Grass-roots genomics

Mark Johnston[1] & Stanley Fields[2]

[1]Department of Genetics, Washington University Medical School, 4566 Scott Avenue, Saint Louis, Missouri 63110, USA. [2]Howard Hughes Medical Institute, Departments of Genetics and Medicine, University of Washington, Seattle, Washington 98195, USA. e-mail: mj@genetics.wustl.edu and fields@u.washington.edu

In these heady days of genomic enterprise, the wet-bench geneticist may wonder what the future holds in store for those who explore biology in the laboratory. We argue that the goal of genomics is to serve geneticists.

What most biologists need is genes, not genomics. If genomicists rise to this challenge, they will succeed in eliminating their field as a distinct area of study. What will follow is grass-roots genomics—with laboratories of any size incorporating genomic strategies into their *modus operandi.*

Consider the path taken by other innovative technologies of the biological sciences. In the late 1970s, recombinant DNA methods and DNA sequencing were in their infancy, and the province of a few specialized laboratories. This was a time when a postdoc might land a job purely on the basis of knowing how to clone a gene or determine its sequence. But these technologies quickly permeated the community, their spread facilitated by detailed how-to manuals, commercial availability of enzymes, reagents, and ultimately step-by-step kits, along with practical courses and dispersal of scientists with the necessary expertise. Whereas dedicated sequencing laboratories have reappeared on the landscape—albeit on a scale unimaginable twenty years ago—adequate sequencing capability is within the reach of biologists. Anyone now arriving at a job interview presenting cloning credentials is far more likely to be expert in transforming mammalian eggs into embryos than bacterial colonies to ampicillin resistance.

If research on *Saccharomyces cerevisiae* serves as a harbinger of the study of other organisms, it is instructive to consider the reagents that will soon be available to the average yeast biologist, who strives to discern the workings of an organism with only 6,000 genes. Residing in trays in the laboratory freezer will be a set of PCR products corresponding to every predicted open reading frame (ORF) in a form that enables facile cloning into any plasmid vector[1]; a set of 6,000 strains, each lacking a specified ORF (ref. 2); a set of plasmids that allows ablation of any gene in other strains; and sets of strains overproducing each protein fused to either an epitope-tag[3], a visible tag (for example, green fluorescent protein), a purification hook[4] (such as glutathione S-transferase), or a DNA-binding or transcriptional activation domain[5]. Stored in not-quite-so-cool a climate will be trays of slides containing tiny spots of DNA (microarrays) that enable the rapid quantitation of transcript levels for every gene[6,7], or the detection of any gene whose alteration is responsible for an observable phenotype[8,9]. If not available in the biologist's own laboratory, these reagents and the equipment to handle them will be available within the department or university, or, at worst, from a facility not far away.

Robin Lovell-Badge

What these reagents are, in essence, is genes—known genes and unknown genes, regulatory genes and structural genes, regulated genes and constitutive genes—all the genes of the organism. Whereas each one of these sets of genes is not yet available to yeast researchers, some have already arrived; the rest will be here soon. (See, for example, refs 10,11.) The only limiting factors will be the cost of the genes and the availability of machines—the robots, arrayers and detectors—to deal with them. As costs go down (or grant funds go up), these tools should rapidly spread throughout the yeast community.

Students beginning their studies of yeast biology take as their birthright the complete genome sequence and plan their experiments accordingly. The awareness that more genomic tools are just over the horizon is already having a palpable effect on the kinds of projects envisioned. And whatever heretofore inconceivable experiments now being contemplated for *S. cerevisiae* with its 6,000 genes can surely be imagined for *Caenorhabditis elegans*, *Drosophila melanogaster* or *Arabidopsis thaliana*, organisms with three- or fourfold more genes. Imagined not by genome scientists in Saint Louis, Stanford or the two Cambridges, but by students in large and small laboratories in universities everywhere. And can the gene sets of mice and humans, perhaps another four- or fivefold larger than those of worms and flies, be that much more difficult to generate and manoeuvre? The goal of genomicists should be to convert this promise into reality, sooner rather than later, by making their methods accessible to all.

Putting the techniques of today's genomicists within reach of the rank-and-file biologist is essential if functional genomics is to fulfil its much-heralded promise. This dispersal is just beginning to happen; a few relatively small laboratories, mostly through independent effort, have acquired the ability to measure gene expression with DNA microarrays[12–14], a technology formerly accessible only to genomicists[6,7]. Advances in DNA microarray fabrication techniques[15,16] promise to reduce the cost of arrays, which should increase their availability to the general community.

But as the din of genomics grows ever louder, a vital note must be sounded above the noise. The independent investigator in the trenches of academia is needed now more than ever, because genomics cannot elucidate gene function:

© 2000 Nature America Inc. • http://genetics.nature.com

© 2000 Nature America Inc. • http://genetics.nature.com

it can only find genes. And the genomicists are placing heaping servings of genes on the platter, the functions of which must be determined if we are to realize the goal implicit in the genome projects: to understand how the products of each of those genes collaborate to construct an organism. Achieving this goal will require the expertise and insight that can be brought to bear on biological problems only by a person with a deep and abiding interest in the particular process(es) executed by a set of genes. Genomic techniques can provide clues to gene function[17], but the information they provide is usually too meagre to be conclusive. These clues can lead down the path of discovery only if followed by someone with the experience and intuition to discern the direction in which they are pointing. It is only in this way—gene-by-gene, process-by-process, researcher-by-researcher—that we will reach the goal of 'solving' an organism. To realize this goal, researchers must get their hands on the gene sets and tools necessary to interrogate them. As the past is our guide, we expect this will happen sooner, rather than later. □

1. Hudson, J.R., Jr. *et al. Genome Res.* **7**, 1169–1173 (1997).
2. Winzeler, E.A. *et al. Science* **285**, 901–906 (1999).
3. Burns, N. *et al. Genes Dev.* **8**, 1087–1105 (1994).
4. Martzen, M.R. *et al. Science* **286**, 1153–1155 (1999).
5. Uetz, P. *et al. Nature* (in press).
6. Derisi, J.L., Iyer, V.R. & Brown, P.O. *Science* **278**, 680–686 (1997).
7. Wodicka, L., Dong, H., Mittmann, M., Ho, M.H. & Lockhart, D.J. *Nature Biotechnol.* **15**, 1359–1367 (1997).
8. Winzeler, E.A. *et al. Science* **281**, 1194–1197 (1998).
9. Cho, R.J. *et al. Proc. Natl Acad. Sci. USA* **95**, 3752–3757 (1998).
10. http://sequence-www.stanford.edu/group/yeast_deletion_project/deletions3.html
11. http://ycmi.med.yale.edu/YGAC/home.html
12. Pearce, D.A., Ferea, T., Nosel, S.A., Das, B. & Sherman, F. *Nature Genet.* **22**, 55–58 (1999).
13. Cox, K.H., Pinchak, A.B. & Cooper, T.G. *Yeast* **15**, 703–713 (1999).
14. Lelivelt, M.J. & Culbertson, M.R. *Mol. Cell. Biol.* **19**, 6710–6719 (1999).
15. Blanchard, A.P. & Friend, S.H. *Nature Biotechnol.* **17**, 953 (1999).
16. Singh-Gasson, S. *et al. Nature Biotechnol.* **17**, 974–978 (1999).
17. Eisen, M.B., Spellman, P.T., Brown, P.O. & Botstein, D. *Proc. Natl Acad. Sci. USA* **95**, 14863–14868 (1998).

© 2000 Nature America Inc. • http://genetics.nature.com

Mediating mismatch repair

Josef Jiricny

Institute of Medical Radiobiology of the University of Zurich and the Paul Scherrer Institute, CH-8008 Zurich, Switzerland. e-mail: jiricny@imr.unizh.ch

The DNA repair picture in humans becomes more complete with the identification of MLH3, a homologue of MutL and a heterodimeric partner of MLH1.

In *Escherichia coli*, the principal players in post-replicative mismatch repair (MMR)—MutS and MutL—are well characterized. The MutS protein is a mismatch recognition factor, whereas MutL mediates cross-talk between MutS and the other members of the repair complex—exonucleases, helicase(s) and polymerase(s)[1]. The MMR system is highly conserved; MutS and MutL functions are retained in eukaryotes. But the situation in higher organisms is complicated by the fact that six MutS homologues (MSH) and numerous MutL homologues (designated either MLH or PMS, owing to their link with post-meiotic segregation in yeast) have been identified. Not surprisingly, it has taken some time to elucidate who does what, and with whom. On page 27, Steven Lipkin *et al.*[2] now provide additional insight into MutL homologue function with their identification and analysis of MLH3, the most recently discovered member of the human MMR family. The authors also note that mouse *Mlh3* lies within the colon cancer susceptibility locus *Ccs1*, which is intriguing given that mutations in some (but not all) MMR genes segregate with hereditary non-polyposis colon cancer (HNPCC).

Sharing partners

In human cells, MMR is mediated by MSH2, MSH3, MSH6, MLH1 and PMS2 (ref. 3). Mismatch recognition in human cells is carried out principally by the heterodimeric combination of MSH2 and MSH6; this complex acts in the repair of mismatches (for example, G:T) and of insertion and deletion loops (IDLs), which can arise during DNA replication through slippage of the primer strand on the template[4]. A second, less abundant, heterodimeric complex composed of MSH2 and MSH3 can mediate the repair of IDLs in the absence of MSH6 (see figure). The loss of MSH2 therefore leads to the accumulation of point mutations and frameshift mutations resulting from unrepaired IDLs, detectable as high microsatellite instability (MSI-H), a phenomenon affecting repeated sequence motifs such as AA[A]nAA or CA[CA]nCA, in which eukaryotic DNA polymerases often 'stutter'[4]. MSI-H is a marker of classical HNPCC tumours. Due to the redundant function of MSH3 and MSH6 in IDL repair, cells lacking the former have no apparent phenotype, whereas cells lacking the latter accumulate predominantly point mutations; germline *MSH6* have been identified in only a few families with 'atypical' HNPCC, with MSI in tumours limited to mononucleotide repeats or absent. A similar situation applies to mice: *Msh2*$^{-/-}$ animals develop colon tumours, whereas *Msh6*$^{-/-}$ animals do not[6] or do so only later in life[7]. Mice with mutations in *Msh3* are not predisposed to cancer, but those lacking Msh3 and Msh6 develop colon tumours owing to total inactivation of the MMR process by elimination of both cognate partners of Msh2 (ref. 6).

As with MSH2 in mismatch recognition, MLH1 is the kingpin of the MutL homologue world. Interactions between MLH1 and PMS1 (ref. 8) and now MLH3 (ref. 2) have been reported, but only the MLH1/PMS2 heterodimer (ref. 9) has been shown to participate in mismatch repair (see figure). The notion that MLH1 might have a partner other than PMS2 was originally based on data from mammalian systems—in HNPCC families, *MLH1* is frequently mutated, whereas only two families with *PMS2* anomalies have been described. Because Mlh3p (of *Saccharomyces cerevisiae*) forms a heterodimer with Mlh1p that acts in combination with Msh2p/Msh3p in the repair of a subset of IDLs (ref. 10), there is reason to believe that MLH3 also acts in IDL repair in human cells (see fig-