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?”
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 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.

“...[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.”

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 µM) 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 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-
nosome 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 than 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


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.


