Licensed Affymetrix Service Provider



Comprehensive Genomic Services

www.expressionanalysis.com

Toll-free 1.866.293.6094

The single best **multiplex**
solution for **quantitative**,
high-throughput, cost-effective
gene **expression** analysis.

**SINGLE SOURCE
SAMPLE TO ANSWER IN 8 HOURS**



9:00 AM
RNA sample preparation
and quant/norm



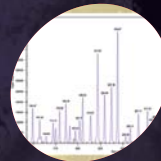
11:30 AM
Multiplex reverse transcription setup
on the Biomek® 3000 liquid handler



1:00 PM
Multiplex PCR setup with
GeXP Start Kit or gene set kit



3:30 PM
Gene expression analysis
with GenomeLab GeXP



5:00 PM
Quantitative gene
expression results

Reduce bottlenecks in your daily workflow.

Finally, quantitative gene expression that's cost-effective too. The GenomeLab™ GeXP Genetic Analysis system utilizes a patented, highly multiplexed PCR* approach to quickly and efficiently look at the expression of gene sets with greater sensitivity and speed. Visit www.beckmancoulter.com/GeXP



Genomics Proteomics Cell Analysis Centrifugation Lab Tools Particle Characterization Bioseparation Lab Automation

Simplify · Automate · Innovate

*The PCR process is covered by patents owned by Roche Molecular Systems, Inc. and F. Hoffmann-La Roche, Ltd.
© 2006 Beckman Coulter, Inc.

Letter from the editor



We are pleased to present *Genome Technology's* second technical reference guide focusing on arrays. This volume, which is designed to complement our previous guide on microarray sample prep, brings together the well-formed thoughts of several experts as they evaluate various confirmation and validation methods.

When it comes to microarray experiments, quality control is vital. Scrupulous confirmation and data analysis measures are prerequisites for generating results that reliably mirror what is going on at the transcript level. The problem is that there are many complexities to consider in any array experiment. This makes the validation of array data — at each step of experimentation — of prime significance to both the array community as a whole and to individual investigators.

Confirmation methods such as real-time PCR assays, northern blots, or *in situ* hybridization are all useful for verifying results, depending on the type of experiment being run. Yet verification, while necessary and important, is only one of several processes used to analyze experimental results. Other data analysis quality control techniques, both conceptual and statistical, are often pulled into the effort to objectively evaluate array performance.

In this guide, you'll find advice on the importance of confirming results, choosing and applying normalization techniques, including replicates and identifying differentially expressed genes, as well as the comparison of array data across different platforms. We've also added a resource guide of data analysis tools and publications to keep on hand if you're hungry for more.

— Jennifer Crebs

Index of experts

Genome Technology would like to thank the following contributors for taking the time to respond to the questions in this tech guide.



Gary Churchill
The Jackson Laboratory



John Quackenbush
Dana-Farber Cancer Institute



Mostafa Ronaghi
Stanford University



Marc Salit
National Institute of
Standards and Technology



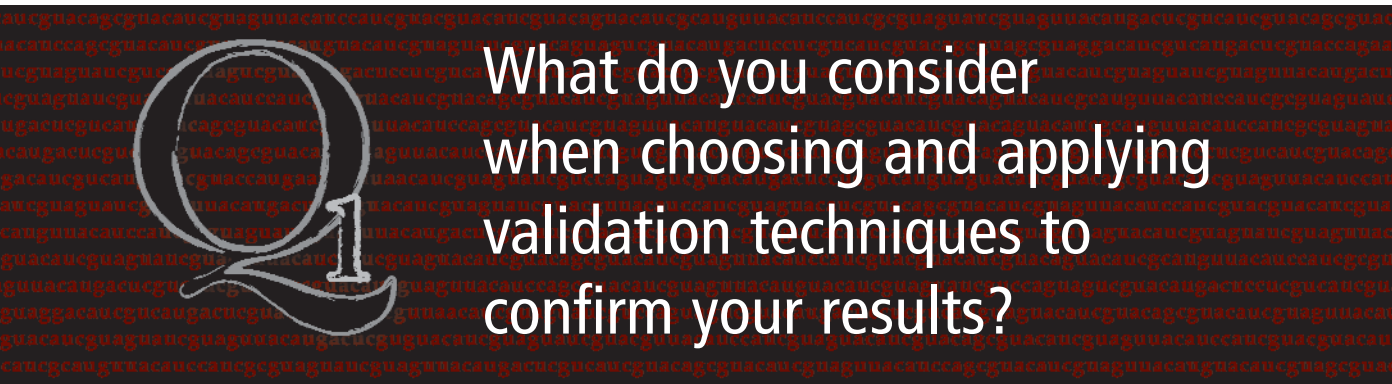
Chris Stoeckert
University of Pennsylvania
Penn Center for Bioinformatics



Lisa White
Baylor College of Medicine



Joshua Yuan
University of Tennessee
UTIA Genomics Hub



What do you consider when choosing and applying validation techniques to confirm your results?

As microarray technology continues to improve, the need for independent validation will certainly decrease, but we will probably always feel compelled to validate the most surprising or important findings. When choosing a method of validation, we consider what might have gone wrong to produce a misleading result. If a particular measurement is suspect, it may be a localized problem such as dust on an array element. Replication and good quality control practices will usually be sufficient to catch these errors.

More often we may question whether the particular probe or probe set is measuring what we think it should be measuring. In this case, validation with an assay that utilizes different sequence features of the transcript is required. The most popular choice is quantitative real-time PCR. When the same RNA sample is assayed, microarray and qRT-PCR results can be highly concordant quantitatively. When the results are discordant, we question the PCR result as much as we do the array.

Different approaches are required when an entire sample may have been compromised. Independent biological replication and good experimental design practices are the best safeguards. Microarrays can be remarkably good at detecting inconsistencies in experimental procedures. Even minor differences in the handling of animals can alter the expression of large numbers of transcripts. In one case, a 'fasted' animal apparently found bits of food in the bedding. Telltale expression of lipases and proteases in the liver clued us in to the problem. Repetition of an entire experiment with new samples on a different array platform is an ideal global validation strategy.

Microarray studies can provide their own internal validation when the pattern of changes in many genes is consistent with prior biological knowledge. The risk is that we may disregard a novel result when it isn't consistent. If something looks unusual, it could pay off to check it out.

— Gary Churchill

1. The purpose of the experiment
2. The purpose of the validation analysis
3. The availability of samples
4. The overall scope of the validation required
5. The cost of the assay

— John Quackenbush

We generally apply two validation approaches: the first being a verification of the direction and magnitude of change using RT-PCR or Molecular Inversion Probe assay, and then an informatics approach to compare our results to other data both internal to the group and published.

We usually use NextBio system for the informatics approach. NextBio has created a great atlas on every single tissue and best validation I would get is to find relevance between the data we have produced and the one other has produced. With that system we can easily compare orthogonal datasets to validate study results.

— Mostafa Ronaghi

The definition of validation is a useful place to start. ISO 9000 defines validation as being "confirmation through examination of a given item and provision of objective evidence that it fulfills the requirement for a stated, intended use." That's sort of

measurement science or accreditation or quality systems gobbledygook. But what stands out is this idea of the "provision of objective evidence" — that the hit a microarray has called is something worth deeper investigation or is scientifically — biologically — meaningful.

So what you're trying to do when you're trying to validate is you're trying to say, 'Okay, I've got this hit — now how can I get further evidence to increase the confidence in the measurement?' Confidence in measurements often is achieved by repeating the measurement with an approach or an assay that uses different principles.

[This is] why people use PCR to validate microarrays. If you use two different assays that give you similar or congruent results, those may in fact be compatible results indicating that a particular gene is being differentially expressed in this pathway.

I think it's important to set out to say, 'What is it I'm trying to do?' and not waste resources trying to validate things that are potentially of marginal interest. Especially because going from a highly multiplexed assay, like a microarray, into a PCR assay — where you're doing things more or less one at a time — can be very expensive. In other words, you don't want to validate a couple hundred hits. Doing pre-screening of the hits that seem to be most strongly interesting makes a lot of sense.

— *Marc Salit*

As a bioinformaticist no longer working in the lab, I look for multiple lines of supporting evidence. As a start, we may use multiple analyses to identify results with overlapping agreement. For differentially expressed genes, we may get real-time quantitative

PCR. Or we may look for alternate computational data sets for support such as phyletic profiles.

— *Chris Stoeckert*

For validation of the platform itself I would consider possibly using another microarray platform to confirm (e.g., first platform spotted glass long oligonucleotide microarray like Agilent confirmed by a second platform with short oligonucleotides like Affymetrix). Validation by qPCR to confirm the platform would probably consist of random selection of genes showing change and no change.

For validation of my experiment on a microarray platform I would select genes of interest that show change due to my experimental treatment. These genes could be validated using qPCR or in some cases by *in situ* hybridization.

One particular factor that should probably be addressed is what you validate. It is important to realize what the assay you are validating (e.g., microarray) is actually assaying especially when designing qPCR primers. If the microarray only assays the extreme 3' end of the transcript and you design the primers for the middle there may be problems with correlation of the two data types.

— *Lisa White*

Real-time PCR, northern blot; considering the limitation of microarray experiments, I consider validation an important part of the experiments.

— *Joshua Yuan*

Digital Expression Profiling

Rapid, Economical, Accurate



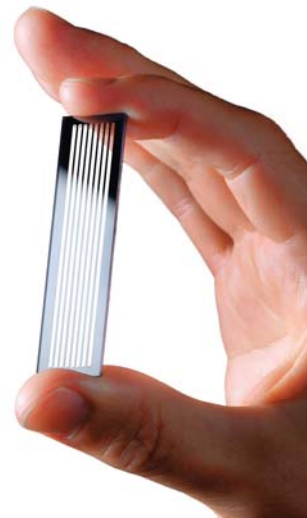
With Solexa's digital expression profiling system you can rapidly and affordably sequence millions of mRNA or miRNA tags per sample

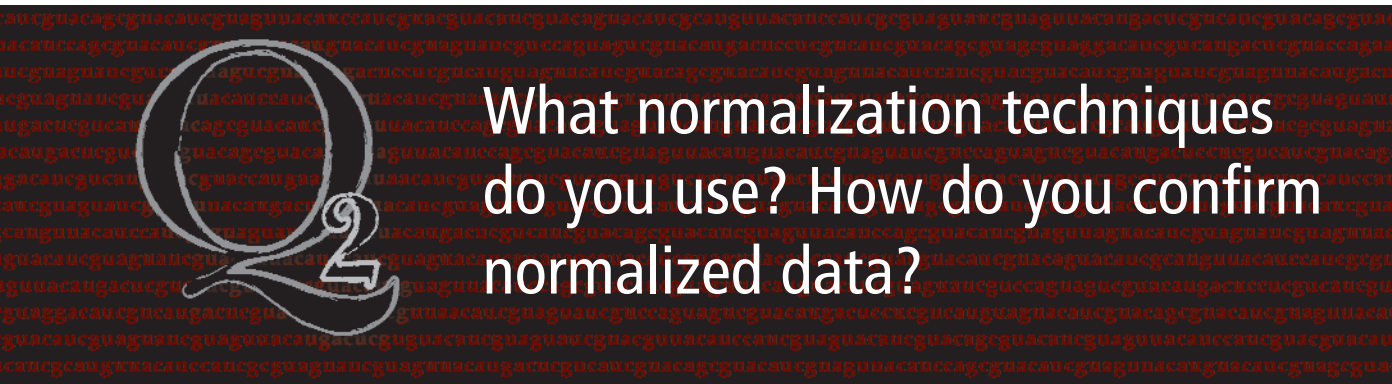
- Simultaneously analyze up to eight samples.
- Generate genome-wide, sequence-based digital expression profiles from any gene, in any organism, without the need for probes.
- Discover novel content in well-annotated genomes.
- Obtain orthologous, whole-genome profiles of expression at costs comparable to hybridization arrays.
- Discover and analyze global profiles of small RNA in any species, rapidly and economically.



Learn more about digital expression profiling from Solexa.

Visit www.solexa.com/compare to download a comparison of expression profiling techniques.





What normalization techniques do you use? How do you confirm normalized data?

Microarray data should be analyzed on a logarithmic scale because the effects of most interest are approximately multiplicative. Mean centering and adjustment for batch effects and other design factors are best done using a linear model-based analysis. These are fairly innocuous normalizations. Additional transformations appear to be required in some cases but the reasons are poorly understood and it is not clear which approaches are best.

When it comes to transforming data, simple is better. Methods that are theoretically sound, empirically validated, and carefully compared to the best competing methods are a must. The best way to normalize data remains an open question and the answer changes with every technical innovation.

A common problem with two-color array data can be diagnosed using the MA-plot. The "hockey stick" shape can be caused by different additive backgrounds in the two dye channels. Direct subtraction of estimated background is almost always a bad idea. At best it will add noise to data and in many cases it will attenuate the signal as well. A LOWESS transformation will correct the curvature and is the best way to deal with background in two-color arrays. Post-LOWESS, the MA-plot will always be flat and zero centered, even when it shouldn't be.

With single-color systems, we use quantile normalization as a means to adjust for uneven intensity distributions across different arrays. All the above caveats apply. When samples have different ranges of expression, quantile normalization can be problematic. The same is true when expression changes are unilateral in direction.

It is always good practice to scrutinize your data. Some problems can be seen easily and other not so

easily. Look at raw images, scatterplots, and MA-plots before and after transformation. Plot log ratios using a color scale on the array coordinates. If there is significant spatial variation, streaks, or bubbles on the array image, throw it out and replace it. Some things are just not worth the price of fixing.

— Gary Churchill

For two-color arrays, a local LOWESS (we use the implementation in MIDAS). For single-color arrays, we use either RMA or dChip.

— John Quackenbush

We mainly use Affymetrix gene expression platforms and dChip has been the method of choice for normalization in our lab. All of the normalized data are then analyzed by using model-based expression analysis to generate the expression values with the dChip software.

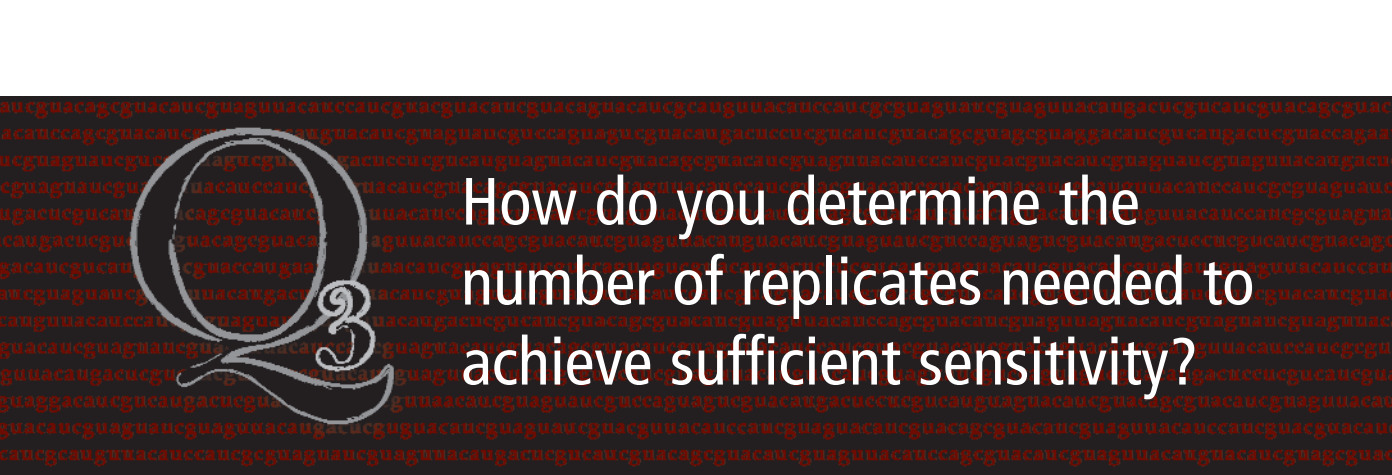
— Mostafa Ronaghi

One should be careful that they're not masking real effects by normalizing. I'm generally uncomfortable with the non-parametric, rank order-based normalization approaches that are ubiquitous in microarray science. [However,] I would hesitate to jump in on any normalization recommendations at this point.

— Marc Salit

For two-channel arrays, we typically look at the MA-plots and perform LOESS (or print-tip LOESS if the MA-plots indicate problems there). LOWESS assumes balanced differences between channels at each intensity so for very

(continued on p.16)



How do you determine the number of replicates needed to achieve sufficient sensitivity?

Cost is always a factor in determining the size of an experiment. The high cost of microarrays has resulted in experiments that are so small that they would be unacceptable in almost any other context. Independent replication within an experiment provides at least three benefits. It increases the precision of estimation. It provides a means to detect sample mix up and contamination. Most importantly, it provides a basis for estimating the degree of error in the measurements and thereby a means for making statistical inference.

As a simple guideline for sample size, I consider the measurements from one single gene and work out the ANOVA table for the proposed experimental design. Having enumerated the potential sources of variation and their associated degrees of freedom, one can subtract these from the number of data points to arrive at the residual degrees of freedom (df).

An ideal experiment will have 10 to 20 residual df. Much more than 20 df is wasteful. Fewer than 10 df and the experiment is likely to be underpowered. With microarrays, as few as four or five residual df may still be acceptable. An experiment with too few residual df is subject to random fluctuations that are familiar to anyone who has applied a t-test to a small microarray experiment. This happens because there is insufficient information to estimate the denominator of the test statistic.

Perhaps the most common microarray experiments are two condition comparisons. With three independent replicates per group, there are four residual df. With four replicates, we are up to six df. At 10 replicates per group, there are 18 df. This may seem like a lot of replicates; the somewhat

surprising reason for this is that the experiment is too simple. Substantial gains come from factorial experiments that examine more than one factor at a time. With fewer replicates per condition and little or no increase in the number of arrays, both the scope and power of an experiment can be increased. For example, an experiment with two sexes, two diets, two strains, and 3x replication per group requires 24 arrays. After accounting for main effects and interactions there are 16 residual df. The experiment is more powerful than three two-condition experiments at 4x replications and uses the same number of arrays. Moreover, one can ask questions about interactions.

— Gary Churchill

This depends on the purpose of the experiment and the strategy for validation. The most widely used power calculation in the world, however, is available dollars divided by the cost per assay.

— John Quackenbush

This pretty much depends on the project and the budget. For the projects in which we are limited with the amount of sample there would be no replicate study. If we have larger sample on some of individuals then we would try to replicate 10 percent of the data. Ideally, if we wouldn't be limited with budget and sample, I still would replicate all the study despite the fact that we have very good process for expression profiling.

— Mostafa Ronaghi

(continued on p.16)

val' i · da' tion



It's a synonym for TaqMan® Gene Expression Assays.

Your single easiest solution

TaqMan® Gene Expression Assays deliver accurate real-time PCR results to validate your microarray discoveries. TaqMan® Assays are designed to run under universal thermal cycling conditions and are formulated into a single 20X solution—less pipetting means less chance of error.

**NOW
AVAILABLE**

Arabidopsis and
Drosophila
Assays!

>600,000 assays, and all very affordable

Virtually every gene and every transcript for human, mouse, rat, Arabidopsis and Drosophila is covered—with no time, reagents or effort spent on design and optimization, you save money on every experiment.

The gold standard in quantitative gene expression analysis

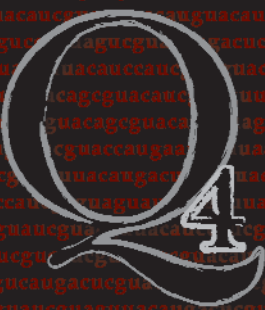
TaqMan Assays provide unmatched sensitivity, specificity and reliability for true gene expression validation—you can publish your data with confidence.

To learn more about TaqMan Gene Expression Assays and how they can help you validate your research, visit www.allgenes.com.

**Applied
Biosystems**



For Research Use Only. Not for use in diagnostic procedures. The PCR process and 5' nuclease process are covered by patent owned by Roche Molecular Systems, Inc. and F. Hoffmann-La Roche Ltd, and by patents owned by or licensed to Applied Biosystems. Further information on purchasing licenses may be obtained from the Director of Licensing, Applied Biosystems, 850 Lincoln Centre Drive, Foster City, California 94404, USA. Applied Biosystems is a registered trademark and AB (Design) is a trademark of Applied Biosystems or its subsidiaries in the US and/or certain other countries. TaqMan is a registered trademark of Roche Molecular Systems, Inc. © 2006 Applied Biosystems. All rights reserved.



What methods do you use to identify differentially expressed genes in replicate experiments?

If you observe a large fold change in an interesting gene, check it out. There is no point letting statistics get in the way of science. But eventually you will have to validate your findings with a statically sound experiment. Most of us will want to use a statistical criterion from the start and the standard t- or F-test are good choices. These tests formally require normality and constant variance across groups but in balanced designs they are quite robust.

An added degree of robustness is provided by using permutations to estimate significance. Small experiments limit the number of unique permutations but the permuted test statistics can be pooled across genes. It is important to first remove the differentially expressed genes before running the permutation analysis. It sounds circular, but if you apply a t-test with a liberal cutoff (0.1 alpha-level from the standard t-distribution) and leave these genes out of the permutations, the pooled statistics provide a valid and robust permutation based p-value. Use at least 1,000 permutations even when pooling the results.

If the residual degrees of freedom are more than 10, the t- or F-test applied one gene at a time is probably all you need. This approach allows each gene to have its own individual variance. This is a biological reality and ignoring it by using a test that assumes a common variance for all genes will lead to erroneous conclusions. For smaller experiments we can take advantage of the fact that microarrays measure many genes simultaneously. The estimated variance for each individual gene has two sources of variability, one is biological variation and the other is statistical estimation error. The statistical component can be reduced by "shrinking" the variance

estimates. Several forms of empirical Bayes tests have been proposed and they all seem to work equally well. For experiments with fewer than 10 residual df, an empirical Bayes test using the shrunken variance estimates can dramatically improve sensitivity.

Multiple test correction has become a mantra and the method of choice in the microarray world is the false discovery rate (FDR). FDR has a simple interpretation and it is generally the right choice for list generation but FDR estimation can be finicky. Make a histogram of the unadjusted p-values. If it doesn't look perfect, something has gone wrong. An unanticipated correlation in the data, perhaps an effect of normalization, is often the cause. At what FDR significance level should you cut off the list? There is no rule. It is up to you to choose but consider that gene lists are just a starting point for downstream analysis and interpretation. Try making several different cuts and follow through with the pathway, GO term, or clustering analysis. The results can be surprisingly different and the different lists may be telling you about different aspects of the biology.

— Gary Churchill

This depends on the goal of the experiment and its design. We typically use a variety of statistical approaches, including t-tests, ANOVA, and SAM.

— John Quackenbush

In the old days, we used to perform RT-PCR or occasionally we used an in-house developed method based on molecular inversion probe assay. In our lab we have been trying to set *(continued on p.16)*



What techniques do you use for replacing missing data or identifying poor quality spots?

Missing data are less of a problem than they have been in the past. An ideal solution would be to implement a statistical missing data algorithm such as multiple imputation or EM. Specifying a model for imputation may be problematic and the computation will also be daunting. To make matters worse you will probably need to run a permutation analysis. In practice we have simply removed whole genes that have any missing data from the analysis.

A quick and dirty solution is to make a single imputation using the average of replicate samples of the same condition, if you have them. If you make up data, keep track of it, and don't let it show up in your published results. If an array generates too many missing data points, replace it.

An important point to make here is that low intensity data, even if it is below the background signal level, is not missing. It is low intensity and can be highly informative. Removing low intensity data points from an analysis is a sure recipe for missing the most significant, all-or-nothing, changes in gene expression.

— Gary Churchill

Missing data? The best replacement is a new assay. We generally treat missing data as missing. If data for a particular gene are missing from too many of the assays, we consider the gene to be uninformative in the experiment.

We continue to try to develop a robust quantitative measure of probe quality but have not settled on a definitive metric.

— John Quackenbush

We don't have missing spots when we use Affymetrix chips. The chip manufacturing is pretty standard now with good quality-control steps. Sometimes the hybridization quality is low though. Since we have a lot of leftover from the expression sample, we would re-hybridize the chip to generate the data. However, these things happen rarely these days in our hands.

— Mostafa Ronaghi

I think there is good work in the image processing community that has long predated the microarray community for doing feature-extraction and doing things like characterizing spot morphology.

What people are using in practice is probably not unreasonable, which is probably [using] the default settings on their scanner software or on the software that they're using to do image feature extraction.

We don't necessarily need new science for that, people just need to use good practice. I think the microarray science community has lots to learn from those in the image processing community who have plowed these fields before us. We already have been leveraging the excellent work that's well established in that field.

As far as replacing missing data, there is no way to create data where none exists! Good practice for a missing value usually comes down to doing typical kinds of things: using medians [and] working with good analysis-of-variance software that can handle unbalanced models.

— Marc Salit

(continued on p.16)



How do you compare array data across different platforms?

Concordance is the best you can do in the absence of a truth standard. Spike-in experiments feel artificial and may not accurately reflect the behavior of real samples. We see a high degree of concordance among most viable expression technologies including spotted, short oligo, and bead array platforms. When results are concordant you can be sure that if something is wrong, everything is wrong in the same way. When discrepancies occur, you can try validation with a qRT-PCR assay. But it still comes down to concordance and majority rule.

In a recent comparison of two commercial platforms we followed up on the handful of discordant data points by mapping the probes back onto the current build of the mouse genome. Many of these pairs of probe mapped to different genomic locations. Thus, although errors are rare, the accuracy of probe annotation appears to be a major factor in reliability of microarray results.

It is good to have access to multiple platforms. In addition to impetus for ever improving tools that is driven by competition, our confidence in results is bolstered when different platforms agree. We know that we are measuring something consistently. Nonetheless, a dose of skepticism is healthy and we should always question just exactly what we are measuring on any microarray platform.

— *Gary Churchill*

Carefully.

It depends on whether we are comparing our own samples across platforms or if we are looking at published data. The first is easier and we published an approach showing that careful analysis using consistent techniques provides good correlation

across platforms. However, in looking at published data, it is difficult to tell what the data really represent and whether the sample classes are accurately described.

To make comparisons across platforms, we first construct a linking table based on RESOURCERER.

— *John Quackenbush*

We usually use NextBio system. NextBio has developed a huge atlas on different tissue in different species and diseases. We start with importing our dataset into the NextBio system and then immediately we can compare our dataset with large number of studies available there. The cool thing here is that suddenly we can put our work in the context of what other people have done. We have found extremely exciting correlations among different diseases. Sometimes data has been generated from SNP genotyping on selected genes and we could immediately validate our expression data with allele frequency of different genes. This kind of data integration is of utmost importance in genetic research today.

— *Mostafa Ronaghi*

I think the science for understanding how to compare microarray results across different microarray platforms is still immature. All of the studies I've seen show that microarray cross-platform comparisons of differentially expressed genes agree anywhere between 40 percent and 60 percent. The real question is whether this is "fit for purpose," or acceptable for the application.

Certainly, if you leave out all of the false negatives, you can get much better agreement. In

other words, you get better agreement if you only consider concordance between two platforms as being between those results that have significant signal detected on both platforms. Some studies have reported up to 90 percent concordance, but they didn't count genes detected on one platform but not the other in the denominator.

At this point, my recommendation is to use caution in interpreting cross-platform comparisons. The field has been filled with them. The seminal comparison done by Maggie Cam at the NIDDK (Tan et al., 2003) was the paper that launched a thousand microarray studies. The variety of designs and styles, making it difficult to compare across the comparisons! Looking at a single study and declaring "Victory!" is likely to be naïve, or even disingenuous.

There is a great interest in understanding repeatability and reproducibility of array measurements. Scientists want to understand how much confidence to put in their array results, and observing and understanding these properties can lead to well-placed confidence. However, I have yet to see a solid, theoretical, quantitative treatment of the comparison of results.

What is reasonable, which is coming into practice, is to use one DNA microarray platform to confirm or corroborate another DNA microarray platform; where both platforms agree, you've certainly got a lot of evidence that something's going on. But where platforms disagree, what you don't have is convincing evidence that nothing is going on.

As the field matures, the next set of questions

will be 'Why are these results like this?' and 'What should they be like?' These are two different ways of looking at the same marble.

— *Marc Salit*

Comparing data requires normalization to put data on an equal footing. For different platforms this generally means transforming probe set intensities or log ratios to a common metric. One approach we favor is to analyze data on the same platform first to generate calls and confidence scores and then compare those analysis results across platforms. With p-values, one can combine them (i.e. multiply them as in Fischer) if the assays are independent.

— *Chris Stoeckert*

Comparing data across platforms is not as simple as it appears. Each platform that you intend to use should have outstanding annotation for each of the genes being assayed. Identifying the shared content across platforms relies entirely on the annotation available for them. Once shared content is identified, data type should be taken into account. Microarray data is expression in relative terms, unlike qPCR. There may not be a direct correlation between the intensity values determined for microarray data and the absolute expression values from qPCR or other platforms.

— *Lisa White*

We don't compare different platforms; we verify with a second technology. Right now, most of our array data can be verified by real-time PCR. We use long-oligo arrays.

— *Joshua Yuan*

Q2: What normalization techniques do you use? How do you confirm normalized data?

(continued from p.9)

different samples we try to include a set of control features (spots) that can be used to generate the curves. For Affymetrix, we've been using gc-RMA which has performed well in recent studies comparing approaches.

We find hierarchical clustering to be very useful in assessing quality of the experiment. Replicates should cluster together. Normalized data should have high correlations with un-normalized input.

— *Chris Stoeckert*

Bulk and LOWESS. We confirm by MA-plot.

— *Joshua Yuan*

Q3: How do you determine the number of replicates needed to achieve sufficient sensitivity?

(continued from p.10)

Ideally, a pilot experiment should be done to determine the level of overall variability (biological and technical). Greater variability requires more replicates. If technical variability is small relative to biological variability, I would advise just doing biological replicates. Another consideration is that methods using permutations to estimate background distributions don't work very well with only three replicates. Five replicates is a good starting point, though two replicates is still minimally useful.

— *Chris Stoeckert*

We generally use three or more biological replicates, which is acceptable by the community. Statisticians always suggest more, and we understand that the more replicates, the easier to sort the noise out. However, the array experiments are very expensive and we've got to be realistic. Moreover, we can increase replicates at validation step.

— *Joshua Yuan*

Q4: How do you identify genes in replicate experiments?

(continued from p.12)

up a very precise protocol to generate reproducible data. In order to do this we had to minimize contamination sources, calibrate pipettes, and perform the assay and hybridization within the same day. The protocol is now published in www.gluegrant.org website. Again I would check the fold change of specific genes of interest in NextBio database.

— *Mostafa Ronaghi*

We'll have better answers to some of this, I think, in 18 months to two years after work that's underway in several labs, including ours, is ready to go. These are the great questions to ask: 'How do you determine what is a differentially expressed gene?' and 'How do you compare two gene lists?'

— *Marc Salit*

We use PaGE (Grant et al, 2005), which is like SAM, only better. We often use SAM (Tusher et al, 2001) for comparison and LODS method (Lonnstedt and Speed, 2002) when there are few replicates.

— *Chris Stoeckert*

Q5: How do you replace missing data or identify poor quality spots?

(continued from p.13)

Generally we don't replace missing data. With sufficient replicates, we can look for outliers (using PaGE) and remove them. We also use clustering and visual inspection to see if certain hybridizations should be discarded.

— *Chris Stoeckert*

There are various methods available for replacing missing data. In my opinion ignoring the entries containing missing values, replacing missing values by zeros or imputing missing values of row averages or medians doesn't work very well. Our statisticians generally use a Bioconductor R package for this called *impute* which utilizes a k-nearest neighbor imputation method.

— *Lisa White*

Sequenom

brings you the solution for

Quantitative Gene Expression Analysis

- ✓ PRECISE and ACCURATE direct measurement
- ✓ MULTIPLEXING up to 24 transcripts
- ✓ SCALABLE throughput
- ✓ Highly SENSITIVE and SPECIFIC



Fast-forward your Quantitative Gene Expression Research!

- Solutions for your microarray follow-up and validation
- Ideally suited for studies up to a few hundred target genes over multiple samples
- FLEXIBLE assay design and rapid turn around time
- No fluorescent labeled primers
- VERSATILE platform that also supports SNP Genotyping and DNA Methylation analysis
- Gene Expression analysis services now available

www.sequenom.com

SEQUENOM®

List of resources

There are a number of Web resources and publications germane to microarray confirmation and validation. In addition to our experts' recommendations, we have rounded up a selection of online tools and books to ensure necessary and sufficient array results.

Articles

Lonnstedt I. and Speed T. (2002)

Replicated microarray data.

Statistica Sinica 12, 31-46.

Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, Howe EA, Li J, Thiagarajan M, White JA, Quackenbush J. (2006)

TM4 Microarray Software Suite.

Methods Enzymol. 411:134-93.

Tan PK, Downey TJ, Spitznagel EL Jr, Xu P, Fu D, Dimitrov DS, Lempicki RA, Raaka BM, Cam MC. (2003)

Evaluation of gene expression measurements from commercial microarray platforms.

Nucleic Acids Res. 31(19):5676-84.

Tsai J, Sultana R, Lee Y, Pertea G, Karamycheva S, Antonescu V, Cho J, Parvizi B, Cheung F, Quackenbush J. (2001)

RESOURCERER: a database for annotating and linking microarray resources within and across species.

Genome Biol. 2(11):SOFTWARE0002.

Tusher VG, Tibshirani R, Chu G. (2001)

Significance analysis of microarrays applied to the ionizing radiation response.

Proc Natl Acad Sci USA. 98(9):5116-21.

Books

Discovering Genomics, Proteomics, and Bioinformatics

by A. Malcolm Campbell, Laurie J. Heyer
(September 2002) Benjamin Cummings;
ISBN: 0805347224

DNA Microarrays and Gene Expression

by Pierre Baldi, Wesley G. Hatfield
(October 2002) Cambridge University Press;
ISBN: 0521800226

Microarrays and Cancer Research

by Janet A. Warrington, Randy Todd, David Wong
(June 2002) Eaton Pub Co; ISBN: 1881299511

Microarray Quality Control

by Wei Zhang, Ilya Shmulevich, Jaako Astola
(April 2004) John Wiley & Sons; ISBN: 0471453447

Online tools

The Fitness for Purpose of Analytical Methods

<http://www.eurachem.ul.pt/guides/valid.pdf>

Inflammation and the Host Response to Injury

<http://www.gluegrant.org>

MIDAS: Microarray Data Analysis System

<http://www.tm4.org/midas.html>

MADAM: Microarray Data Manager

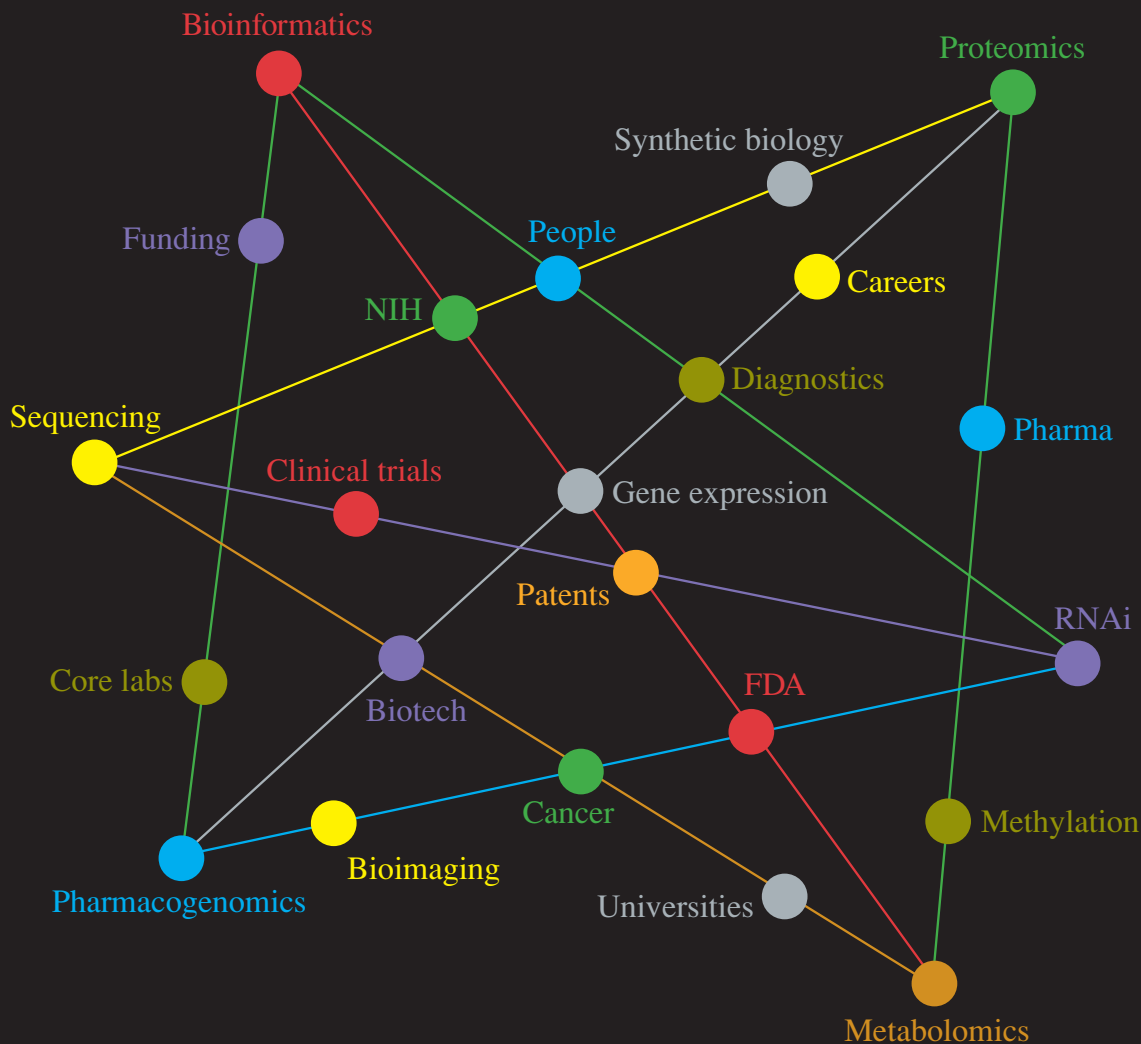
<http://www.tm4.org/madam.html>

NextBio System

<http://www.nextbio.com/index.html>

RESOURCERER 12.0 (July 2005 release)

<http://pga.tigr.org/tigr-scripts/magic/r1.pl>



Connecting the dots
for scientists worldwide.



illumina Gene Expression. It's time to buy a different vowel.

You have a choice in gene expression.
Switch to “i” and:

- ▶ Spend just \$160 per whole-genome array
- ▶ Process multiple whole-genome arrays per BeadChip
- ▶ Use 1/20th the amount of sample
- ▶ Get 100% oligo and array quality control
- ▶ Profile partially degraded RNA (FFPE samples) using the DASL[®] solution

Join the growing Illumina gene expression user community and enjoy the support of over 500 Illumina employees who are committed to your success.

Visit www.switchtoi.com
and find out why it makes
sense to switch to “i”.

www.switchtoi.com

