

Fun with Microarrays Part III: Integration and the End of Microarrays as We Know Them

Paul C. Boutros

Microarray experiments are an important method for the high-throughput investigation of biological phenomenon. A wide variety of techniques and algorithms exist for analyzing and extracting information from microarrays. This is the last of a series of reviews outlining the technological, computational, and experimental aspects of microarray experiments. This part focuses on cutting-edge applications of microarrays, with examples from drug discovery, evolutionary biology, and polymorphism analysis. A discussion of the future of array-based research is also given. Overall, microarray technologies are rapidly reaching maturation, and the focus is on their clever application to address previously intractable biological problems.

Prologue

THE YEAR IS 2050, and summer is approaching. We are in a classroom, watching senior high-school students study biology. The country, the city, the language — they do not matter. The topic is integrated cellular biology, and the teacher is trying to excite her restless class in the painstaking research that created an integrated cellular network by using that reliable hook: a real-world example.

Hal hazards a guess. “Is it a chemosensing plastic that recognizes changes in food pH, like they use in mood-clothing?” “That’s a good guess, Hal, but I’m sorry that’s incorrect.” Hal blushes and shifts in

his seat. “Actually, those sensors work using microarrays tuned to the concentration of bacterial DNA species present in the food product. As that concentration rises, the seal gradually changes colour to indicate freshness.”

The class appears mildly interested in this revelation, and the teacher capitalizes on the moment. “Did you know that only 40 years ago microarrays were a major research tool? That people used microarrays to elucidate signaling pathways, to determine transcription-factor targets, and even to develop early versions of personalized therapy?”

Paul C. Boutros | Division of Cancer Genomics and Proteomics, Ontario Cancer Institute, Toronto, Canada, M5G 2M9. Division of Signaling Biology, Ontario Cancer Institute, Toronto, Canada, M5G 2M9. Department of Medical Biophysics, University of Toronto, Toronto, Canada, M5G 2M9
Correspondence: paul.boutros@utoronto.ca

Received: 08/01/2007, Accepted: 09/03/2007

Posted online: 07/04/2008

© 2008 Paul C. Boutros. This is an Open Access article distributed by *Hypothesis* under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

At last the class laughs. Those were the old days — nobody was that silly anymore.

Introduction

New biological data is being uncovered at an accelerating pace. This alone should give us great hope for future understanding of disease, but there is an even greater hope. New research *technologies* are being developed even faster than our ability to apply these technologies. Forty-five

“...[M]icroarrays are becoming a standard part of the toolkit for molecular biologists. Surprisingly though, today we are also seeing the first hints that microarray technologies will soon be obsolete.”

years ago the structure of DNA was first being solved, and the concept of sequencing whole genomes and building massive sequence databases was fantastical — computers did not even exist to assemble or align such information! Fifteen years ago the first microarrays, containing only a few hundred genes (all that were known at the time!) were printed, allowing parallel quantitation of mRNA levels. Today, microarrays are becoming a standard part of the toolkit for molecular biologists. Surprisingly though, today we are also seeing the first hints that microarray technologies will soon be obsolete.

This review is the final part in a series discussing microarrays. The first part discussed the basic technology (*Boutros, 2006*),

while the second discussed the basic aspects of data analysis (*Boutros, 2007*). This third review attempts to provide insight into the new types of questions that can be answered with microarrays, or indeed with alternative technologies that might replace them.

I focus on three specific examples of new types of questions being answered by microarray approaches. The first is the “Connectivity Map,” which links mRNA expression responses to small molecules and disease states (*Lamb et al., 2006*). The second uses microarrays to understand and assess evolutionary trends (*Gilad et al., 2006*). The third uses microarrays to identify a large fraction of the genome as being subject to variations in copy-number.

Microarrays for Drug Discovery

A preponderance of molecular biological research is focused on the identification of novel therapeutics. Because most, if not all, diseases manifest themselves in changes at the mRNA level, it might be thought that drugs that can “reverse” these mRNA changes would be effective therapeutics. Yet millions of small molecules exist, and these compounds can have dose-and-tissue-specific effects (*Shioda et al., 2006*). So, can drugs that reverse disease-specific changes be identified?

One solution is to develop a large, public database of mRNA responses to a range of small molecules that many researchers can query to search for small-molecules related to their disease of interest. The “Connectivity Map,” developed by Todd Golub’s group at the Broad Institute is just such a database (*Lamb et al., 2006*). To get around the problem of dose-and-tissue-specific effects, the Golub group focused on one cell

line (MCF7, an epithelial breast cancer cell line) and profiled each compound at a single dose (typically 10 μ M) and time-point (6 hours). Small numbers of compounds were assayed in additional cell-lines, at other doses, or at a second time-point (12 hours). In total, the dataset covers “only” 164 compounds hand-selected to include major drug-classes and physiological compounds (e.g. estrogens). Notably, when the controls and additional cell-line, dose, and time-points were considered, a total of 564 arrays were required, implying a cost of at least \$400,000 for arrays alone.

This is an admirable attempt to profile a useful portion of the small-molecule space, but can such a small dataset truly be useful for discovering new therapeutics for specific diseases? Outside of the initial report of the Connectivity Map, the authors also published two detailed applications of the database to drug discovery, both in the context of cancer therapeutics.

The first application involves glucocorticoid resistance in acute lymphoblastic leukemia (ALL). Glucocorticoids are a first-line treatment for ALL, and tumours that are resistant to this therapy exhibit dramatically reduced prognosis (*Styczynski and Wysocki, 2002*). If glucocorticoid resistance could be reversed, then first-line therapy could be extended and patient outcome improved. To search for a compound that could accomplish this, the authors performed mRNA profiling on 13 ALL samples from patients sensitive to glucocorticoid treatment and 16 ALL samples from patients resistant to treatment (*Wei et al., 2006*). They identified a set of 157 genes that differentiated these two groups, and searched the Connectivity Map for compounds that mimicked this signature.

Surprisingly, rapamycin — an inhibitor of the mTOR pathway — significantly reversed glucocorticoid resistance. The authors proceeded to verify this functional finding and to partially elucidate the mechanism. Thus, in this case the Connectivity Map allowed identification of a novel therapeutic, despite the small size of the database.

The second application involves the treatment of hormone-refractory prostate cancer. Initially, most prostate cancers are dependent on the androgen receptor for growth, and treatments that reduce androgen receptor activity are highly effective. At some stage, though, tumours can become independent of androgen receptor activation, and therapies directed at this target fail. The authors used a small-molecule

“...[A] public database of mRNA responses to a range of small molecules that many researchers can query to search for small-molecules related to their disease of interest. The “Connectivity Map”, developed by Todd Golub’s group at the Broad Institute is just such a database...”

screen to identify two novel inhibitors of androgen receptor function, then used the expression signatures of these inhibitors to generate mechanistic hypotheses from the Connectivity Map (*Hieronymus et al., 2006*). The two inhibitors both showed very strong similarities to known HSP90 inhibi-

tors (the androgen receptor interacts with HSP90 in the cytoplasm and hormone activation dissociates this interaction to allow translocation to the nucleus). The authors proceeded to verify this mechanism and to demonstrate the potential efficacy of their new compounds. In this case, the Connectivity Map was used to elucidate the mechanism of a new drug.

It is hard to evaluate the utility of the Connectivity Map from these two success stories because we do not know how many attempts were made to yield these two validations. Nevertheless, the concept of using small-molecule profiles appears robust, and as the Connectivity Map is extended to additional compounds, doses, time-points, and cell-lines, its utility will increase.

...Intriguingly, transcription factors were particularly prone to showing human-specific alterations in expression, suggesting that changes in gene regulation are essential for speciation.

Microarrays for Evolutionary Biology

The idea that changes in mRNA expression are responsible for speciation is decades old, but testing this concept was challenging before the advent of microarrays (Gibson, 2002). In a recent study, Gilad and coworkers employed microarrays specifically designed to identify orthologous genes across four primate species (Gilad et al., 2006). That is, each probe on the array was capable of identifying the same gene in all four species,

but with slightly different hybridization affinities. This method allows an assessment of the species-specific changes in mRNA expression levels. Because the phylogenetic relationships amongst primate species is known, standard methods (Baldauf, 2003) could be used to investigate genes whose expression levels showed human-specific patterns. For example, the authors identified a series of genes whose expression remained constant in three primate species across 65 million years of evolutionary time, but diverged significantly in the 5 million years of human evolution. Intriguingly, transcription factors were particularly prone to showing human-specific alterations in expression, suggesting that changes in gene regulation are essential for speciation.

This study highlights the power of microarrays to probe questions in diverse fields of biology, and to shed light on the biology of ancestral species now extinct. While the Gilad study described here is one of the most comprehensive primate studies currently available, similar work has also been done to study the evolution of various species of *Drosophila* (Ranz et al., 2003).

Microarrays for Copy-Number Analysis

The best characterized and most prevalent type of intra-species genetic variation is the single-nucleotide polymorphism (SNP). SNPs occur, on average, at least once every 200 bp throughout the human genome (Schneider et al., 2003; Stephens et al., 2001), and can profoundly alter the way our cell responds to external stimuli (Okey et al., 2005a; Okey et al., 2005b) or disease states (Iida et al., 2002; Knudsen, 2006; Zhu et al., 2004). It has recently been shown that these single-nucleotide variations in the ge-

nome are not the only major source of genetic diversity amongst individuals. Rather, a significant fraction of the human genome has been shown to vary in *copy-number*. That is, a single region can be repeated only once in some individuals, but many times in others (Feuk *et al.*, 2006; Sebat *et al.*, 2004). A recent large-scale study strove to analyze the frequency of copy-number polymorphisms in a large cohort of individuals to assess just how often this phenomenon occurs (Redon *et al.*, 2006).

The authors of this seminal work selected 270 individuals from three different continents of ancestry (Europe, Africa, and Asia). Two separate microarray technologies were used to estimate which regions of the genome displayed either gains or losses amongst individuals. It is important to note that these 270 individuals were thought to be broadly healthy — the differences in copy-number observed are believed to be naturally occurring, rather than resulting from a disease, such as cancer. Indeed, for 180 of these individuals are in parent-offspring triplets (two parents and one offspring, thus 60 separate families in total). In these cases, it was possible to eliminate artifacts and spontaneous changes from the analysis.

In total, this work identified 12% of the human genome as being copy-number polymorphic in 1,447 separate regions. Intriguingly, gaps in the standard assemblies of the human genome are particularly likely to be associated with copy-number variations. Genes associated with copy-number polymorphisms are particularly enriched for cell adhesion and smell-related genes and depleted for cell-signaling and cell-proliferation functions. In a separate study the authors employed copy-number analysis of families with autism-affected individuals to

identify a copy-number variation associated with this disease (Szatmari *et al.*, 2007), thereby demonstrating the applicability of copy-number variations to the study of disease states.

...a significant fraction of the human genome has been shown to vary in copy-number. That is, a single region can be repeated only once in some individuals, but many times in others

Conclusions

Microarrays have been an important method for understanding biology in a fast, high-throughput fashion. The applications described here represent clever ways in which microarrays can be used to answer diverse biological questions. I did not highlight any of the more common, yet still important, applications of microarrays. For example, microarrays have been extensively applied to the search for prognostic markers or biomarkers for disease states (Li *et al.*, 2005; Raponi *et al.*, 2006; Simon *et al.*, 2003; van de Vijver *et al.*, 2002; Wang *et al.*, 2005), for studying the mechanism of a toxic or therapeutic compound (Rubins *et al.*, 2004; Waring *et al.*, 2001; Waring *et al.*, 2002), and for assessing genome-wide transcription-factor binding (Andrau *et al.*, 2006; Carroll *et al.*, 2005; Kim *et al.*, 2005; Odom *et al.*, 2006; Phuc Le *et al.*, 2005; Zhu *et al.*, 2006). Similarly, I did not highlight the innovative work being undertaken in non-mammalian model-organisms.

The three examples chosen here — drug-discovery, evolution, and human diversity —

indicate the diversity of microarray applications. But they also highlight a change in the way microarray experiments are conceived. Microarrays are not an end unto themselves, but a technology that can be used to address important open questions in all branches of biology.

As alluded to in the introduction, the future of microarray-based technologies may be in doubt with the advent of more powerful sequencing methods. It may soon be feasible to directly sequence cellular mRNA or DNA extracted from ChIP reactions. If this transpires, microarrays could easily be relegated to niche markets, like multi-species analysis for microbial detection. The first wave of papers exploiting this advanced sequencing technology show very promising results (*Barski et al., 2007; Bhinge et al., 2007; Johnson et al., 2007; Robertson et al., 2007*).

But the moral of this review, if reviews can be said to be moral, is that the technology is, in some sense, irrelevant. The questions addressed today with microarrays will remain important, even when they are later addressed with another, more powerful technology. The importance of understanding the foundations of our research tools, as outlined in part one of this series of reviews, is independent of the change and evolution of technologies. In short, the better we understand our tools, the better we can use them.

References

- Andrau, J.C., van de Pasch, L., Lijnzaad, P., Bijma, T., Koerkamp, M.G., van de Peppel, J., Werner, M., and Holstege, F.C. (2006). Genome-wide location of the coactivator mediator: Binding without activation and transient Cdk8 interaction on DNA. *Mol Cell* 22, 179-192.
- Baldauf, S.L. (2003). Phylogeny for the faint of heart: a tutorial. *Trends Genet* 19, 345-351.
- Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell* 129, 823-837.
- Bhinge, A.A., Kim, J., Euskirchen, G.M., Snyder, M., and Iyer, V.R. (2007). Mapping the chromosomal targets of STAT1 by Sequence Tag Analysis of Genomic Enrichment (STAGE). *Genome Res* 17, 910-916.
- Boutros, P.C. (2006). Fun with Microarrays Part I: Of Probes and Platforms. *Hypothesis* 4, 15-21.
- Boutros, P.C. (2007). Fun with Microarrays II: Data Analysis. *Hypothesis* 5, 15-22.
- Carroll, J.S., Liu, X.S., Brodsky, A.S., Li, W., Meyer, C.A., Szary, A.J., Eeckhoute, J., Shao, W., Hestermann, E.V., Geistlinger, T.R., et al. (2005). Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* 122, 33-43.
- Feuk, L., Carson, A.R., and Scherer, S.W. (2006). Structural variation in the human genome. *Nat Rev Genet* 7, 85-97.
- Gibson, G. (2002). Microarrays in ecology and evolution: a preview. *Mol Ecol* 11, 17-24.
- Gilad, Y., Oshlack, A., Smyth, G.K., Speed, T.P., and White, K.P. (2006). Expression profiling in primates reveals a rapid evolution of human transcription factors. *Nature* 440, 242-245.
- Hieronymus, H., Lamb, J., Ross, K.N., Peng, X.P., Clement, C., Rodina, A., Nieto, M., Du, J., Stegmaier, K., Raj, S.M., et al. (2006). Gene expression signature-based chemical genomic prediction identifies a novel class of HSP90 pathway modulators. *Cancer Cell* 10, 321-330.
- Iida, A., Saito, S., Sekine, A., Mishima, C., Kitamura, Y., Kondo, K., Harigae, S., Osawa, S., and Nakamura, Y. (2002). Catalog of 605 single-

Boutros

- nucleotide polymorphisms (SNPs) among 13 genes encoding human ATP-binding cassette transporters: ABCA4, ABCA7, ABCA8, ABCD1, ABCD3, ABCD4, ABCE1, ABCF1, ABCG1, ABCG2, ABCG4, ABCG5, and ABCG8. *J Hum Genet* 47, 285-310.
- Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. *Science* 316, 1497-1502.
- Kim, T.H., Barrera, L.O., Zheng, M., Qu, C., Singer, M.A., Richmond, T.A., Wu, Y., Green, R.D., and Ren, B. (2005). A high-resolution map of active promoters in the human genome. *Nature* 436, 876-880.
- Knudsen, K.E. (2006). The cyclin D1b splice variant: an old oncogene learns new tricks. *Cell Div* 1, 15.
- Lamb, J., Crawford, E.D., Peck, D., Modell, J.W., Blat, I.C., Wrobel, M.J., Lerner, J., Brunet, J.P., Subramanian, A., Ross, K.N., et al. (2006). The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313, 1929-1935.
- Li, W., Kessler, P., Yeger, H., Alami, J., Reeve, A.E., Heathcote, R., Skeen, J., and Williams, B.R. (2005). A gene expression signature for relapse of primary wilms tumors. *Cancer Res* 65, 2592-2601.
- Odom, D.T., Dowell, R.D., Jacobsen, E.S., Nekudova, L., Rolfe, P.A., Danford, T.W., Gifford, D.K., Fraenkel, E., Bell, G.I., and Young, R.A. (2006). Core transcriptional regulatory circuitry in human hepatocytes. *Mol Syst Biol* 2, 2006 0017.
- Okey, A.B., Boutros, P.C., and Harper, P.A. (2005a). Polymorphisms of human nuclear receptors that control expression of drug-metabolizing enzymes. *Pharmacogenet Genomics* 15, 371-379.
- Okey, A.B., Franc, M.A., Moffat, I.D., Tijet, N., Boutros, P.C., Korkalainen, M., Tuomisto, J., and Pohjanvirta, R. (2005b). Toxicological implications of polymorphisms in receptors for xenobiotic chemicals: The case of the aryl hydrocarbon receptor. *Toxicol Appl Pharmacol* 207, 43-51.
- Phuc Le, P., Friedman, J.R., Schug, J., Brestelli, J.E., Parker, J.B., Bochkis, I.M., and Kaestner, K.H. (2005). Glucocorticoid receptor-dependent gene regulatory networks. *PLoS Genet* 1, e16.
- Ranz, J.M., Castillo-Davis, C.I., Meiklejohn, C.D., and Hartl, D.L. (2003). Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. *Science* 300, 1742-1745.
- Raponi, M., Zhang, Y., Yu, J., Chen, G., Lee, G., Taylor, J.M., Macdonald, J., Thomas, D., Moskaluk, C., Wang, Y., et al. (2006). Gene expression signatures for predicting prognosis of squamous cell and adenocarcinomas of the lung. *Cancer Res* 66, 7466-7472.
- Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., et al. (2006). Global variation in copy number in the human genome. *Nature* 444, 444-454.
- Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., Euskirchen, G., Bernier, B., Varhol, R., Delaney, A., et al. (2007). Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nature methods* 4, 651-657.
- Rubins, K.H., Hensley, L.E., Jahrling, P.B., Whitney, A.R., Geisbert, T.W., Huggins, J.W., Owen, A., Leduc, J.W., Brown, P.O., and Relman, D.A. (2004). The host response to smallpox: analysis of the gene expression program in peripheral blood cells in a nonhuman primate model. *Proc Natl Acad Sci U S A* 101, 15190-15195.
- Schneider, J.A., Pungliya, M.S., Choi, J.Y., Jiang, R., Sun, X.J., Salisbury, B.A., and Stephens,

- J.C. (2003). DNA variability of human genes. *Mech Ageing Dev* 124, 17-25.
- Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Maner, S., Massa, H., Walker, M., Chi, M., et al. (2004). Large-scale copy number polymorphism in the human genome. *Science* 305, 525-528.
- Shioda, T., Chesnes, J., Coser, K.R., Zou, L., Hur, J., Dean, K.L., Sonnenschein, C., Soto, A.M., and Isselbacher, K.J. (2006). Importance of dosage standardization for interpreting transcriptomal signature profiles: evidence from studies of xenoestrogens. *Proc Natl Acad Sci U S A* 103, 12033-12038.
- Simon, R., Radmacher, M.D., Dobbin, K., and McShane, L.M. (2003). Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. *J Natl Cancer Inst* 95, 14-18.
- Stephens, J.C., Schneider, J.A., Tanguay, D.A., Choi, J., Acharya, T., Stanley, S.E., Jiang, R., Messer, C.J., Chew, A., Han, J.H., et al. (2001). Haplotype variation and linkage disequilibrium in 313 human genes. *Science* 293, 489-493.
- Styczynski, J., and Wysocki, M. (2002). In vitro drug resistance profiles of adult acute lymphoblastic leukemia: possible explanation for difference in outcome to similar therapeutic regimens. *Leuk Lymphoma* 43, 301-307.
- Szatmari, P., Paterson, A.D., Zwaigenbaum, L., Roberts, W., Brian, J., Liu, X.Q., Vincent, J.B., Skaug, J.L., Thompson, A.P., Senman, L., et al. (2007). Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat Genet* 39, 319-328.
- van de Vijver, M.J., He, Y.D., van't Veer, L.J., Dai, H., Hart, A.A., Voskuil, D.W., Schreiber, G.J., Peterse, J.L., Roberts, C., Marton, M.J., et al. (2002). A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 347, 1999-2009.
- Wang, Y., Klijn, J.G., Zhang, Y., Sieuwerts, A.M., Look, M.P., Yang, F., Talantov, D., Timmermans, M., Meijer-van Gelder, M.E., Yu, J., et al. (2005). Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 365, 671-679.
- Waring, J.F., Ciurlionis, R., Jolly, R.A., Heindel, M., and Ulrich, R.G. (2001). Microarray analysis of hepatotoxins in vitro reveals a correlation between gene expression profiles and mechanisms of toxicity. *Toxicol Lett* 120, 359-368.
- Waring, J.F., Gum, R., Morfitt, D., Jolly, R.A., Ciurlionis, R., Heindel, M., Gallenberg, L., Buratto, B., and Ulrich, R.G. (2002). Identifying toxic mechanisms using DNA microarrays: evidence that an experimental inhibitor of cell adhesion molecule expression signals through the aryl hydrocarbon nuclear receptor. *Toxicology* 181-182, 537-550.
- Wei, G., Twomey, D., Lamb, J., Schlis, K., Agarwal, J., Stam, R.W., Opferman, J.T., Sallan, S.E., den Boer, M.L., Pieters, R., et al. (2006). Gene expression-based chemical genomics identifies rapamycin as a modulator of MCL1 and glucocorticoid resistance. *Cancer Cell* 10, 331-342.
- Zhu, X., Wiren, M., Sinha, I., Rasmussen, N.N., Linder, T., Holmberg, S., Ekwall, K., and Gustafsson, C.M. (2006). Genome-wide occupancy profile of mediator and the Srb8-11 module reveals interactions with coding regions. *Mol Cell* 22, 169-178.
- Zhu, Y., Spitz, M.R., Amos, C.I., Lin, J., Schabath, M.B., and Wu, X. (2004). An evolutionary perspective on single-nucleotide polymorphism screening in molecular cancer epidemiology. *Cancer Res* 64, 2251-2257.