

NEWS AND VIEWS

Weighing our measures of gene expression

John Quackenbush

Dana-Farber Cancer Institute, Boston, MA, USA

Molecular Systems Biology 14 November 2006; doi:10.1038/msb4100096

On a hot day in almost any city in the world, if you ask someone what the temperature is, you might not be surprised to hear them answer '35'. We all know what this means—the temperature is 35°C and the report is a product of the environment and the instrument we use to measure its properties. If, however, you were to ask the same question on an equally hot day in a city in the United States, where we stridently avoid SI units in our every day lives, the answer you would receive is '95' (degrees Fahrenheit for those of you who are not familiar with our antiquated system of measurement). Same level of heat, same average molecular kinetic energy, but a different answer to the same question. Does that mean we should abandon measurement of temperature? No. What it means is that we have to understand how the measurements are made and how to relate them to each other. The thermometers being used are not at fault, they are just measuring the same thing in slightly different ways. It is the type of problem we as scientists deal with all the time.

Which is why I am so perplexed by the way in which DNA microarrays have been singled out as being of questionable value in expression analysis. First, we have to recognize that no technology we have available measures gene expression. At best, what microarrays, or Northern blots, or SAGE, or QRT-PCR measure is RNA abundance. But even that is not quite true. We are not given an accounting of individual molecules in a sample (and *no*, SAGE doesn't do this either, even though the output is a numeric count of observed tags). What each technology gives us is a *surrogate* measure of RNA abundance—a fluorescence intensity, a band intensity, a tag count, or a C_t value. And each of these has its own biases, limitations, and weaknesses.

Second, among the publications in which various applications of DNA technology have been compared and found to be lacking, the authors generally either find that the expression measures between two conditions depend on the particular microarray platform being used or that measurements made on the same 'biological system' (tumor versus normal, for example) in two different studies produce lists of significant genes that are not highly concordant. In the former case, much like in measuring temperature, we have to understand whether we are measuring the same thing in the same way (or the same thousands of things in the same thousands of ways). For the latter, even in studies that use the same instruments, it is not clear that the samples under study, even if they carry the same labels, are truly the same and the problem of comparing lists can be effectively handled if well annotated high-quality data with information regarding the samples are available in public repositories so that we can constantly improve our conclusions.

The question of measurement and bias in DNA microarrays was recently addressed by Kuo *et al* (2006) in the most comprehensive analysis to date. Using 10 different microarray platforms and two QRT-PCR methods for comparison, Kuo *et al*'s analysis reveals some interesting facts. First, each of the platforms alone is highly consistent and reproducible. This observation is of crucial importance because it demonstrated that the basic technology is, for each probe on the array, making measurements that can be replicated. The same is true when comparing array results with QRT-PCR analyses or when comparing different QRT-PCR technologies.

Second, although the correlation between platforms is fairly good, it improves significantly if one limits the analysis to probes that overlap with each other. This goes beyond simply assigning probes to the same gene. Using correlation across platform as a measure, probes from the same RefSeq exon outperform those mapping to the same RefSeq gene, which are in turn better than those mapping to the same Locus Link accessions, and finally, these outperform those mapping to the same UniGene cluster. There are many potential reasons for this, ranging from alternate splicing to different sequence specificity, but the bottom line is that probes are more likely to correlate in their measurements if they are measuring the same thing.

One should note that manufacturers of the arrays are not at fault here. The problem here is that we are still working in partial darkness. Despite the fact that the human genome sequence has once again been declared 'finished' (Gregory *et al*, 2006), we still do not have a comprehensive catalog of all of the genes, the transcripts they encode, their variants, or their genomic structures (and we can still debate exactly what a gene is or how many exist in the human or other genomes). In many ways, arrays themselves have the best potential to solve this problem, as tiling arrays are, at present, the best experimental tool for elucidating transcripts and their variants on a global scale. And as the genome sequence and its annotation evolve, so will array technologies.

Third, the quality of commercial arrays has improved so that, in this study, the commercial arrays outperformed those made 'in house.' This again is not surprising as manufacturers whose goal is producing a product are generally better positioned to assure high quality than researchers whose primary interest is in the use of those products. Indeed, commercial organizations should be commended for increasing quality while decreasing price so that more laboratories have the opportunity to apply DNA microarray expression analysis in their experimental programs.

Finally, the correlation between platforms is best for genes expressed at moderate to high levels. Again this is not

surprising but it is important. Many of the genes in which we may be most interested, including many transcription factors, are expressed at relatively low levels. However, when the signal is close to the background noise, making precise measurements is difficult and this is true of any technology.

So are microarrays reliable? The weight of the evidence presented by Kuo and his colleagues demonstrates that if the experiments are performed carefully and if the data are analyzed in a consistent manner, then the signal from the biology dominates the choice of technology—particularly if each technology platform is measuring the same thing. This is because, fundamentally, every measurement is a convolution of the quantity being measured and the instrument being used to measure it. If we fully understand this principle, then appropriate experimental design and optimal use of the data will enable us to exploit better our measurements for discovery.

References

- Gregory SG, Barlow KF, McLay KE, Kaul R, Swarbreck D, Dunham A, Scott CE, Howe KL, Woodfine K, Spencer CCA, Jones MC, Gillson C, Searle S, Zhou Y, Kokocinski F, McDonald L, Evans R, Phillips K, Atkinson A, Cooper R, Jones C, Hall RE, Andrews TD, Lloyd C, Ainscough R, Almeida JP, Ambrose KD, Anderson F, Andrew RW, Ashwell RIS, Aubin K, Babbage AK, Bagguley CL, Bailey J, Beasley H, Bethel G, Bird CP, Bray-Allen S, Brown JY, Brown AJ, Buckley D, Burton J, Bye J, Carder C, Chapman JC, Clark SY, Clarke G, Clee C, Copley V, Collier RE, Corby N, Coville GJ, Davies J, Deadman R, Dunn M, Earthrowl M, Ellington AG, Errington H, Frankish A, Frankland J, French L, Garner P, Garnett J, Gay L, Ghorri MRJ, Gibson R, Gilby LM, Gillett W, Glithero RJ, Grafham DV, Griffiths C, Griffiths-Jones S, Grocock R, Hammond S, Harrison ESI, Hart E, Haugen E, Heath PD, Holmes S, Holt K, Howden PJ, Hunt AR, Hunt SE, Hunter G, Isherwood J, James R, Johnson C, Johnson D, Joy A, Kay M, Kershaw JK, Kibukawa M, Kimberley AM, King A, Knights AJ, Lad H, Laird G, Lawlor S, Leongamornlert DA, Lloyd DM, Loveland J, Lovell J, Lush MJ, Lyne R, Martin S, Mashreghi-Mohammadi M, Matthews L, Matthews NSW, McLaren S, Milne S, Mistry S, Moore MJF, Nickerson T, O'Dell CN, Oliver K, Palmeiri A, Palmer SA, Parker A, Patel D, Pearce AV, Peck AI, Pelan S, Phelps K, Phillimore BJ, Plumb R, Rajan J, Raymond C, Rouse G, Saenphimmachak C, Sehra HK, Sheridan E, Shownkeen R, Sims S, Skuce CD, Smith M, Steward C, Subramanian S, Sycamore N, Tracey A, Tromans A, Van Helmond Z, Wall M, Wallis JM, White S, Whitehead SL, Wilkinson JE, Willey DL, Williams H, Wilming L, Wray PW, Wu Z, Coulson A, Vaudin M, Sulston JE, Durbin R, Hubbard T, Wooster R, Dunham I, Carter NP, McVean G, Ross MT, Harrow J, Olson MV, Beck S, Rogers J, Bentley DR (2006) The DNA sequence and biological annotation of human chromosome 1. *Nature* **441**: 315–321
- Kuo WP, Liu F, Trimarchi J, Punzo C, Lombardi M, Sarang J, Whipple ME, Maysuria M, Serikawa K, Lee SY, McCrann D, Kang J, Shearstone JR, Burke J, Park DJ, Wang X, Rector TL, Ricciardi-Castagnoli P, Perrin S, Choi S, Bumgarner R, Kim JH, Short GF, Freeman III MW, Seed B, Jensen R, Church GM, Hovig E, Cepko CL, Park P, Ohno-Machado L, Jenssen TK (2006) A sequence-oriented comparison of gene expression measurements across different hybridization-based technologies. *Nat Biotechnol* **24**: 832–840