INNOVATIONS

New technologies to assess genotype–phenotype relationships

Barry R. Bochner

The accelerating pace of the discovery of genes has far surpassed our capabilities to understand their biological function — in other words, the phenotypes they engender. We need efficient and comprehensive large-scale phenotyping technologies. This presents a difficult challenge because phenotypes are numerous and diverse, and they can be observed and annotated at the molecular, cellular and organismal level. New technologies and approaches will therefore be required. Here, I describe recent efforts to develop new and efficient technologies for assessing cellular phenotypes.

Ever since Gregor Mendel used the observable traits of pea plants to define and follow units of genetic inheritance, the definition and testing of phenotypes has had a key role in genetic analysis. Phenotypes are important for several reasons. They allow us to observe genetically inherited traits and events, and aid in genetic manipulations. Genetic changes that confer a growth or survival advantage, or a trait that can be scored physically, have been exploited to great advantage. Examples include the use of selectable drug-resistance genes (with drugs such as tetracycline, kanamycin and geneticin) and the selection and scoring of clones on the basis of β-galactosidase activity[1]. Phenotypes that confer a growth or survival disadvantage are also useful. They allow dissection of functional relationships by providing conditions for selecting suppressors that compensate for the disadvantage. Finding, identifying and understanding suppressors has been an important method for getting from a gene of interest to other genes (proteins) that interact with it. Phenotypes are also crucial because they are the expression of genotypes and reveal gene function. In this regard, phenotypes are an essential intermediate in the pathway from basic genetics to biological understanding.

Importance of phenotypes in genomics

In the past decade, we have witnessed an explosion in the availability of new genetic analysis tools and genomic information. Sequencing technology has provided us with complete genomic sequences for species ranging from microbes to plants and animals[2–8] — including that of the human[9,10]. These projects were accompanied by efforts to locate, enumerate and annotate genes and to assign known or putative biochemical functions to them. However, from the most thoroughly studied and 'simple' bacterial cells[2] to man[9,10], only about two-thirds of all genes have an assigned biochemical function and only a fraction of those are associated with a phenotype[11–13]. Even when phenotypes are assigned, they might represent only a partial understanding of the role of the gene. The function of a gene cannot be fully understood until it is possible to predict, describe and explain all the phenotypes that result from the wild-type and mutant forms of that gene.

Phenotypes often cannot be predicted on the basis of the biochemical function of a gene alone because it is not clear how a catalytic or regulatory activity will affect the biology of the cell or the whole organism. However, if a gene has a biological function then, for every identified gene, it should be possible to define at least one phenotype. A second layer of genomic annotation could then follow, in which every gene is described biologically by the phenotypes that it produces (shown conceptually in FIG. 1). A first step in producing a so-called 'phenomic map' has been made for *Escherichia coli* by LaRossa[14] who has tabulated ~1,000 phenotypes that correspond to various genes that have been studied. In diploid and higher organisms in particular, this will be complicated by the fact that several genes can affect gene expression[15], and the resulting phenotypes[16] of each other, leading to epistasis, complex traits and multifactorial diseases.

Along with phenomic maps, there is a need for phenotypic standardization that has already been recognized by breeding and stock centres[17]. Several projects[18,19] have begun to develop a standardized approach to developing annotation and databases. Just as comparative genomics has allowed powerful extrapolation of gene and protein function from one cell type to another[12,13], it will be important to develop a coordinated effort to standardize phenomic nomenclature to facilitate database searches, comparisons and extrapolations. Such a system of comparative phenomics would facilitate the progression of knowledge throughout model biological systems from bacteria to humans.

Many scientists are coming to the conclusion that advances in genetic and genomic analysis are being hindered by the slow pace at which our understanding of biology is progressing. Simply put, biological (that is, phenotypic) information is not keeping pace with genomic information. In 1989, I predicted that global phenotypic analysis would soon be needed to complement the massive amounts of genetic data being obtained[20], and, in 1996, Brown and Peters called attention to 'the phenotype gap' in mouse research[21]. The Nobel laureate Sydney Brenner, in a recent keynote address (at a joint Cold Spring Harbor Laboratory/Wellcome Trust Genome Informatics Conference held at Hinxton in the UK on 9 September 2002) emphasized

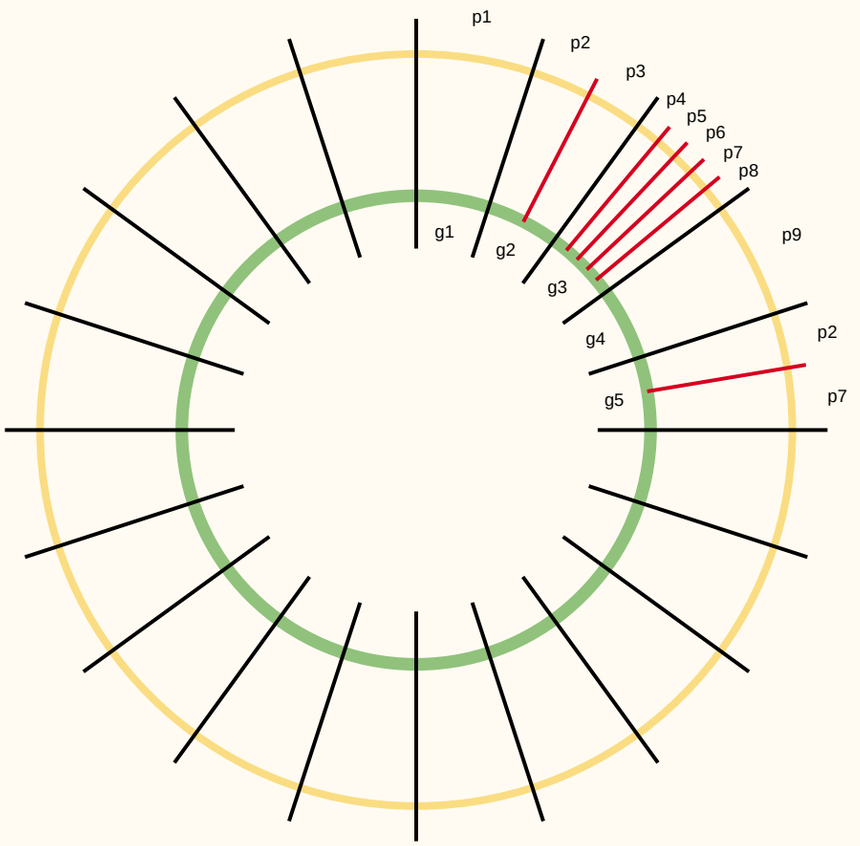


Figure 1 | **Genotypic and phenotypic maps.** A phenotypic map (yellow) can be generated to correspond to any genomic map (green). Some genes, such as gene1 (g1), have only one corresponding phenotype (p1), whereas most genes have many corresponding phenotypes. Phenotypes can be coded for by more than one gene, as shown by p2, which is affected by g2 and g5.

that approaches that relied heavily on genome sequences and bioinformatic extrapolation had too much noise and were becoming non-productive. Instead, he called for a renewed focus on cellular studies and the creation of function-based cell maps in a variety of cell types by the year 2020.

However, generating phenotypic maps will not be easy. Scientists generally test and measure phenotypes one at a time, which is too slow. Almost every model system in which the genome has been sequenced has functional genomics projects to associate the genome with the biology, and this typically includes some efforts that involve phenomics. Many large-scale projects are being carried out both in publicly funded research projects (for example, for animals[22–24] and for plants[25]) and in corporations (such as Lexicon Genetics, Inc., Deltagen, Inc., Phenomix Corporation, SurroMed, Inc. and Paradigm Genetics, Inc.). These projects generally use and adapt diverse existing phenotypic technologies that range from animal autopsies to MASS SPECTROMETER analysis of cellular metabolites.

Although cellular phenotyping does not replace plant or animal phenotyping, it can provide a more rapid, efficient and cost-effective method by which to begin to understand the phenotypes of the tens of thousands of non-annotated genes. The testing of cell suspensions is more amenable to large-scale high-throughput testing and can be implemented with modern robotics and instrumentation. However, so far, robotics has been used primarily to automate small numbers of phenotypic assays. There are few reports of efforts to test many phenotypes simultaneously. To maintain momentum and productivity in biological research, we need much more comprehensive and efficient tools for testing cellular phenotypes. The remainder of this article discusses recent efforts to develop better technologies for assessing genotype–phenotype relationships in cellular systems.

"…a system of comparative phenomics would facilitate the progression of knowledge throughout model biological systems from bacteria to humans."

Phenotyping in single-cell systems

The most complete gene annotation is available for simple microbial-cell model organisms such as *E. coli*[2] and *Saccharomyces cerevisiae*[3]. There are many advantages to large-scale phenotyping in single-cell systems, especially microbial cells, in which it is easier to standardize the biology and to alter genes and assess phenotypes. The phenotypes that are measured are typically biochemical and, therefore, can be easily related to specific enzymatic activities. Gene functions that are initially determined in these models can provide the basis for extrapolation to more complex life forms in which phenotypic testing presents further levels of complexity.

S. cerevisiae researchers have taken the lead in 'genomic-scale phenotyping'. Efforts began in 1996, when a consortium of yeast researchers undertook a project to construct ISOGENIC knockouts of most of the ~6,000 known genes[26]. Hampsey[27] published an overview of yeast phenotypes, and several groups took up the challenge of phenotyping knockout strains as a method of determining the function of various genes. The approaches that were taken are summarized in TABLE 1.

Although the efforts with yeast set a direction for large-scale phenotyping, their results have left many open issues and unanswered questions. A high percentage of the knockout strains that were assayed showed phenotypic changes. This was surprising, as the largest number of phenotypes assayed was 300 and most studies measured ≤20 phenotypes. For example, Hegemann and co-workers[28] tested just 20 phenotypes but found changes in one-third of the strains, and two-thirds of the conditional mutants had multiple phenotypes. Clearly, one problem with most of these approaches is that the phenotypes that were tested, such as growth in rich or minimal media, were not specific. When a change is detected, we can postulate little, if anything, about the gene function. Hegemann and co-workers concluded that the provision of the mutants to the scientific community was likely to be of more use than the phenotypes that were detected, but they expressed the hope that "...experts in specific areas of yeast cell biology will be able to analyze the relatively few phenotypes in which they are experts"[28].

Table 1 | Large-scale phenotyping projects in *Saccharomyces cerevisiae*

Laboratory	Number of strains tested	Number of phenotypes tested	Test format	References
Brown	268	7	Growth assays	34
Davis	5,916	2–6	Growth competition assays with strain bar-coding	35–37
Slonimski	100	300	96-well microplates with agar media	38–40
Hegemann	Hundreds	20	96-well microplates with agar media	28,41
Snyder	8,000	20	Growth in 96-well microplates replica-plated to agar media	42
Lindquist	14	150	Growth on agar media	43
Harashima	465	11	Suspensions in 96-well microplates replica-plated to agar media	44
Blomberg	–	98	Growth in microwell plates in 350 µL liquid cultures	45

In general, the efforts to phenotype yeast mutants have not provided a basis for solving the general need for comprehensive and detailed cellular phenotyping. At most, 300 phenotypes were tested, the specific tests used are not readily adaptable for other types of cells, the technologies are still cumbersome for high-throughput applications and, in many cases, the phenotypes are still qualitative rather than quantitative.

Phenotype MicroArray technology

In 1998, our group began a programme to devise a phenotyping technology that had attributes that were missing from previous approaches: it could assay ~2,000 distinct culture traits; it could be used with a wide range of microbial species and cell types; it would be amenable to high-throughput studies and automation; it would allow phenotypes to be recorded quantitatively and stored electronically, to facilitate comparisons over time; it would give a comprehensive scan of the physiology of the cell; and, by providing global cellular analysis, it would provide a complement to genomic and proteomic studies (FIG. 2).

Instead of using growth-based assays, we have used a TETRAZOLIUM REDOX CHEMISTRY that produces a colour change in response to cell respiration[29] in each well of 96-well microplates. This gives an accurate reflection of the physiological state of the cell, and can be used in some important assays that do not depend on growth. The technology is feasible for high-throughput analyses because the microplates are manufactured with a stable dry chemistry ready for inoculation. Also, the monitoring and recording of data is automated, standardized and quantitative. The result of these efforts is a new technology that we have called Phenotype MicroArrays (PMs) (REF. 30; BOX 1).

The initial objective of PM technology was to allow the testing of thousands of phenotypes. One simple reason for having thousands of tests is that microbial cells have thousands of genes, and we expect that each gene will be responsible for one or more phenotypes. Furthermore, we wanted our selection of phenotypes to provide a comprehensive analysis of the basic physiology of the cell, and to use specific phenotypes that could point towards specific cellular pathways and biological functions. Nearly 2,000 tests could be accomplished by using 20 96-well microplates, tested simultaneously, and with detailed kinetics recorded.

Expanded phenotypic analyses

An example of a phenotypic comparison of two isogenic strains of *E. coli* is shown in FIG. 3. In this example, MG1655 — the genomically sequenced strain[2] — is compared with an isogenic derivative that contains a knockout of the *malF* gene caused by the insertion of a Tn10 (tetracycline resistance) transposon. The *malF* gene encodes a protein that is involved in the uptake of maltosides, so we would expect to see phenotypic defects related to maltose metabolism as well as resistance to tetracyclines. The PM analysis detects both types of phenotypic changes: the loss of maltose, maltotriose and dextrin metabolism (red lines in FIG. 3) and the gain of resistance to a variety of tetracyline antibiotics (green lines in FIG. 3).

Whereas mutation of a specific transport or metabolic function might result in a small number of easily interpretable phenotypic changes, mutation of a global regulatory gene might alter many phenotypes, so interpretation might be complex. We have previously published an example of an adenylate cyclase (*cya*) mutant of *E. coli* (REF. 30). More recently, Xiang-He Lei in our laboratory has analysed knockouts of 32 two-component regulatory genes of *E. coli* in collaboration with Zhou and Wanner at Purdue University (L. Zhou and B. Wanner, unpublished observations); nineteen of these were found to have detectable phenotypic changes. The number of phenotypes ranged from as few as one change, to as many as 50 changes for *arcA* and *arcB* deletions. Some of the phenotypes were expected, but others were not and remain to be explained. We have also analysed mutant strains for a number of other laboratories working on *E. coli* and have completed a phenotypic comparison of several wild-type *E. coli* strains that are in common use (X.-H. Lei *et al.*, unpublished observations). Applications of this technology are not limited to *E. coli*.

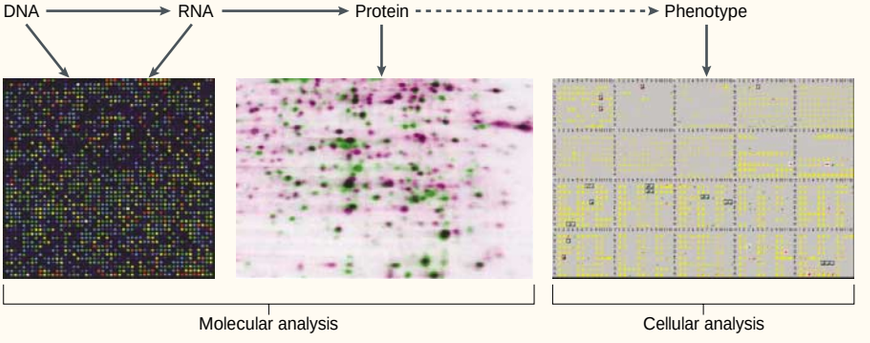


Figure 2 | **Global cellular analysis.** The information in cells flows from the level of genotype to the gene and protein expression levels, and results in cellular phenotypes. Modern tools for global analysis are beginning to provide a way to study and understand this process in greater detail.

Amalia Franco-Buff has analysed isogenic strains with alterations in the *prfA* gene of *Listeria monocytogenes* in collaboration with Jose Vazquez-Boland at the University of Bristol (A. Franco-Buff, unpublished observations). This is a particularly interesting regulatory gene because it regulates the biological functions that are essential for pathogenicity in this bacterium[31]. In another project, Richard Kostriken in our laboratory has analysed gene knockouts of human disease gene homologues in *S. cerevisiae* (R. Kostriken, unpublished observations). Over the past year, we have shown that we can use our current set of PMs to test other Gram-negative genera such as *Salmonella*, *Pseudomonas*, *Burkholderia*, *Vibrio* and *Sinorhizobium*; Gram-positive genera such as *Bacillus*, *Staphylococcus*, *Streptococcus* and *Enterococcus*; yeast such as *Candida* and *Cryptococcus*; and filamentous fungi such as *Aspergillus nidulans*. We have also had success in adapting this technique for bacteria that require incubation in special gas atmospheres (such as *Helicobacter pylori*).

Comprehensive phenotyping with PM technology is useful for many other types of comparison. In addition to knockouts, it is possible to compare the phenotypic consequences of gene underexpression or overexpression, as well as interesting alleles of genes such as those that encode regulatory proteins that lock the circuitry in which they are involved in the 'on' or 'off' state. Modern controllable promoters, such as the arabinose system in bacteria[32], can be used to vary the level of gene expression, including that of essential genes. Alternatively, a controllable promoter can be used to produce varying amounts of antisense RNA *in vivo*[33]. Another possibility involves the introduction of one or more genes into a cell line to determine the phenotypic consequences of their expression, which can reveal their function. An example of detecting the function of introduced genes (in this case, the tetracycline

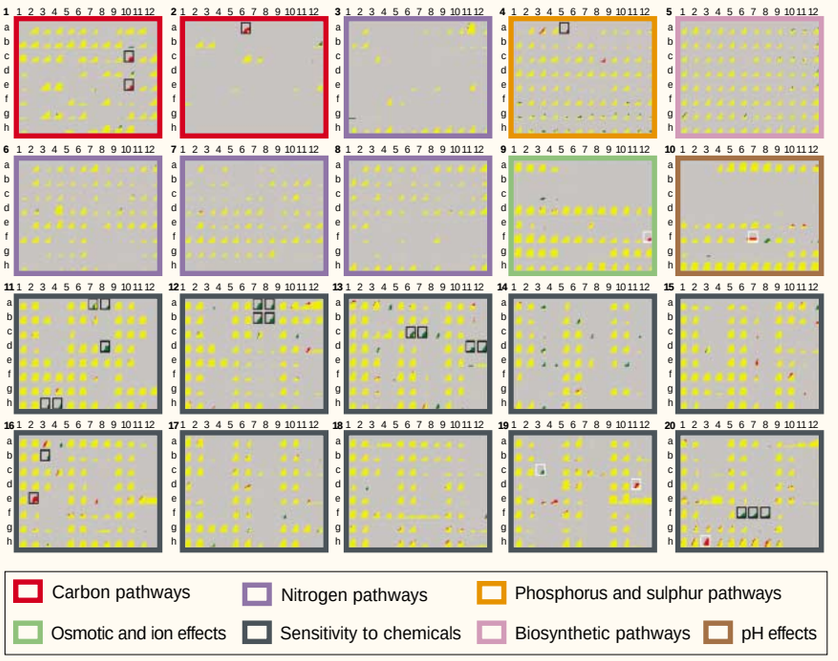
Box 1 | Phenotype MicroArray technology

Phenotype MicroArrays (PMs) are a simple tool for testing hundreds or thousands of cellular traits simultaneously. The PMs that are available at present contain ~2,000 tests that are selected to approximate a comprehensive scan of the known cellular pathways.

The layout of the 2,000 PM tests is summarized in panel **a**. PMs 1–8 test the main catabolic pathways in cells for carbon, nitrogen, phosphorus and sulphur, as well as biosynthetic pathways. PM9 tests osmotic and ion effects on the cell. PM10 primarily tests pH growth range and pH regulation. The remaining 10 PMs test the sensitivity of cells to a wide range of chemicals, including antibiotics, anti-metabolites, membrane-active agents, respiratory inhibitors and toxic metals. Antibiotics and anti-metabolites, with different modes of action, target the cell wall, membranes, ribosomes, RNA and DNA polymerases, and diverse metabolic pathways. Membrane-active agents and respiratory inhibitors probe the chemistry, structure and function of membrane-associated processes, such as respiration and protein localization. Toxic metals can be present in the environments of most cells, which are likely to have cellular systems for handling them.

To analyse a microbial strain, a cell suspension is prepared and inoculated into the set of microarrays. As shown in panel **b**, in which a pair of isogenic strains are compared, the PM panel sets are then placed inside the OmniLog — an incubator/reader instrument that cycles the arrays in front of an imaging head every 15 minutes, measuring and recording the colour formed from reduction of the tetrazolium dye in each well. Computer software plots kinetic graphs of colour formation against time for each well and each strain. When two strains are compared, the reference strain is plotted in red and the mutant strain in green. This is analogous to labelling the RNA from two strains with red and green dyes in gene-expression analysis. The software can compare the kinetic phenotypes by overlaying the kinetic graphs and colouring areas of overlap (no change) in yellow. The result is a red–green–yellow array in which phenotypes that are lost are coloured red, phenotypes that are gained are coloured green and unchanged phenotypes are coloured yellow. Thresholds can be set to disregard small and insignificant changes, and all of the wells with changes that exceed this threshold are marked with a black box.

a



b

resistance transposon Tn10) is shown in FIG. 3. Many laboratories have libraries of HETEROLOGOUS GENES that have been cloned in *E. coli*. If the cloned genes have promoters that are operative in *E. coli*, it might be possible to assay directly for the function of these genes using the bacterial cell as a surrogate. Other interesting blocks of DNA, such as plasmids, viruses and PATHOGENICITY ISLANDS, can be added to cells and tested for phenotypes that they have engendered. Isogenic cells can be compared for epigenetic effectors, such as changes in DNA methylation, histone acetylation, prion effects and so on. Useful information can also be gained by comparing non-isogenic cells such as multiple isolates from one species, pathogenic versus non-pathogenic strains and PASSAGED STRAINS versus recent natural isolates.

Limitations of large-scale phenotyping

An appealing aspect of phenotypic analysis is the simplicity and directness of its interpretation. For many applications, it relies on the validity of comparing the biology of isogenic strains, which has a substantial and proven record. A change in genotype leads to one or more changes in phenotype. To fully understand the function of a gene, we need to be able to enumerate and explain all the phenotypic changes that result from changes in that gene. But implicit in these types of isogenic analysis is that the cells, plants or animals that are being compared are truly isogenic (with the exception of the intended genetic change). This means that the genetic techniques that are used to create the strains must be precise, otherwise some of the phenotypes that are detected might be due to other unintended genetic differences between mother and daughter cell. As large-scale phenotyping technologies move towards the goal of measuring all the phenotypes of a cell, they also approach a level of sensitivity at which they can serve to assay the precision of genetic manipulation. In fact, in our work over the past two years using PM technology on *E. coli*, we have found several examples of strains that had been produced by proven genetic techniques that contained extraneous genetic changes. Even with accurate genetic manipulation, the accumulation of secondary suppressors in isogenic lines can be problematic. Two methods can be used to gain confidence that a phenotype is tied to a genotype: restoring the allele back to wild type and showing that the phenotypic change goes away, or assaying at least two independently constructed strains and showing that both isolates show the same phenotypic changes.

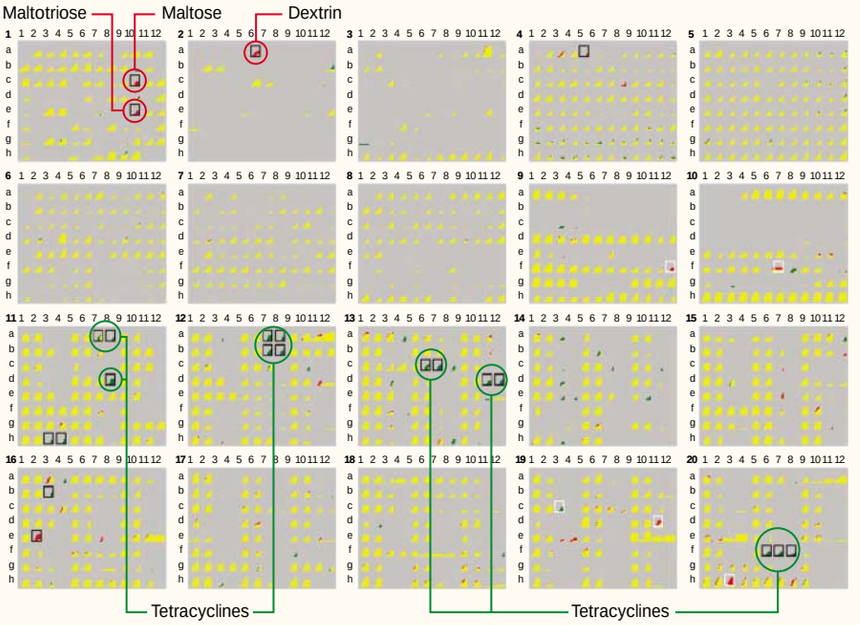


Figure 3 | **Phenotype MicroArray comparison of two isogenic strains of *E. coli*.** Phenotype MicroArray analysis of isogenic *E. coli* strains *E. coli* malF::tn10 versus MG1655. The mutant strain is shown in green and the parental MG1655 strain is shown in red. Knockout of the *malF* gene leads to the loss of catabolism of maltose, maltotriose and dextrin. Insertion of the TN10 CASSETTE leads to the gain of resistance to a number of tetracycline antibiotics.

An important impetus for the development of large-scale phenotypic analysis has been to determine the function of the remaining genes for which no function is known. We can expect these efforts to be partially successful — our limited experience, so far, using PM technology with genes of unknown function indicates a success rate of ~65% in *E. coli* (B.B. *et al.*, unpublished observations). When phenotypes are found, they can indicate anything from a precise enzymatic function to a vague allocation of the gene to an area of cellular physiology.

Surveying all of the phenotypes of a cell is a theoretical concept and goal. In reality, cells have too many phenotypes for us to be able to define, let alone test. It is important to acknowledge and be aware of the limitations of large-scale phenotyping techniques such as PM analysis. There are a number of reasons why, in its present form, PM analysis and other phenotyping technologies will not discover all of the phenotypes. First, the phenotyping sets available at present are not all-inclusive. For microbial cells, PM technology is likely to miss phenotypes that specifically involve intracellular structures (for example, the cytoskeleton and organelles) and surface structures and functions such as flagella, attachment, biofilm formation, motility and chemotaxis, as well as functions turned on only under anaerobic

Glossary

HETEROLOGOUS GENE
A gene that is transferred into a cell but originated in a cell from a different species.

ISOGENIC
Cells or organisms that are derived from the same parent and have almost identical genomes.

MASS SPECTROMETRY
An analytical tool for determining the molecular weight of a chemical.

MULTI-STATE AUTOMATON
A self-acting and self-responding machine that has the ability to change itself into multiple states.

PASSAGED STRAINS
Cells that have been repeatedly subcultured, typically under artificial *in vitro* laboratory-culture conditions and not in more natural *in vivo* conditions.

PATHOGENICITY ISLAND
A contiguous block of genes, found in pathogenic microorganisms, in which at least a subset of the genes code for virulence factors.

TETRAZOLIUM REDOX CHEMISTRY
A dye chemistry that absorbs the electrons produced by cellular respiration, causing a colour change as the tetrazolium dye is reduced.

TN10 CASSETTE
A contiguous block of genes that is derived from the bacterial transposon Tn10, which confers resistance to tetracycline antibiotics.

conditions. Also, it is not possible to test the cellular functions that still remain to be discovered — there are undoubtedly gaps in our knowledge of the whole spectrum of cellular functions. Second, the effects of some genes might be cryptic and only have a function under highly specific cellular conditions. Many microbial phenotypes might be expressed only when the microbe interacts with an animal or plant. And third, we might not be able to discern phenotypes for some genes because there are redundant cellular functions that compensate in their absence.

Phenotypic analyses are likely to provide an important complement to gene-expression and proteomic analyses of genetically altered cells. Molecular analyses enumerate a large number of biochemical changes, but cellular analyses show what these changes mean at the biological level. To illustrate the complementarity of these two approaches, consider the comparison of a mutant and wild-type cell tested by gene-expression analysis. Data from this analysis compares the level of thousands of genes under a single growth condition and state of the cell. The same comparison done by PM analysis looks at only a single gene, but under thousands of growth conditions and cell states. It is very important to appreciate that a cell line is not a single static entity. Every cell is a MULTI-STATE AUTOMATON with the capacity to change in minutes. A cell constantly senses its environment and adapts to changes by altering its gene-expression pattern, protein content, membrane constitution, surface receptors and so on. In each growth state, the cell becomes a different cell, sometimes markedly different. Understanding this fluid 'landscape' will challenge biologists for many decades to come. FIGURE 2 depicts PM analysis as a technology in the stream of, and complementary to, DNA microarrays and proteomic analysis. By using these genomic technologies, and others that are derived from and added to this set, we will continue to move our knowledge forward.

Phenotyping in higher eukaryotes

Here, I have emphasized studies of microbial cells as model systems. Other, more complex model systems, such as *Caenorhabditis elegans*, that can be cultured in microwells could be amenable to modified versions of these phenotypic technologies. Higher plants and animals could certainly be targeted next by adapting the technologies for cell cultures. We are working to extend PM technology to mouse and human cells, and prototype PMs for testing carbon and energy metabolism in human and mouse liver and blood cells have already been devised and successfully tested (A. Morgan, unpublished observations). In the near future, we will see the capability for simulating part of the metabolic and cell biology of a mouse or human by using the large-scale phenotypic analysis of cells derived from most of the main tissues. Another approach would involve the detailed phenotyping of embryonic stem cells in culture, to see how useful the cellular phenotypes are in predicting the phenotypes of animals. Relevant to this technology are the recent advances in gene inactivation using RNAi, which provide a method of specifically inactivating gene function. We forsee immediate applications of phenotypic analysis in genetics, physiology, toxicology and the study of ageing, differentiation and disease.

Barry R. Bochner is at Biolog, Inc., 3938 Trust Way, Hayward, California 94545, USA.
email: bbochner@biolog.com

doi: 10.1038/nrg1046

1. Miller, J. H. *A Short Course in Bacterial Genetics: A Laboratory Manual and Handbook for Escherichia coli and Related Bacteria* (Cold Spring Harbor Lab. Press, New York, 1992).
2. Blattner, F. R. *et al.* The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1474 (1997).
3. Goffeau, A. Life with 6,000 genes. *Science* **274**, 563–567 (1996).
4. Kaul, S. *et al.* Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815 (2000).
5. Goff, S. A. *et al.* A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* **296**, 92–100 (2002).
6. The *C. elegans* Sequencing Consortium. Genome sequence of *C. elegans*: a platform for investigating biology. *Science* **282**, 2012–2018 (1998).
7. Adams, M. D. *et al.* The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195 (2000).
8. Waterston, R. H. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
9. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
10. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
11. Nadeau, J. H. *et al.* Functional annotation of mouse genome sequences. *Science* **291**, 1251–1255 (2001).
12. Bassett, D. E. *et al.* Genome cross-referencing and XREFdb: implications for the identification and analysis of genes mutated in human diseases. *Nature Genet.* **15**, 339–344 (1997).
13. Rubin, G. M. *et al.* Comparative genomics of eukaryotes. *Science* **287**, 2204–2215 (2000).
14. LaRossa, R. A. in *Escherichia coli and Salmonella Cellular and Molecular Biology* 2nd edn Vol. 2 Ch. 39 (ed. Neidhart, F. C.) 2527–2587 (ASM, Washington DC, 1996).
15. Cheung, V. G. & Spielman, R. S. The genetics of variation in gene expression. *Nature Genet.* **32** (Suppl.), 522–525 (2002).
16. Jansen, R. C. Studying complex biological systems using multifactorial perturbation. *Nature Rev. Genet.* **4**, 1–7 (2003).
17. Rosenthal, N. & Ashburner, M. Taking stock of our models: the function and future of stock centres. *Nature Rev. Genet.* **3**, 711–717 (2002).
18. Eppig, J. T. Algorithms for mutant sorting: the need for phenotype vocabularies. *Mamm. Genome* **11**, 584–589 (2000).
19. Pargent, W. *et al.* MouseNet database: digital management of a large-scale mutagenesis project. *Mamm. Genome* **11**, 590–593 (2000).
20. Bochner, B. R. Sleuthing out bacterial identities. *Nature* **339**, 157–158 (1989).
21. Brown, S. D. & Peters, J. Combining mutagenesis and genomics in the mouse — closing the phenotype gap. *Trends Genet.* **12**, 433–435 (1996).
22. Paigen, K. & Epping, J. T. A mouse phenome project. *Mamm. Genome* **11**, 715–717 (2000).
23. Moldin, S. O. *et al.* Trans-NIH neuroscience initiatives on mouse phenotyping and mutagenesis. *Mamm. Genome* **12**, 575–581 (2001).
24. Beckers, J. & Hrabe de Angelis, M. Large-scale mutational analysis for the annotation of the mouse genome. *Curr. Opin Chem. Biol.* **6**, 17–23 (2001).
25. Somerville, C. & Dangl, L. Genomics: plant biology in 2010. *Science* **290**, 2077–2078 (2000).
26. Oliver, S. G. A network approach to the systematic analysis of yeast gene function. *Trends Genet.* **12**, 241–242 (1996).
27. Hampsey, M. H. A review of phenotypes of *Saccharomyces cerevisiae*. *Yeast* **13**, 1099–1133 (1997).
28. Niedenthal, R. *et al.* Systematic analysis of *S. cerevisiae* chromosome VII genes. *Yeast* **15**, 1775–1796 (1999).
29. Bochner, B. R. 'Breathprints' at the microbial level. *ASM News* **55**, 536–539 (1989).
30. Bochner, B. R., Gadzinski, P. & Panomitros, E. Phenotype MicroArrays for high-throughput phenotypic testing and assay of gene function. *Genome Res.* **11**, 1246–1255 (2001).
31. Vazquez-Boland, J. A. *et al. Listeria* pathogenesis and molecular virulence determinants. *Clin. Micro. Rev.* **14**, 584–640 (2001).
32. Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177**, 4121–4130 (1995).
33. Ji, Y. *et al.* Identification of critical staphylococcal genes using conditional phenotypes generated by antisense RNA. *Science* **293**, 2266–2269 (2001).
34. Smith, V. *et al.* Functional analysis of the genes of yeast chromosome V by genetic footprinting. *Science* **274**, 2069–2074 (1996).
35. Shoemaker, D. D. *et al.* Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. *Nature Genet.* **14**, 450–456 (1996).
36. Winzeler, E. A. *et al.* Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901–906 (1999).
37. Giaever, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387–391 (2002).
38. Rieger, K. J. *et al.* Large-scale phenotypic analysis — the pilot project on yeast chromosome III. *Yeast* **13**, 1547–1562 (1997).
39. Rieger, K. J. *et al.* in *Methods in Microbiology* Vol. 28 (ed. Craig, A.G.) 205–227 (Academic Press, London, 1999).
40. Rieger, K. J. *et al.* Chemotyping of yeast mutants using robotics. *Yeast* **15**, 973–986 (1999).
41. Entian, K. D. *et al.* Functional analysis of 150 deletion mutants in *Saccharomyces cerevisiae* by a systematic approach. *Mol. Gen. Genet.* **262**, 683–702 (1999).
42. Ross-MacDonald, P. *et al.* Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature* **402**, 413–418 (1999).
43. True, H. L. & Lindquist, S. L. A yeast prion provides a mechanism for genetic variation and phenotypic diversity. *Nature* **407**, 477–483 (2000).
44. Sakumoto, N. *et al.* A series of double disruptants for protein phosphatase genes in *Saccharomyces cerevisiae* and their phenotypic analysis. *Yeast* **19**, 587–599 (2002).
45. Warringer, J. & Blomberg, A. Automated screening in environmental arrays allows analysis of quantitative phenotypic profiles in *Saccharomyces cerevisiae*. *Yeast* **20**, 53–67 (2003).

Acknowledgements

The author gratefully acknowledges and thanks his colleagues that have participated in the development of Phenotype MicroArray technology: X. H. Lei, A. Franco-Buff, R. Kostriken, J. Argyle, L. Wiater, J. Carlson, A. Morgan, C. Gorman, P. Gadzinski, E. Olender, E. Panomitros and L. He. We are also grateful for partial financial support of this project by Small Business Innovation Research Grants from the National Institutes of Health/National Institute of General Medical Sciences and from the National Aeronautics and Space Administration

Online links

DATABASES

The following terms in this article are linked online to:
Entrez: http://www.ncbi.nlm.nih.gov/entrez
malF | *prfA*

FURTHER INFORMATION

Deltagen, Inc.: http://www.deltagen.com
Lexicon Genetics, Inc.: http://www.lexicongenetics.com
Paradigm Genetics, Inc.: http://www.paradigmgenetics.com
Phenomix Corporation: http://www.phenomixcorp.com
SurroMed, Inc.: http://www.surromed.com
Access to this interactive links box is free online.