Redefining genomics

J. David Grausz

Advances in genome sciences have occurred so rapidly that it is difficult to keep track of developments. In this lightning tour of the technologies and active companies, the author explores the implications of the rapid evolution of genomics for the pharmaceutical and diagnostics industries.

Both the number of drug targets and the cost of pursuing a drug specific to a particular target have increased almost exponentially over the past decade (from $54 million dollars per molecule in 1976 to more than $450 million today). Concentrating upon the 100 or so complex diseases that are the most frequent causes of medical intervention in Europe, Japan and North America (diseases such as heart disease, breast or prostate cancer, and noninsulin-dependent diabetes), genomics has massively extended the number of molecules to be tested by identifying potential new gene targets and surrogate physiological modifiers. Clearly, the challenge for the research laboratories, biotech (including genomic) companies and the pharmaceutical industry is to select the most relevant targets, and translate that choice into effective, safe and rapid treatments.

High-throughput methods have to be applied to functional genomics early in the chain of drug discovery, to identify the best gene targets to pursue. To pursue function one can either catalog what is already there, working at the level of protein and mRNA, or alter the instructions (possibly search for 'natural' alterations), working at the level of the genome.

Technological breakthroughs – who is doing what?

Altering function

Functional Genomics I dissects the role of a known sequence. There are several model systems that provide such information (see Table 1):

- The most sophisticated examples come from work with transgenic mice. Here one can actually talk of 'chromosome' engineering (for recent examples, see Refs 1,2; transgenic/targeted mutation database, see Ref. 3; *cre–loxP*, see Ref. 4; conditional expression in transgenics, see Ref. 5).
- Knock-out or double knock-in (amplification) of genes in yeast can be very informative and has provided proof by analogy of function and/or mechanism[6,7].
- Zebrafish (Figure 1) are an excellent source of embryological information concerning the development of the CNS. They reproduce rapidly and are transparent permitting easy identification of malfunctions[8].

Functional Genomics II details the role of an unknown sequence. Again, mice are extremely useful because of their tendency to conserve linkage groups with man. Lexicon Genetics (The Woodlands, TX, USA) have developed a rapid method of generating desired transgenic mice, and their recently concluded $8 million collaboration with the Merck Genome Research Institute (Whitehouse Station, NJ, USA) will make 150 to-be-selected transgenics available at cost to the research community. Their newly created OmniBank library will eventually include 500,000 murine embryos harboring mutations in virtually every gene in the genome. The possibility to target expression to particular cells[3,4] or delay expression of defects using promoters sensitive to rapamycin[5] greatly increases the ability to pinpoint function in transgenic mice.

The fact that many of these observations and/or alterations can first be made on mammalian cells in culture has not escaped genomic biotechs:

- Oncogene Science (Uniondale, NY, USA), an off-shoot of research at Cold Spring Harbor, has a key patent for the use of reporter genes (e.g. luciferase) to detect transcription factors and drugs that modulate them. The company

J. David Grausz, 50 Avenue Mathurin Moreau, 75019 Paris, France. fax: +33 1 44 05 1970, e-mail: jd.grausz@chu-stlouis.fr

  1359-6446/98/$19.00. PII:S1359-6446(97)01127-6

has established drug discovery deals with Wyeth-Ayerst (Princeton, NJ, USA) and Hoechst (Frankfurt-am-Main, Germany; division of Hoechst Marion Roussel), and has just launched a joint venture with Roche (Basle, Switzerland) called Helicon Therapeutics. Oncogene Science already had a coventure with Pfizer (New York, NY, USA).

Aurora Bioscience's (La Jolla, CA, USA) recent successful IPO is a good indicator. Their partners, including Sequana (La Jolla, CA, USA: merged with Arris on 3 November 1997 to form Axys Pharmaceuticals), Alanex (San Diego, CA, USA; recently acquired by Agouron), ArQule (Medford, MA, USA), Bristol-Myers Squibb (Princeton, NJ, USA) and Roche Biosciences (Palo Alto, CA, USA), want to use their expertise with fluorescence detection of relevant proteins in cultured mammalian cells to dissect function.

Genotyping

Knowing the genotype has many applications. First, and most obvious, it is necessary to diagnose patients at risk for disease: mutations in *BRCA1* and *BRCA2* indicate susceptibility to breast cancer, for example[9]. Second, the genotype can indicate which treatments are appropriate to which patients[10]. Certain treatments will cure only a subpopulation of those suffering from a disease, such as the genetically homogeneous group in which hypertension can be mastered by following a low-salt diet (work of Myriad Genetics, Salt Lake City, UT, USA). Moreover, some patients are more likely to suffer complications, such as diabetes-associated nephropathy or retinopathy, or be prone to metastasis and, consequently, have to be treated more rapidly and severely. Third, genetic background can be crucial to interpreting susceptibility to environmental influences (e.g. radiation or stress), drugs, allergies and even putative genetic defects. It determines the degree to which genes interact with each other and with the environment.

Rats, cystic fibrosis transmembrane conductance regulator (CFTR) and 'phenomics'

Dr Howard Jacob (University of Wisconsin, Madison, WI, USA) terms the study of the influence of 'secondary' genes 'phenomics'. He is applying high-throughput genotyping and cDNA correlation methods to identify culprit genes in rats and find their human equivalents. Because the influence of such loci is small, large populations must be studied, and inbred rodents are ideal. There are eight strains of naturally hypertensive rats, most of which can be ascribed to mutations in both copies of a single culprit gene (i.e. recessive inheritance). Heterozygote rats and/or those not bearing 'fully functional' alleles at the culprit, or other, related, loci are more susceptible to hypertension either when the environment (e.g. stress or diet) or the genetic make-up are altered. However 'phenomics' is not confined to hypertension. The genetic background of some CFTR-defective individuals protects them from ever developing cystic fibrosis. Similarly, alcohol or cigarettes are more dangerous to some individuals than to others.

Correlating sequence with function

A battery of tests is under development to correlate sequence with function, analyzing both human and nonhuman expression patterns. These tests must be robust, universally applicable (even when studying samples of mutant tissue diluted 10–100-fold by normal tissue, as in the case of cancer), accurate and rapid. In certain applications it is desirable to scan kilobases of genomic DNA and/or its coded mRNA for discrepancies. It can be important to determine the presence and/or quantity of a protein or its transcript. A profile of transcripts is useful, especially when it can be correlated with proteins expressed. Two biotech companies focus on correlating genotype with treatment effectiveness: Genaissance (New Haven, CT, USA) and Genset (Paris, France); see below.

Table 1. Model systems in functional genomics

Organism	Model	Observe	Application	Refs
Fugu fugu	Genome	Organization, sequence	Dystrophin	16
Saccharomyces cerevisiae, yeast	Double hybrid (A+B)	A = DNA binding B = catalytic domain	Identify proteins that interact, and new members of known families	6,7
Zebrafish	Development	Visible alterations	Central nervous system	8
Transgenic mouse:				
– knock out	Gene function	Embryo and adult	Telomerase	17,18
– *cre–loxP*	Gene function	Tissue-specific knockout	Developing T cells in thymus	4
– induction	Gene function	Delayed expression	Rapamycin	5

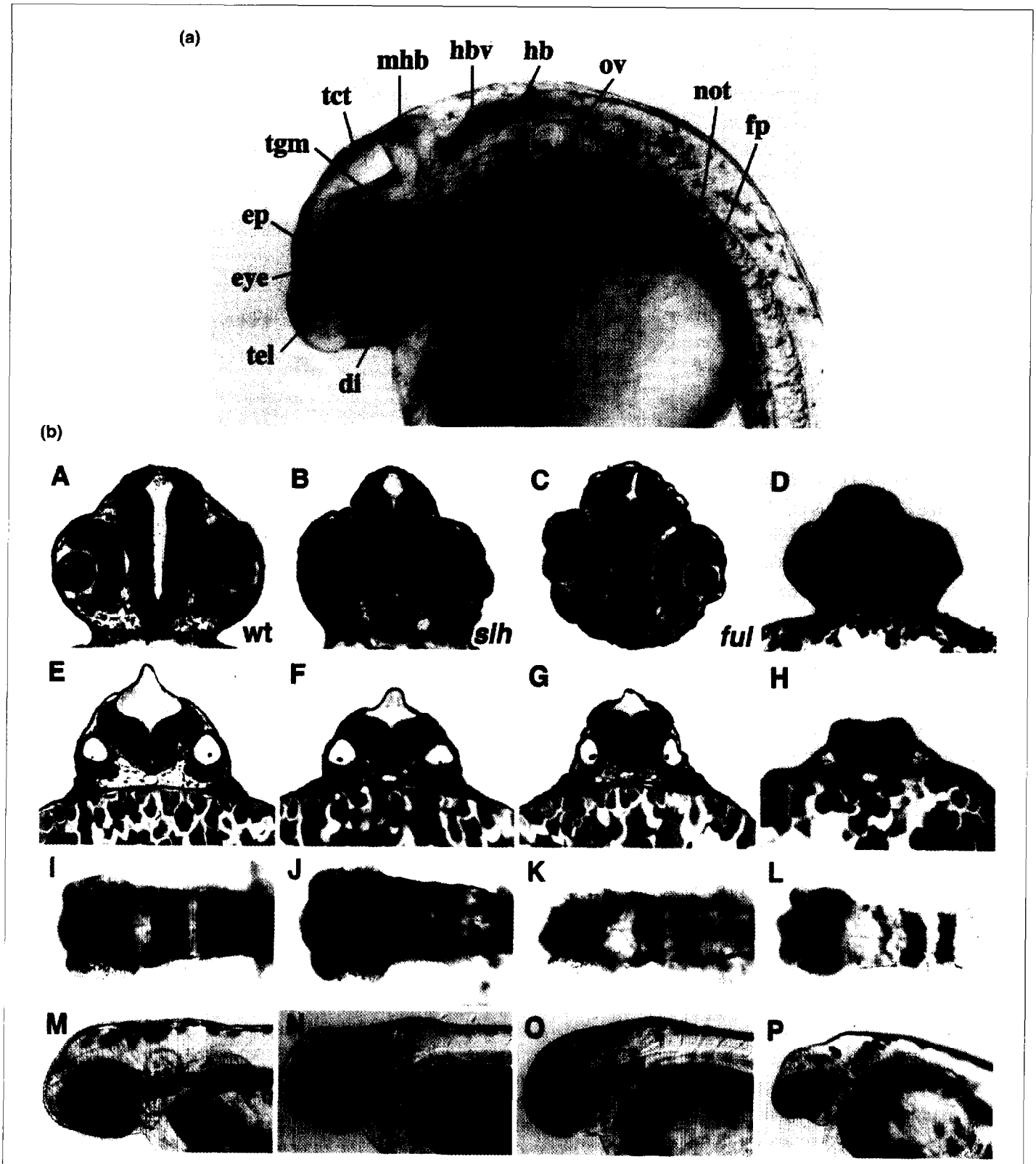

Figure 1. *Zebrafish – an excellent model for brain development. (a) Morphology of the embryonic brain (28 h postfertilization). (b) Transparent embryo permits the distinction of anomalies in brain development, specifically ventricle enlargement (four genetic anomalies are shown). Courtesy of Prof. W. Driever (reproduced, with permission, from Ref. 8).*

Chip technology

One of the better known methods employs oligonucleotides arrayed on solid surfaces, DNA chips, beads or slides. A few examples are considered below.

Affymetrix (Palo Alto, CA, USA) are the closest to commercializing their product: GeneChip®. Recent articles attest to the ability of these chips to detect mitochondrial mutations (scanning 16 kb of mitochondrial genome)[11] and mutations across 3 kb of the expressed part of the *BRCA1* gene[12]. In a collaboration with Incyte (Palo Alto, CA, USA) they are looking for the frequency of proprietary transcripts in normal and pathological tissue samples. With Genetics Institute (Cambridge, MA, USA), and with data shared by Bayer (Leverkusen, Germany), Chiron (Emeryville, CA, USA), Kirin Brewery (Tokyo, Japan) and Genentech (South San Francisco, CA, USA), in the DiscoverEase program they are looking at the functional significance of several thousand putative secreted and receptor protein transcripts. Over the past 2 years the Affymetrix technology has evolved not only in terms of interactive, analytical software, but also (and especially) in 'chipware'. From a 50 μm resolution Molecular Dynamics (Sunnyvale, CA, USA) system (65,000 sequences per chip) to a 20 μm resolution Hewlett Packard (Palo Alto, CA, USA) system (400,000 sequences per chip), they are currently packing a million short (25mers) oligonucleotide sequences on a 1.28 mm × 1.28 mm chip (Figure 2).

Another chip-based strategic player, Hyseq (San Diego, CA, USA), was one of the first to propose using oligonucleotide arrays (8mers) to sequence and/or compare cDNA libraries. They presently have collaborations with Chiron and Perkin-Elmer (Foster City, CA, USA) to apply their technology to both cDNA and genomic sequencing and increase the utility of the

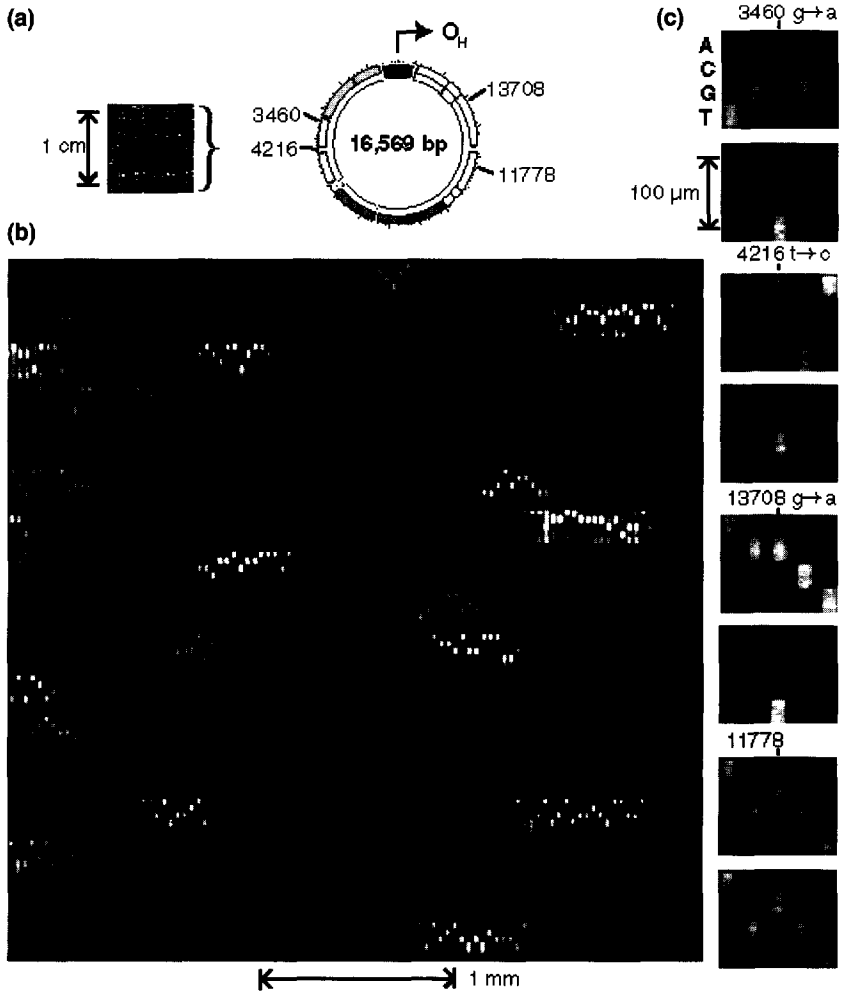

Figure 2. *Human mitochondrial genome on a chip (courtesy of Affymetrix, Inc.). (a) An image of the array hybridized to 16.6 kb of mitochondrial target RNA (L strand). The 16,569 bp map of the genome is shown, and the H strand origin of replication (O_H), located in the control region, is indicated. (b) A portion of the hybridization pattern magnified. In each column there are five probes: A, C, G, T, and Δ, from top to bottom. The Δ probe has a single base deletion instead of a substitution and hence is 24 instead of 25 bases in length. The scale is indicated by the bar beneath the image. Although there is considerable sequence-dependent intensity variation, most of the array can be read correctly. The image was collected at a resolution of ~100 pixels per probe cell. (c) The ability of the array to detect and read single base differences in a 16.6 kb sample is illustrated. Two different target sequences were hybridized in parallel to different chips. The hybridization patterns are compared for four different positions in the sequence. Only the P25, 13 probes are shown. The top panel of each pair shows the hybridization of the mt3 target, which matches the chip P0 sequence at these positions. The lower panel shows the pattern generated by a sample from a patient with Leber's hereditary optic neuropathy (LHON). Three known pathogenic mutations, LHON3460, LHON4216, and LHON13708, are clearly detected. For comparison, the fourth panel in the set shows a region around position 11,778 that is identical in both samples.*

database HyGene (both companies have equity investments in HySeq). For the year 2001, they are preparing a new automated strategy for genotyping and information access for SmithKline Beecham (King of Prussia, PA, USA) with the help of Orchid Biocomputer (Palo Alto, CA, USA).

GeneLogic (Columbia, MD, USA) profiles transcripts in tissue samples and cell cultures, correlating these signatures with development and disease. The result is a time-lapse recoding of cDNA evolution. They detect the mRNA binding to their chips using a fluorescent oligo-dT probe. They have academic partnerships in the fields of osteoporosis, prostate cancer, schizophrenia and Alzheimer's disease in order to enhance their database. They have also used their multiplexing technology to identify key control elements, transcription factors and their binding sites. Their first deal was recently concluded with Proctor & Gamble (Cincinnati, OH, USA) for more than $75 million. Digital Technologies (La Jolla, CA, USA) uses a proprietary solid-phase polymerase chain reaction (PCR) hybridization system (called TOGA, total gene analysis) to identify unusual messenger patterns and correlate them with tissue and disease. They have an alliance with Recordati SpA (Rome, Italy) in atherosclerosis.

Lynx Pharmaceuticals (Hayward, CA, USA) generates multiple parallel sequences of available transcripts. These sequences can then be used to generate probes (that could eventually be bound to a solid surface or chip) to define the functional status of a tissue. Their strategic partners include Hoechst and BASF (Ludwigshafen, Germany). Very similar to Lynx, the biotech company Synteni (Fremont, CA, USA) has introduced a two-color fluorescence method for comparison of the cDNA derived from two closely related cell populations. They expose their arrays of approximately 100,000 probes first to one cDNA source (labeled with green fluorescence) and then to another (labeled with red fluorescence) to indicate sequences for which over- or underexpression could correlate with a disease state.

Visible Genetics (Toronto, Canada) is also developing a solid-phase hybridization system (OpenGene®) for rapid detection of mutations in the clinic. They recently filed a patent for their p53 mutation detection system.

Software is a very important component in all cases, both to analyze and to array the data obtained. It is, however, important to note two limitations of chips. Firstly, they are not designed to deal with repetitive families of sequences, and repetitions can be very important in disease (e.g. the CNG triplet amplifications seen in several nervous disorders including Huntington's disease). Secondly, with few exceptions, such as Synteni, they are designed to work best on normalized cDNA libraries, where each message is represented an equal number of times. Thus, the original messenger frequencies can no longer be measured.

Matching and mismatching

If a culprit mutation or transcript is known (diagnostic applications such as the δ508 mutation in CFTR or the detection of HIV-coded messenger in AIDS) there are various methods of monitoring it other than oligochips. The TAQ-Man approach is probably the most sensitive. This has been developed in a sealed-tube format by scientists at ABI-Perkin-Elmer to screen for 25 well-known mutations and to detect minute quantities of infectious viral DNA (e.g. HIV). The method monitors probe degradation (accompanied by fluor activation; i.e. release from quenching because the probe contains coupled fluorescent and quenching moieties that are separated by degradation; see Figure 3a), which occurs, during PCR amplification, only when the target sequence (e.g. a mutation or an HIV-specific mRNA) is present.

Gen-Probe (San Diego, CA, USA) has a similar, but somewhat more robust, method of genotyping, measuring the disappearance of chemoluminescence (aeridium ester degradation) from a DNA probe hybridized near the site of a mutation. True hybridization protects the signal, which is hydrolyzed when there is a mismatch within 5 bp of the probe – a hybridization protection assay. They have also developed an isothermal transcription-mediated amplification protocol, which shows higher fidelity than standard PCR. Tests on tumor-derived material give 97–98% accuracy even when 0.1% of the tissue is estimated to be affected. An alternative method developed by Behring Diagnostics (Marburg, Germany), who recently entered into the Dade-Behring, Inc., joint venture, relies on detection of the strong fluorescent signal generated in the presence of a mutation, when two weakly fluorescent primer-generated products associate. Finally, oligoligase amplification systems (developed by Uwe Lundegren and Lee Hood, University of Washington, Seattle, USA) have proven effective in detecting as many as ten mutations simultaneously in routine patient screening[13]. A ligase-based method on chips is being exploited by HySeq in collaboration with ABI-Perkin-Elmer.

Other methods based upon mismatch repair are more effective in scanning for unknown discrepancies (or identifying new single basepair polymorphisms) and counting repeat amplifications or deletions such as those critical in inherited nervous disorders. Ambion (Austin, TX, USA) and Avitech

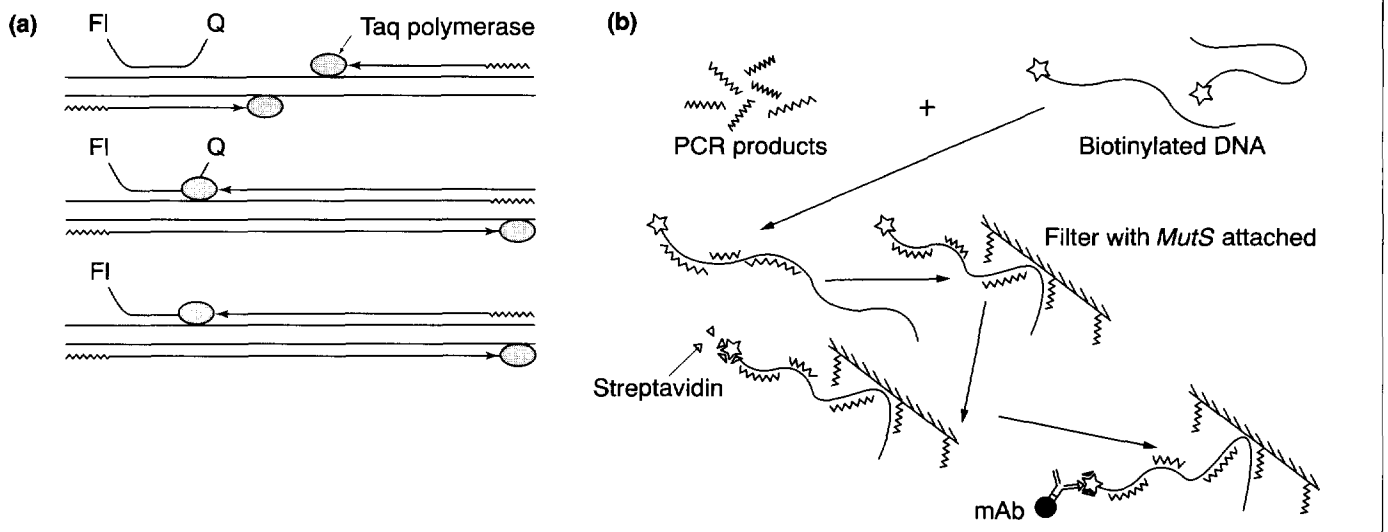

Figure 3. *Schematic diagram of three methods of genotyping. (a) In the TAQ-Man method, a quenched fluorescent probe is used, in which a group Q on one end of the DNA probe prevents the fluorescent group Fl on the other end from fluorescing when bound to its target sequence. Performing PCR on a template that has been annealed to this quenched probe causes the probe to be cut by the advancing Taq polymerase, allowing the fluorescence to be detected. (b) Using* MutS *on a solid support allows the selection of DNA molecules from solution that fail to bind to the probe PCR products; binding to the filter only occurs when there is a 1–2 bp mismatch between the probes and the biotinylated DNA. Once DNA is bound, the filter can be washed and the binding of biotinylated DNA detected using streptavidin and fluorescence or* ^{125}*I-labeled monoclonal antibody (mAb). (c) Using the SpectroChip® oligomer chip: test DNA serves as the template in a polymerization reaction with only 3 of the 4 dNTPs (e.g. dATP, dCTP, dTTP). Polymerization is complete when the first polymorphism (requiring the fourth dNTP, which is only available as the dideoxy form, e.g. ddGTP) is reached, or a microsatellite (e.g. repetition of CA, CT or CAT) ends. The extended oligomers are then removed from the chip, using a laser pulse and their lengths analysed by MALDI-TOF mass spectrometry.*

(Philadelphia, PA, USA) use two very different enzyme cocktails (RNase and DNA endonuclease 7, respectively) to excise and detect mispaired base(s). Avitech's kit, the enzyme mutation detection (EMD®) assay, has been shown to be almost 100% accurate in double-blind tests of potential colon cancer patients (whose status was known by extensive sequencing). It is good for mutation/polymorphism scanning because it can indicate the exact site of the mutation, can detect <5% mutant tissue mixed with 'normal' tissue (e.g. in cancer patients) and has routinely been used to scan 3 kb fragments of DNA.

Genecheck (Fort Collins, CO, USA) detects mismatches between a biotinylated probe and PCR products of genomic DNA by taking advantage of the selective binding of immobilized *MutS* (*Escherichia coli*) to 1–2 bp mismatches (Figure 3b). The fact that the *MutS* is attached to a solid support reduces background substantially[14,15]. Sequenom (San Diego, CA, USA) have taken advantage of advances in mass spectrometry (MALDI-TOF) to detect mutations and polymorphisms as length variations in the products of oligomers that fail to extend across the site of a putative mutation (Figure 3c).

Sensitivity

Both the chip and mismatch detection techniques are designed for high-throughput application. They are meant as shortcuts to labor-intensive sequencing and denaturing gel-based technologies. The major stumbling block is sensitivity. In order to

Table 2. Diagnostics partnerships

Company	Partner	Focus	Date	Deal value ($ million, up front)
Abbott	Genset	Drug response genes	July 1997	42.5
Amersham Pharmacia	Avitech	EMD scanning	September 1997	ND
	Molecular Dynamics	Microarrays	July 1997	ND
	Sequana	Microarrays	July 1997	ND
Behring Diagnostics	Dade International	Acquisition	July 1997	ND
Bristol-Myers Squibb	Millennium & Whitehead		July 1997	40
Eli Lilly	Myriad		July 1995	ND
Roche	Corange (Boehringer-Mannheim)	Acquisition	May 1997	11
	Oncogene Science	Helicon (joint venture)	July 1997	ND
SmithKline Beecham	Incyte (will give $10m)	Form diaDexus as 50:50 partners		25 (10+15)

ND = not disclosed

ensure that absolutely no false-positive diagnoses are made, or overoptimistic predictions of physiological/pathological relevance proposed, backgrounds are reduced radically. Inevitably this raises the level of false negatives, overlooking real positives and true functional candidates. BioTraces (Fairfax, VA, USA) have invented a solution that increases sensitivity of both chip-based and mismatch repair genotyping. They can detect several thousandfold less probe, and the system can be adapted to most of the methods already described, as long as fluorescence is replaced by an isotope (can be low specific activity) of radioactive phosphorus, iodine or other high-energy emitter. The technology permits double, triple or quadruple labeling.

Using these more sensitive detection strategies offers three distinct advantages: speed of detection is enhanced; sample size is reduced; and the accuracy of the result is increased. Thus, difficult genotypes can be determined and the expression of minority transcripts monitored. Repetitive sequences can be counted because backgrounds can be reduced. Quantitative measurements of relatively low cDNA levels are possible; thus, cDNA libraries do not have to be normalized. Finally, *in vivo* studies at the level of individual cells are possible because the levels of radiation used are those encountered in the environment.

Two other fields will benefit from such enhanced detection techniques. Firstly, monitoring proteins, in parallel with RNA and DNA. Protein cataloguing, correlation with corresponding mRNA, nuclear versus cytoplasmic localization of the proteins expressed (using, for example, microdissection techniques or quantitative 2D gel analysis) deserve attention. Secondly, long-range PCR (across 30 or more kb) products can be quantified, vastly improving the scanning range of the Avitech and GeneCheck approaches, for example.

Partnerships point to future directions

The safe, effective and rapid molecules that evolve, for these are the three criteria applied by the patent office to assure novelty, are the next generation of drug and diagnostic targets. The high-throughput methods of target selection are being integrated in both academic and industrial laboratories (see Table 2 for partnerships). Consider the example of SmithKline Beecham, which has chosen two small biotech companies, HySeq and Orchid, to provide robotic methods by the year 2001, or Dow Chemical (Midland, MI, USA) partnering with Caliper Technologies (Palo Alto, CA, USA) to produce custom chips. MIT/Whitehead Institute, an academic center that has led in the invention and application of high-throughput screening (HTS) technology, have just entered into a $40 million consortium with Bristol-Myers Squibb (New York, NY, USA), Affymetrix and Millennium (Cambridge, MA, USA) to accelerate functional genomic studies. Orchid and Caliper are the major proponents of 'flow chips'. These are laboratories on a chip designed to perform chemical and enzymatic reactions in microscale, for high throughput at low cost.

Millennium has launched a separate venture, Millennium Therapeutics, to which Eli Lilly (Indianapolis, IN, USA) has already allied itself in an effort to validate protein targets and develop small-molecule drugs.

Most provocative, diaDexus was recently created (August 1997) as a 50:50 joint venture by Incyte and SmithKline Beecham to apply their, until now, mutually exclusive human cDNA databases to DNA diagnostics. Since SmithKline Beecham benefits from its access to the HGS (Human Genome Sciences, Rockville, MD, USA) database, the new company should be in a prime position to bring us into the

age of personalized treatment and preventive medical care. The SmithKline Beecham contract with HGS required modification for the SmithKline Beecham contribution to diaDexus to proceed.

Genset signed a pact (in July 1997) whereby it will receive up to $42.5 million from Abbott Pharmaceuticals (Abbott Park, IL, USA) to correlate drug effectiveness with genotype for some of Abbott's products. Genaissance, a new biotech company launched in June, intends to work along similar lines to individualize treatment. Once they succeed, clinical trials will be streamlined, predefining the appropriate patient populations and subdividing a disease phenotype into its ontological components.

Roche acquired Corange (Boehringer-Mannheim) in May 1997 for $11 million to become one of the most powerful players in the DNA diagnostics business. They combine the PCR technology of the former with the reagent and diagnostic kit experience of the latter to make a formidable package.

TIGR, Craig Venter's genome research group, which left NIH to initiate expressed sequence tag sequencing on human cDNAs and found HGS, has again exercised its independence, ending the association with HGS in July 1997. TIGR has made significant progress in applying high-throughput selection methods to sequencing and is now analyzing microbial genomes.

Microbial and fungal genomes are targets for high-throughput sequencing by Genome Therapeutics Corporation (GTC) and Incyte. Both companies are entering into nonexclusive partnerships to share their microbial databases. GTC is sharing the data with Bayer Pharma and Schering-Plough (in the program PathoGenome™). Incyte's PathoSeq™ was recently released in updated version to subscribers, including Abbott, Eli Lilly, Roche and Zeneca (Alderley Edge, UK), and includes fungal sequences. HGS sells microbial sequence information to Pharmacia & Upjohn and Roche. Incyte recently announced the completion of 60% of the *Candida* genome. GTC started to release *Clostridium* sequence in September 1997.

Summary

While it is unlikely that drug development costs will decrease over the next decade, the above innovations should radically decrease development times, for two major reasons. Firstly, the HTS methods described will be combined with molecular modeling and combinatorial chemistry techniques to generate good candidates more rapidly following gene identification (which is already considerably more sensitive and efficient). Secondly, the preclinical and clinical parameters can be evaluated rapidly and quantitatively using the HTS and genetic predisposition prediction techniques. Screening not only selects the most relevant targets but it also indicates the critical parameters and their dynamic range for differential diagnosis; knowing what to measure (especially the approximate size of the variation), when to expect that variation and where it will occur in the evolution of disease and/or in the return to normal physiological development is critical.

The challenge to reduce spiraling costs and respond to the opportunities for new treatments generated by genomic discovery is appreciated by the biotechnology industry and academic scientists. Goals are clear-cut, and HTS solutions are coming online. Moreover, informatics is one of the most important players both in increasing sensitivity and in processing data. It is the key component in the formation of diaDexus, and is behind the Genset deal. It features heavily in almost all of the methods and collaborations discussed, and adds precision (e.g. BioTraces) and power to the technology (e.g. to Roche's PCR and Affymetrix GeneChips®). Just as the most productive biological scientists had to learn molecular biology in the 1970s and 1980s, they must learn computing skills today.

Most important, functional genomics is more than measurement and sequencing (accumulating data), it is knowing enough physiology and medicine to choose the best parameters and the right genes. HTS technology and sequence data are extending the biology that is the basis of physiology and clinical medicine.

REFERENCES

1 DiFiglia, M. *et al.* (1997) *Science* 277, 1990–1993
2 Zhang, P. *et al.* (1997) *Nature* 387, 151–158
3 Jacobson, D. (1997) *TBASE* danj@gdb.org
4 Gu, H. *et al.* (1994) *Science* 265, 103–106
5 Rivera, V.M. *et al.* (1996) *Nat. Med.* 2, 1028–1032
6 Goffeau, A. *et al.* (1996) *Science* 274, 546–568
7 Shoemaker, D.D. *et al.* (1996) *Nat. Genet.* 14, 450–456
8 Driever, W. (1996) *Development* 123, 165–178
9 Gayther, S.A. *et al.* (1995) *Nat. Genet.* 11, 428–433
10 Drews, J. and Ryser, S. (1997) *Drug Discovery Today* 2, 365–372
11 Chee, M. *et al.* (1996) *Science* 274, 610–614
12 Hacia, J.G. *et al.* (1996) *Nat. Genet.* 14, 441–447
13 Tobe, V.O. *et al.* (1996) *Nucleic Acids Res.* 24, 3728–3732
14 Wagner, R. *et al.* (1995) *Nucleic Acids Res.* 23, 3944–3948
15 Bellanné-Chantelot, C. *et al.* (1996) *Mutat. Res. (Genomics)* 6, 997–1003
16 Elgar, G. *et al.* (1996) *Trends Genet.* 12, 145–150
17 Capecchi, M.R. (1989) *Science* 244, 1288–1293
18 Smithies, O. (1994) *J. Biol. Chem.* 269, 27155–27159