References

1 Prusiner, S.B. (1982) Novel proteinaceous infectious particles cause scrapie. *Science* 216, 136–144
2 Wickner, R.B. (1994) [URE3] as an altered Ure2 protein: evidence for a prion analog in *Saccharomyces cerevisiae*. *Science* 264, 566–569
3 Zhouravleva, G.A. *et al.* (2002) To stick or not to stick: prion domains from yeast to mammals. In *Recent Research Developments in Molecular and Cellular Biology* (Vol. 3) (Pandalai, S.G., ed.), pp. 185–218, Research Signpost
4 Chernoff, Y.O. (2004) Cellular control of prion formation and propagation in yeast. In *Prions and Prion Diseases: Current Perspectives* (Telling, G. ed.), pp. 257–303, Horizon Scientific Press
5 Osherovich, L.Z. *et al*. (2004) Dissection and design of yeast prions. *PLoS Biol.* 2, E86
6 Kushnirov, V.V. and Ter-Avanesyan, M.D. (1998) Structure and replication of yeast prions. *Cell* 94, 13–16
7 Chernoff, Y.O. *et al*. (1995) Role of the chaperone protein Hsp104 in propagation of the yeast prion-like factor. *Science* 268, 880–884
8 Glover, J.R. and Lindquist, S. (1998) Hsp104, Hsp70 and Hsp40: a novel chaperone system that rescues previously aggregated proteins. *Cell* 94, 1–20
9 Wegrzyn, R.D. *et al*. (2001) Mechanism of prion loss after Hsp104 inactivation in yeast. *Mol. Cell. Biol.* 21, 4656–4669
10 Ferreira, P.C. *et al*. (2001) The elimination of the yeast [PSI$^+$] prion by guanidine hydrochloride is the result of Hsp104 inactivation. *Mol. Microbiol.* 40, 1357–1369
11 Jung, G. and Masison, D.C. (2001) Guanidine hydrochloride inhibits Hsp104 activity *in vivo*: a possible explanation for its effect in curing yeast prions. *Curr. Microbiol.* 43, 7–10
12 Borchsenius, A.S. *et al*. (2001) Yeast prion protein derivative defective in aggregate shearing and production of new seeds. *EMBO J.* 20, 6683–6691
13 Bailleul, P.A. *et al*. (1999) Genetic study of interactions between the cytoskeletal assembly protein Sla1 and prion-forming domain of the release factor Sup35 (eRF3) in *Saccharomyces cerevisiae*. *Genetics* 153, 81–94
14 Crist, C.G. *et al*. (2003) [PHI$^+$], a novel Sup35-prion variant propagated with non-Gln/Asn oligopeptide repeats in the absence of the chaperone protein Hsp104. *Gene Cells* 8, 603–618
15 Fischer, M. *et al*. (1996) Prion protein (PrP) with amino-proximal deletions restoring susceptibility of PrP knockout mice to scrapie. *EMBO J.* 15, 1255–1264
16 Owen, F. *et al*. (1989) Insertion in prion protein gene in familial Creutzfeldt–Jakob disease. *Lancet* 1, 51–52
17 Goldfarb, L.G. *et al*. (1991) Transmissible familial Creutzfeldt–Jakob disease associated with five, seven, and eight extra octapeptide coding repeats in the PRNP gene. *Proc. Natl. Acad. Sci. U. S. A.* 88, 10926–10930
18 Chiesa, R. *et al*. (2000) Accumulation of protease-resistant prion protein (PrP) and apoptosis of cerebellar granule cells in transgenic mice expressing a PrP insertional mutation. *Proc. Natl. Acad. Sci. U. S. A.* 97, 5574–5579
19 Parham, S.N. *et al*. (2001) Oligopeptide repeats in the yeast protein Sup35p stabilize intermolecular prion interactions. *EMBO J.* 20, 2111–2119
20 Bradley, M.E. and Liebman, S.W. (2004) The Sup35 domains required for maintenance of weak, strong or undifferentiated yeast [PSI$^+$] prions. *Mol. Microbiol.* 51, 1649–1659

0167-7799/$ - see front matter
doi:10.1016/j.tibtech.2004.09.001

GenCompass: a universal system for analysing gene expression for any genome

Nicholas M. Luscombe[1] and M. Madan Babu[2]

[1]Department of Molecular Biophysics and Biochemistry, Yale University, PO Box 208114, New Haven, CT 06520-8114, USA
[2]MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB22QH, UK

Microarrays have become indispensable tools for studying the gene expression of particular organisms on a genomic scale. However, despite its widespread use, there are several draw-backs to the current technology. First, it requires prior knowledge of the DNA sequence encoded in the organism of interest, and second, chips must be designed specifically for each genome, greatly increasing the initial cost incurred in manufacturing the arrays.

To obviate the need for organism-specific arrays, studies have explored the design of generic microarrays containing a comprehensive set of short DNA probes [1]. Practical applications, however, have been hampered by technological difficulties. Van Dam and Quake previously showed that oligonucleotide probes would have to be 10–16 bp long to provide sufficient sequence specificity [2] and a generic microarray would contain over one million probes (4^{10}) to cover all possible DNA sequences. Furthermore, computational methods have so far been unable to account satisfactorily for non-specific hybridisation to short oligonucleotides.

The feasibility of a universal microarray system was recently demonstrated in an elegant study by Lizardi and co-workers [3], which combines standard enzymatic manipulation methods, a universal microarray representing all possible DNA hexamers, and advanced bioinformatics techniques. Having developed the technology, the authors tested its effectiveness by profiling gene expression changes in yeast. Here, we provide an overview of the platform developed by Lizardi and co-workers, and propose future applications for expression profiling.

The GenCompass method

The universal microarray technology, named GenCompass by the authors, uses a microarray containing a universal set of probes to detect short sequence tags

Corresponding authors: Nicholas M. Luscombe (nicholas.luscombe@yale.edu), M. Madan Babu (madanm@mrc-lmb.cam.ac.uk).
Available online 8 September 2004

derived from transcripts. The experimental process consists of three main stages (Figure 1): (a) fragmentation, (b) amplification and (c) hybridisation.

Fragmentation of transcript-derived cDNAs

cDNAs are produced for each transcript by applying random decamer primers and reverse transcriptase to total mRNA extracted from cells (Figure 1a). The fragmentation procedure uses a type II restriction enzyme (e.g. MboI) to cleave the cDNA. These enzymes typically cleave within a specific recognition site to leave a four-base overhang of a particular sequence. Recognition sites are distributed so that transcripts of ~2 kb in length would on average be cleaved into 12–16 fragments.

Next a first adaptor molecule (a biotin labelled short sequence of DNA that contains a type IIs restriction enzyme recognition site and a PCR primer site) is ligated to the overhanging bases. The modified fragments are then treated with a type IIs endonuclease (e.g. FokI), which cleaves non-specifically at a fixed distance away from the recognition site. The resulting four-base overhangs are derived from the original cDNAs, and are made up of one of 256 (4^4) possible sequences.

The resulting sample is divided into 80–120 pools in a microtiter plate. Each pool receives one of 256 (4^4) different second adaptors containing a specific four-base overhang and a second PCR primer site. These second adaptors ligate only with cDNA ends exposing complementary overhangs and result in pools containing sub-populations of cDNA-derived fragments of equal length (~70 bp). Each fragment contains common primer-binding sites from the two adaptors, and an intervening segment (typically a ~14 bp sequence comprising the type II enzyme site, a short section of the cDNA sequence, and the four base overhang from the type IIs cut) that serves as the unique identifier for the fragment (Figure 1d).

Prior to the second ligation, the sample could theoretically be separated into a maximum of 256 pools for every

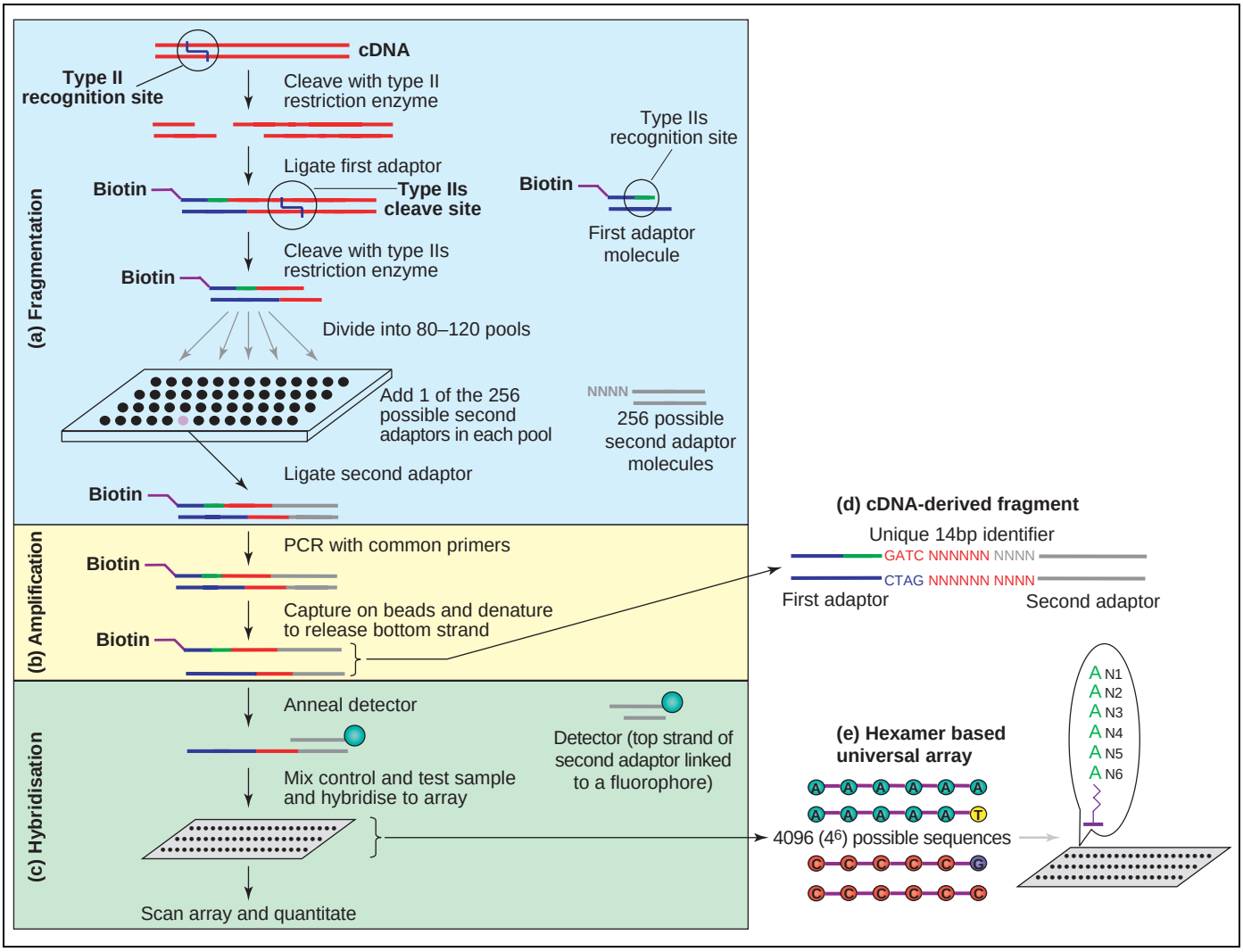

Figure 1. The GenCompass procedure. The experimental procedure comprises three main steps: **(a)** fragmentation, **(b)** amplification, and **(c)** hybridisation. During fragmentation, transcript-derived cDNAs are cleaved successively by two different restriction enzymes, and ligated with two adaptor molecules to produce a cDNA-derived fragment that is ~70bp long. As shown in **(d)**, each fragment contains a 14bp sequence at the centre, which serves as the unique identifier of the transcript from which it originates. During amplification, fragments are amplified from the common primers located in the ligated adaptor molecules by PCR. Finally, in hybridisation, amplified fragments are ligated to a fluorophore-linked detector and hybridised to **(e)** a generic microarray representing all possible DNA hexamers. By identifying the hybridised hexameric probe on the microarray and the ligated adaptors, the unique 14 bp sequence of the fragments can be deduced. Once known, this sequence is searched against a database to ascertain the expressed genes. Adapted with permission from from [3]. (http://www.nature.com).

possible second adaptor; but the authors demonstrate that they achieve 90% coverage of yeast genes (i.e. have at least one fragment per gene) by using a substantially smaller number of pools. By selecting the enzyme pair carefully from databases such as REBASE (http://rebase.neb.com/rebase/rebase.html) [4] one could maximise the coverage of the genome if its sequence is already known.

Amplification of cDNA-derived fragments

cDNA-derived fragments containing both primer-binding sites from the adaptors are amplified by PCR (Figure 1b). Differential amplification is minimised because all fragments are equal in size, and have similar base compositions. After PCR, amplicons are captured on beads and denatured. The released bottom strand is then annealed to a fluorescence labelled ligator detector to produce partial duplexes in preparation for hybridisation to the array.

Hybridisation to the universal microarray

In the final step, corresponding pools from control and test samples of cDNA-derived fragments are mixed and hybridised to a universal microarray representing all 4096 (4^6) combinations of six-base sequences (Figure 1c,e). Each experimental set-up requires the same numbers of hybridisations as there are pools.

The 14-base sequence of the cDNA segment in each fragment can be deduced from the identity of the hexameric probe to which the fragment hybridised, plus the four-base overhangs generated by the two enzymatic cleavages. This short sequence is then searched against a database to identify the gene from which it originates, and the relative hybridisation levels from the test and control samples identify those that are differentially expressed.

Computationally intensive data analysis

After performing the experiments, there are substantial computational challenges for interpreting the resulting data. In addition to the now common analysis methods such as intensity normalisation [5], the authors developed two procedures that are particularly important for the GenCompass system.

Isolation of correct hybridisation signals

The first procedure characterises the hybridisation specificity of the cDNA-derived fragments. Non-specific hybridisation to mismatching DNA probes is well documented [6], and the problem is more acute for the current system because the oligonucleotide probes are short, and all possible hexamers are available for hybridisation. Because of the importance of determining the exact six-base sequence encoded in the DNA fragments, the authors developed a mathematical model to isolate signals generated by correct and mismatching hybridisation events.

Database search for hybridised cDNA-derived fragments

The second procedure is to interpret the hybridisation signature by making gene assignments to the expressed cDNA-derived fragments. This is conceptually simple and involves a sequence search against the appropriate database. However the process is complicated by the need to assemble short sequence fragments into a description of which complete genes are being expressed. This can undoubtedly be automated by building a software tool that performs sequence searches against databases such as GenBank [7], and compiles gene assignments.

Obviously the results ultimately depend on the completeness of the sequence information for the organism under study. However, an intriguing consequence of this set-up is that results can be continually improved as databases are updated, without having to repeat the experiments each time.

Technical validation using *Saccharomyces cerevisiae*

The authors validated their technology successfully by profiling gene expression levels in *Saccharomyces cerevisiae*. It must be stressed that in testing, no changes were made to the protocol or microarray design, highlighting the applicability of the system to any genome. In the first experiment, the authors measured the galactose response of yeast; most genes previously associated with this cellular condition were identified, including key galactose inducible genes such as GAL1. They also detect expression of 36 short novel genes, which have only recently been identified in the genome; these genes would have been missed by conventional microarrays unless updated to include new annotations. In a second experiment, the authors engineered a yeast strain containing the T7 T-cell receptor, and successfully detected the expression of the foreign gene. Again, such an experiment would have been much harder using conventional microarrays, which would have to be modified to include probes for each new gene.

Future applications

Clearly the strength of the technology lies in its versatility, which allows it to be used for a wide range of organisms and experimental set-ups. As the authors demonstrated, the system works well for profiling gene expression levels in model organisms, including those that have been engineered with foreign genes. However the potential of the system is not limited to such applications and it can possibly be applied to many further studies. These include: (i) improving current genome annotations by confirming or identifying novel transcripts; (ii) identifying non-coding RNA by restricting sequence searches to non-protein-coding regions of the genome; (iii) monitoring alternative splice forms of genes in different tissue types; (iv) identifying transcripts bound to specific RNA-binding proteins by precipitating such complexes before hybridisation and (v) comparative gene expression profiling across several organisms, and identifying gene expression signatures in related pathogens.

In summary, GenCompass is a system that elegantly combines traditional molecular biological techniques with microarray technology, and integrates intensive computation to give biological meaning to the results. It represents a significant technological advance, and its ability to perform large-scale experiments – previously mainly confined to model systems – on poorly characterised genomes will surely enable it to make a significant contribution to genomic biology.

Acknowledgements
M.M.B. would like to acknowledge financial support from the MRC Laboratory of Molecular Biology, Trinity College, Cambridge and the Cambridge Commonwealth Trust.

References

1 Gunderson, K.L. *et al*. (1998) Mutation detection by ligation to complete n-mer DNA arrays. *Genome Res.* 8, 1142–1153
2 van Dam, R.M. and Quake, S.R. (2002) Gene expression analysis with universal n-mer arrays. *Genome Res.* 12, 145–152
3 Roth, M.E. *et al*. (2004) Expression profiling using a hexamer-based universal microarray. *Nat. Biotechnol.* 22, 418–426
4 Roberts, R.J. *et al*. (2003) REBASE: restriction enzymes and methyltransferases. *Nucleic Acids Res.* 31, 418–420
5 Luscombe, N.M. *et al*. (2003) ExpressYourself: A modular platform for processing and visualizing microarray data. *Nucleic Acids Res.* 31, 3477–3482
6 Housby, J.N. and Southern, E.M. (1998) Fidelity of DNA ligation: a novel experimental approach based on the polymerisation of libraries of oligonucleotides. *Nucleic Acids Res.* 26, 4259–4266
7 Benson, D.A. *et al*. (2004) GenBank: update. *Nucleic Acids Res.* 32, D23–D26

0167-7799/$ - see front matter
doi:10.1016/j.tibtech.2004.08.013

Rewiring the cell: synthetic biology moves towards higher functional complexity

Michael L. Simpson

Molecular-Scale Engineering and Nanoscale Technology Research Group, The Oak Ridge National Laboratory, P.O. Box 2008, MS 6006, Oak Ridge, Tennessee 37831–6006, USA

A steady stream of research has fueled excitement in the field of synthetic biology. Logic gates, oscillators, and memory elements constructed using genetic and biochemical components have all been demonstrated. However, the nagging question remains as to how higher levels of complexity can be designed into these synthetic systems. A recent paper from Collins' group provides some answers to this question.

The intertwined development of a fundamental science and an engineering discipline both centered on information processing in genetic circuits and networks is remarkable. It is true that the fundamental science came first as investigators endeavored to describe regulation and control of cellular processes. But even in these early efforts there was a realization that an engineering discipline lay ready for development as Jacob and Monod opined that gene-regulatory circuits with many different functions could be constructed using networks of simple regulatory elements [1]. Although they are still focused on describing natural genetic systems, many recent efforts have turned to engineering concepts. For example, the competition between the two protein populations that control the selection of lytic or lysogenic pathways in λ phage infection has been likened to critical races between nearly simultaneous electrical signals in digital circuits [2]. The role of stochastic fluctuations in the λ switch has been described using the concept of dithering, in which noise is intentionally added to improve the performance of analog-to-digital converters [3,4] and autoregulated gene circuits have been analyzed using concepts borrowed from electronic amplifier design [5].

Corresponding author: Michael L. Simpson (SimpsonML1@ornl.gov).
Available online 30 September 2004

Only a small step from borrowing engineering concepts for description, we are now seeing the emergence of a design discipline that focuses on manipulating information processing and transport within genetic circuits and networks. There have been numerous efforts to create engineered gene circuits that function according to human design using a 'silicon mimetic' approach, in which engineered genetic or biochemical systems are made to emulate the functionality of semiconductor devices such as logic gates, latches, and oscillators [6–14]. Combinatorial logic gates have been realized using approaches that include random assembly followed by functional screening [8], the rational design of synthetic gene logic circuits and networks [15], and directed evolution [16]. Adding to the toolbox of elemental gene circuits, an oscillator [6] and a toggle switch [7] have been reported. To date most efforts in synthetic biology have focused on the development of individual functional components or relatively small networks of these components. However, complex functionality is more a result of interconnection than fundamental component complexity. In other words, complexity arises from the massive interconnection of simplicity, and the development of flexible and parallel ways to interconnect these circuits is a daunting challenge for the next crucial leaps forward in synthetic gene-network design. Recent work from Collins' group [17] demonstrates that one promising approach is to leverage the existing complexity of natural gene circuits and networks while using synthetic circuits to customize function.

Higher levels of complexity

A major obstacle in generating highly complex function using the synthetic gene circuits described above is the requirement of long ('deep') logic cascades [18]. Many electronic circuits are constructed as deep logic cascades