Pergamon

European Journal of Cancer, Vol. 35, No. 14, pp. 1895–1904, 1999

Printed in Great Britain
0959-8049/99/$ - see front matter

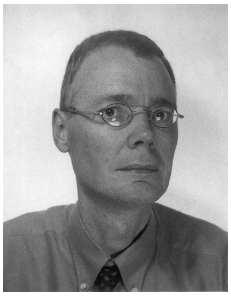
PII: S0959-8049(99)00253-1

European Journal of Cancer Millennium Review 2000

Molecular Pathology and Future Developments

J.J. Going [1] and B.A. Gusterson[2]

[1]Department of Pathology, University of Glasgow, Glasgow Royal Infirmary, Castle Street, Glasgow G4 OSF; and [2]Institute of Cancer Research, The Breakthrough Toby Robins Breast Cancer Research Centre, 237 Fulham Road, London SW3 6JB, U.K.



James Going trained in Pathology in Edinburgh and Glasgow, where he is Senior Lecturer in Pathology (Glasgow University) and Consultant Pathologist in North Glasgow University Hospitals. His research interests include origins of neoplasia in breast, oesophagus and stomach.

Professor Gusterson is Chairman of the Section of Cell Biology and Experimental Pathology at the Institute of Cancer Research in London. He is Founding Director of the Toby Robins Breast Cancer Research Centre and has an international reputation spanning from basic research on breast development and gene function to clinical trials.

There has already been a 'molecular' revolution in pathology. Demonstrating transcription of specific single genes or small gene sets and their protein products by *in situ* hybridisation and immunocytochemistry is routine in diagnostic and experimental pathology. A perhaps-greater revolution is imminent with the application of more recently established and emergent technologies in pathology. These include new approaches to polymerase chain reaction (PCR); simultaneous studies of multiple genes and their expression using oligonucleotide and cDNA arrays; serial analysis of gene expression (SAGE); expressed sequence tag (EST) sequencing, subtractive cloning and differential display; high-throughput sequencing; comparative genomic hybridisation, multiplex fluorescence *in situ* hybridisation (FISH) (spectral karyotyping); reverse chromosome painting; knockout and transgenic organisms;

Correspondence to J.J. Going, e-mail: going@udcf.gla.ac.uk
Received and accepted 30 Sep. 1999.

laser microdissection and micromachining; and new methods in bioinformatics, 'data mining' and data visualisation. Molecular methods will profoundly change diagnosis, prognosis and treatment targeting in oncology and elucidate fundamental mechanisms of neoplastic transformation. Individual susceptibility to specific diseases will become assessable and screening will be refined. The new molecular biology will be most fruitful in partnership with classical approaches to pathology: the expectation that molecular methods alone will answer all pathological questions is unrealistic. A further challenge for the biomedical community in the 'genome era' will be to ensure that the benefits of these sophisticated technologies are enjoyed globally.

Key words: molecular pathology, cDNA arrays, tissue arrays, serial analysis of gene expression (SAGE), multiplex FISH, comparative genomic hybridisation (CGH), real-time PCR, laser microdissection, proteomics, bioinformatics
Eur J Cancer, Vol. 35, No. 14, pp. 1895–1904, 1999

INTRODUCTION

'MOLECULAR PATHOLOGY' is not new. When Perutz and Kendrew began to elucidate the structure of haemoglobin 40 years ago they paved the way for the rich pathophysiology of the haemoglobinopathies. Histochemistry has demonstrated specific tissue enzymes for over 50 years and remains central to the diagnosis of muscle disorders. Labelled antibodies have revealed proteins (and therefore gene expression) in tissues since the work of Coons in the 1940s, and DNA has been specifically and quantitatively demonstrated by the classical technique of Feulgen and Rossenbeck for 75 years [1].

Monoclonal antibodies [2] are dependable reagents indispensable to all branches of modern diagnostic pathology. In haematopathology immunophenotype defines disease entities under the REAL classification, such as mantle cell lymphoma, which is characteristically CD5,20,43+, CD10,23−, and overexpresses cyclin D1 as a consequence of its typical t(11;14) translocation. Demonstrating light chain restriction by *in situ* hybridisation with probes for κ and λ is routine in the diagnosis and staging of plasma cell tumours, and PCR demonstrates clonal rearrangement of immunoglobulin and T-cell receptor genes in B cell and T cell tumours, respectively.

As only a very narrow definition of 'molecular pathology' would exclude these established activities, 'molecular pathology' has been practised for years. Nevertheless, there is feeling that we are on the threshold of a revolution in understanding disease, including the neoplastic disorders and the way they are diagnosed and treated, and perhaps prevented. Novel technologies, generating data on an unprecedented scale, will offer remarkable investigative opportunities; how these are used will depend on our ability to formulate and test biologically appropriate hypotheses and to handle torrents of data. A further challenge—perhaps the greatest—will be to ensure that benefits of the 'genome era' are not the privilege of wealthy nations.

This review seeks to identify molecular technologies recently established or emerging in the biological sciences, and to explore the possibilities they seem to offer for basic understanding of mechanisms and processes of neoplastic disease, in diagnostic practice, and in treatment targeting.

MUTATIONAL ANALYSIS

The potential to study gene expression by new technologies such as cDNA arrays and serial analysis of gene expression is complemented by analysis of DNA for mutations associated with tumours. This is not the place to review the vast literature on genetic changes associated with neoplastic transformation, but to explore some recent developments in demonstrating these events.

Classical cytogenetic analysis has shown many characteristic chromosomal alterations since the Philadelphia chromosome was defined 40 years ago, and improved methods ensure new discoveries.

Comparative genomic hybridisation (CGH)

In CGH, two DNA samples (e.g. from tumour and blood) are labelled with contrasting fluorescent molecules and simultaneously hybridised to normal metaphase chromosome spreads [3]. The quantitative ratio of the differently labelled sequences in the two samples determines the intensity of labelling by the contrasting fluorescence which is measured along each chromosome to assess relative DNA sequence copy number at low resolution but across the whole genome. This has revealed novel deletions, amplifications, and polysomies responsible for changes of gene dosage in tumours. Although more easily applied to unfixed tissues, it has been successfully used with fixed tissues [4] including *in situ* lesions [5]. Incorporation of CGH into routine diagnostic use has been predicted [6], but has not yet come about. Whole genome amplification has been used to apply CGH to single cells [7] and CGH of small samples would increase its diagnostic scope, but artifacts of whole-genome amplification are a concern. CGH 'on a chip' has been described [8].

Fluorescence in situ *hybridisation (FISH) and multiplex FISH*

Classical cytogenetic banding techniques have been supplemented by whole-chromosome and chromosome-arm painting [9] which identify specific chromosomes and chromosome arms in metaphase and interphase cells. Interphase cytogenetics has a role in diagnostic pathology. Abnormalities of 12p are frequent in human germ cell tumours, and were shown with 12p and 12q specific chromosomal paints [10], and other characteristic translocations can be demonstrated by interphase chromosome painting in fixed tissues.

Reverse chromosome painting [11] has been used to disentangle complex karyotypes in breast cancer cell lines which standard cytogenetics and chromosome painting did not manage [12], and in acute myeloid leukaemia [13]. In this approach, chromosome paints are prepared from the abnormal chromosomes and used to paint normal metaphase spreads, revealing the constituent elements of the abnormal chromosomes.

More recently, instrumental advances have enabled false-colour discrimination of all 24 human chromosomes by multiplex FISH (also called spectral karyotyping). This employs 24 chromosome 'paints' each labelled with a different combination of fluorochromes; with appropriate filters, charge coupled device (CCD) cameras and software, these can be uniquely discriminated [14, 15]. This approach has again permitted complex abnormal karyotypes to be resolved which would have been difficult or impossible with classical approaches (Figure 1). Well-established banding techniques (Giemsa, 4′,6-diamidino-2-phenylindole [DAPI]), site-specific FISH and CGH provide complementary information and may resolve ambiguities [16].

A consequence of the new cytogenetics may be that diagnostically and perhaps prognostically significant translocations may be documented for the commoner solid epithelial tumours, as well as for haematological and soft-tissue neoplasia where specific translocations are often highly characteristic. Carcinomas often have very complex karyotypes and it is possible that this complexity has obscured significant lesions, although it is also possible that the importance of specific translocations differs fundamentally between carcinomas and the leukaemias, lymphomas and sarcomas.

Mutation screening with arrays

As well as their applications to gene expression, cDNA and oligonucleotide arrays can probe DNA for mutations; this subject is discussed in connection with expression arrays, below.

GENE EXPRESSION—TRANSCRIPTION

Immunohistochemistry revolutionised diagnostic cellular pathology (histopathology and cytopathology) which were, up to that point, based largely on morphology, with few confirmatory tests available. Immunocytochemistry made a diagnosis such as 'anaplastic carcinoma' of the thyroid into a testable hypothesis: whilst neoplastic epithelial cells could express epithelial markers, if not thyroglobulin, lymphoid markers such as leucocyte common antigen ought to be negative. This led to increased recognition that many so-called 'anaplastic carcinomas' of the thyroid were in fact thyroid lymphomas. Lymphomas are now routinely characterised by their immunophenotype as well as their morphology. *In situ* hybridisation studies of gene expression are useful if abundant proteins may have been taken up by tumour cells from their surroundings: demonstrating albumin mRNA in hepatocellular carcinoma is a case in point. Figure 2 shows light chain restriction in a small focus of residual myeloma following chemotherapy, by *in situ* hybridisation using digoxigenin-labelled probes.

Modern diagnostic pathology of tumours relies heavily on gene expression studies to test diagnostic hypotheses. The ability to interrogate tumour cells for their expression of many hundreds or thousands of specific, ultimately well-characterised genes, encoding gene products increasingly of known function, will be a powerful tool for the tumour pathologist.

Both mRNA and proteins will be studied by a wide range of techniques. It is likely that interpretation of analyses will be dependent on the purity of samples. A tumour often contains diverse inflammatory, stromal, vascular and other cells as well as the tumour cells themselves, which may be greatly outnumbered. This important question is addressed below in the section on microdissection and laser machining of tissue sections.

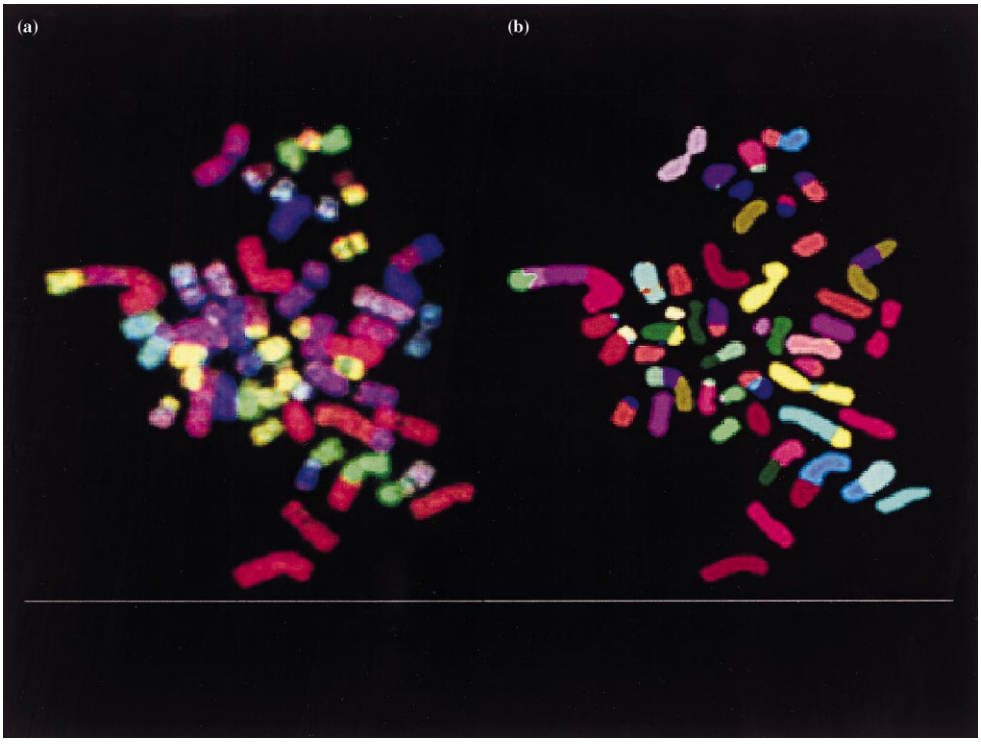


Figure 1. Spectral karyotyping of breast carcinoma cells. (a) is a metaphase painted with chromosomal paints specific for all 24 human chromosomes. (b) shows the same chromosomes with false colours indicating the chromosomal constituents of this complex abnormal karyotype with numerous translocations. (Courtesy of Paul Edwards and Joanne Davidson, Department of Pathology, University of Cambridge, U.K.).

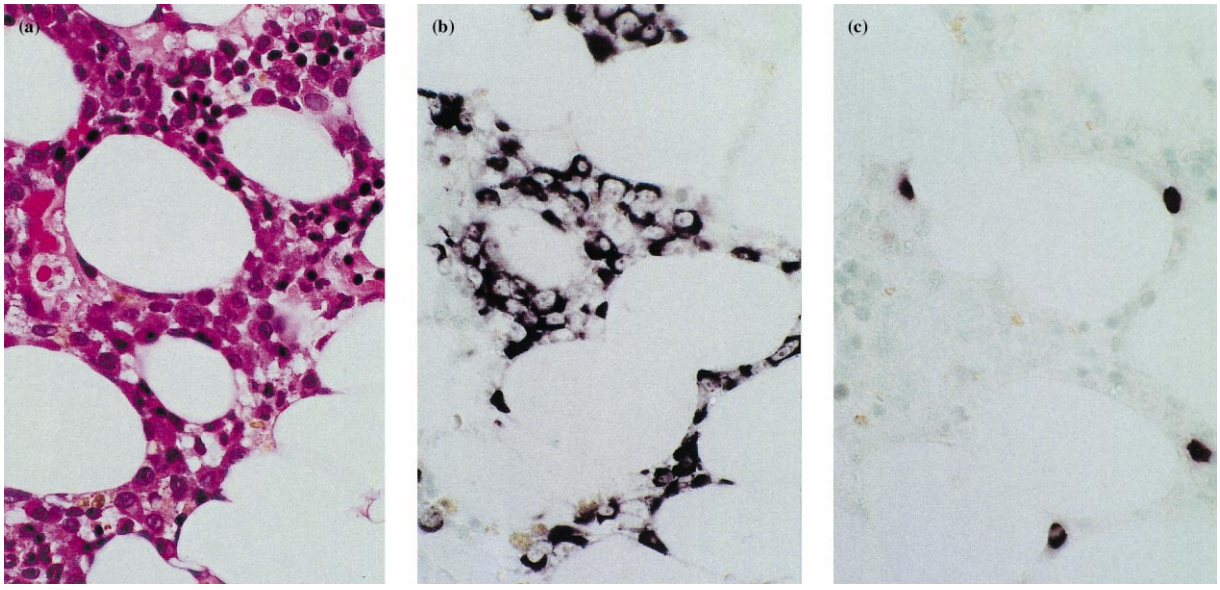


Figure 2. Minimal residual myeloma in bone marrow following chemotherapy. Haematoxylin and eosin histology (a) discloses a few atypical plasma cells. *In situ* hybridisation with probes specific for immunoglobulin light chains κ (b) and λ (c) shows that κ positive cells outnumber λ many-fold; this 'light chain restriction' indicates a neoplastic clone. (*In situ* hybridisation by Maura Farquharson, Department of Pathology, Glasgow Royal Infirmary, U.K.).

cDNA and oligonucleotide arrays

Array technology has been around for several years, but its full impact will make itself felt as commercial pressures improve methods and drive down prices [17]. Some arrays are referred to as 'chips' by analogy with silicon semiconductor chips. This may be an appropriate analogy; the price of computing power has halved every 18 months for nearly 20 years now; a similar phenomenon in the gene chip market would have equally fundamental effects.

In general, two approaches are taken to the manufacture of gene arrays. One is robotic 'spotting' of nucleic acids such as cDNA clones [18] on a substrate such as nylon membrane or glass [19]; the other is to synthesise oligonucleotide arrays *in situ* using complex masks and 'photolithographic' techniques to build the required oligonucleotides *in situ*, layer by layer [20]. These approaches each have their advantages and disadvantages. 'Spotting' requires that the necessary cDNA clones are available in the required degree of purity. Direct on-chip oligonucleotide synthesis generates shorter sequences for hybridisation reactions to occur. Molecular interactions in oligonucleotide microarrays are subject to intense scrutiny [21].

Array hybridisation can interrogate cells and tissues for gene expression, by making cDNA from mRNA and hybridising the labelled cDNA to the array. Figure 3 provides a typical working scheme for the application of this technology and Figure 4 demonstrates its application to the analysis of a human breast cancer. Competitive hybridisation is often performed, with cDNA prepared from two different samples, e.g. cells before and after hormonal stimulation or other treatment, and differentially labelled. Glass arrays are read using laser confocal microscopy; an advantage of arrays on glass is the dimensional stability of the substrate, which makes possible smaller features on the array, which in turn makes it feasible to create ever more dense arrays. Feature sizes of ~20μ achieved at present allow 300 000 different oligonucleotides to be synthesised in an array 12.8×12.8 mm [20] but there is no reason in principle why features of the order of 2μ should not be created, which would offer a 100-fold increase in the density of information on the array, and a corresponding reduction in the total amount of material required for hybridisation, which is at present a disadvantage of array technologies: they require rather large quantities of mRNA to start with. This is also a problem with other approaches such as serial analysis of gene expression (SAGE) (see below). Nevertheless, it can be anticipated that better technology will require less material, and that better methods of preparation will permit comprehensive gene-expression profiling of pure preparations of cells from complex tissues in which many different cell types interact via the short and intermediate-range processes by which tissue homeostasis is maintained.

Serial analysis of gene expression (SAGE)

Cloning and sequencing of 3′-directed cDNA libraries has been used for the quantitative study of gene expression. 3′ sequences are not necessarily translated into protein. 5′-directed cDNA libraries have been used to discover expressed genes with interesting structural motifs (expressed sequence tags: [22]) and many new genes have been characterised by this approach. Serial analysis of gene expression (SAGE) [23] is analogous to the first of these approaches and identifies transcripts from short sequence tags (9 or 10 base pairs) of known location in relation to the polyadenylate tail. Tags prepared from double-stranded cDNA are concatenated for cloning and sequencing. This ingenious but complex method has required relatively large amounts of RNA; efficiency improvements [24, 25] may permit SAGE of microdissected tissue. An advantage of SAGE is that it appears to yield good data about transcript abundance [26]; a consequent disadvantage is that much of the associated sequencing effort determines high-abundance transcripts, which may not be of the greatest biological relevance. SAGE has been successfully applied to characterising gene expression profiles in human cancers [27].

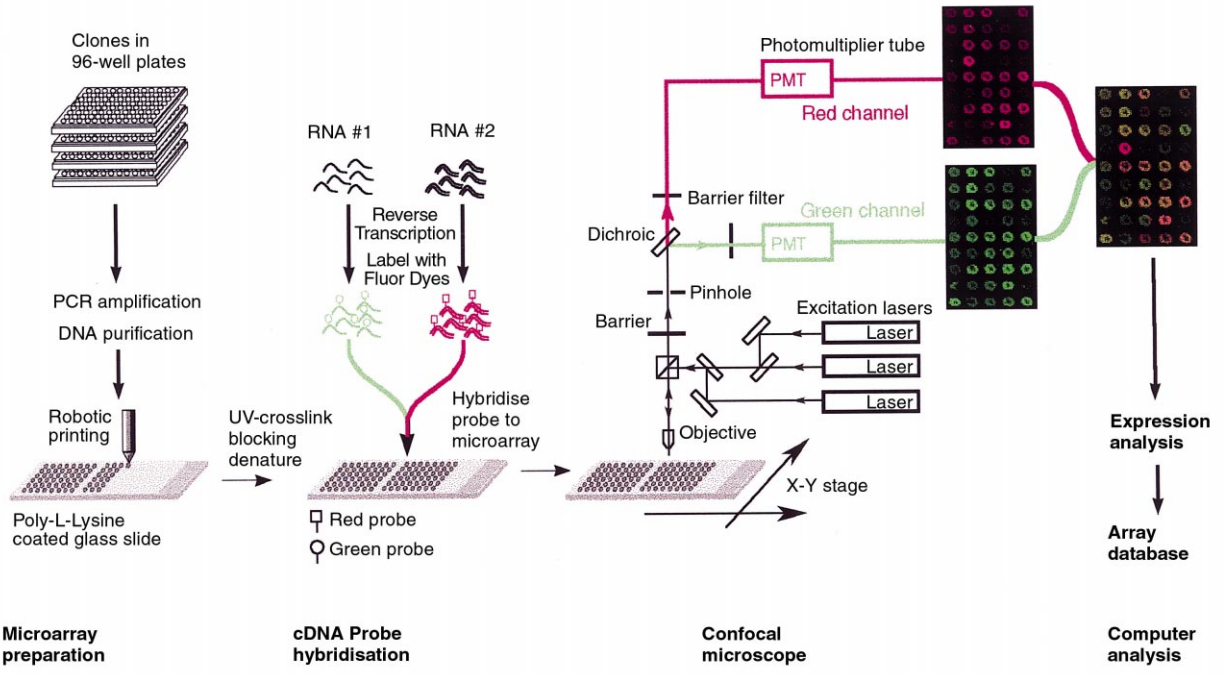


Figure 3. cDNA microarray schema. Templates for genes of interest are amplified by polymerase chain reaction (PCR). After purification and quality control, aliquots are printed on coated glass microscope slides by a computer-controlled, high-speed robot. Total RNA from samples 1 and 2 are fluorescently labelled with either Cye3- or Cye5-DUTP by a single round of reverse transcription. The fluorescent targets are pooled and hybridised under stringent conditions to the probes on the microarray. Laser excitation of the fluorophores incorporated in the bound targets yields emissions with characteristic spectra, which are measured using a scanning confocal laser microscope. Monochrome images from the scanner are imported into software, pseudo-coloured and merged. Data about the genes, including gene name, clone identifier, intensity values, intensity ratios, normalisation constant and confidence intervals are attached to each probe. (Figure adapted from [18]).

Detecting mutations and polymorphisms with arrays

As well as their applications to studies of gene expression, arrays are also able to detect mutations and polymorphisms [28]. Approaches include gain-of-signal analysis with allele-specific oligonucleotide probes, loss-of-signal analysis in competitive hybridisation with differentially labelled test and reference probes, and 'minisequencing' in which fluorescent dideoxynucleotide triphosphates (ddNTPs) extend the array probe sequence to identify the adjacent nucleotide in bound DNA. Combined loss- and gain-of-signal approaches have suggested that a large percentage of *BRCA1* mutations can be detected in high-density oligonucleotide arrays [28]. This approach will also lend itself to detecting single nucleotide polymorphisms.

Differential display and representational difference analysis

Arbitrarily primed polymerase chain reaction (PCR) mRNA differential display [29, 30] is a relatively robust and simple technique for examining differences in expression between samples, but is sensitive to the degradation of mRNA, which is often a problem with human pathological material. Its strengths and weaknesses have usefully been reviewed [31]. An advantage is that it may be applied to minimal quantities of starting material. Alternatives to differential display include representational difference analysis, in which subtractive hybridisation and PCR are combined to amplify differences in expression between two DNA [32] or mRNA samples [33] by factors up to 10^6. This is a relatively complex technique.

GENE EXPRESSION: TRANSLATION—PROTEOMICS AND MASS SPECTROMETRY

Gene expression does not end with transcription. Levels of message do not necessarily reflect protein concentrations, and protein functions are changed by glycosylation, phosphorylation and other modifications. Proteomics concerns itself with the totality of protein expression and modification in cells, and it too commands powerful new tools, including two-dimensional (2-D) gel electrophoresis, high performance liquid chromatography (HPLC) and mass spectrometry. The combination of mass spectrometric (MS) techniques, which include electrospray ionisation (ESI) and matrix-assisted laser desorption ionisation time-of-flight (MALDI-TOF) MS, and 2-D electrophoresis is a powerful tool for identifying proteins and their post-translational modifications such as glycosylation [34] and phosphorylation. ESI has been applied to analysis at femtomolar/attomolar levels [35]. Mass spectrometric sequencing of proteins, peptides and nucleic acids is well established and has, for instance, revealed new products of the pro-opiomelanocortin gene [36]. Proteomics has been applied in a study of keratin expression in metaplastic bladder urothelium [37] and protein profiling of normal tissues and tumours has begun. The main limitation is still sensitivity but high-resolution mapping is possible with 2000 analysable proteins per gel (Figure 5). Cell fractionation into cytoplasmic, nuclear and membrane-bound proteins improves sensitivity. Many abnormalities in cancer are associated with changes to signalling pathways and mass spectrometry will have a major part in the investigation of tumours and mechanisms of drug action.

Mass spectrometry is also a powerful sequencing tool for nucleic acids and has been used to derive reliable sequencing data for *TP53* exons 5–8 from femtomolar samples by MALDI-TOF MS [38].

TISSUE ARRAYS

Tissue microarray technology was designed as a high throughput method for analysing changes involved in the development and progression of cancers at both the nucleic acid and the protein level. The uniqueness of this technique is the ability to simultaneously analyse large numbers of human tumours with minimal consumption of rare archival specimens.

Tissue microarrays are made by retrieving 0.6 mm cores of tumour biopsies from different tumour blocks and precisely arraying them in a single paraffin block [39]. Negligible damage is caused to the donor blocks. Since the cores are only 0.6 mm in diameter [40], approximately 600 different tumours can be united in a single block. This capacity would allow simultaneous screening of many biopsies [41] in clinical trials for testing and targeting new diagnostic or prognostic markers and new therapeutic drugs. Another advantage of increased capacity is reduced screening time and use of reagents. Heterogeneity of most tumours makes 100% concordance between results from cores and whole sections or biochemical assays impossible, but 84% concordance was reported between immunohistochemistry (IHC) of oestrogen receptor (ER) of cores and of a biochemical homogenate, and similar results have been shown with ERBB2 data from tumour arrays compared with disaggregated nuclei from whole mount sections. Tissue microarrays are suitable for use in IHC, FISH, PCR and reverse transcriptase (RT)-PCR, and cores can be used for DNA and RNA extraction. It is estimated that DNA from a 0.6 mm core is sufficient for 50 PCR runs.

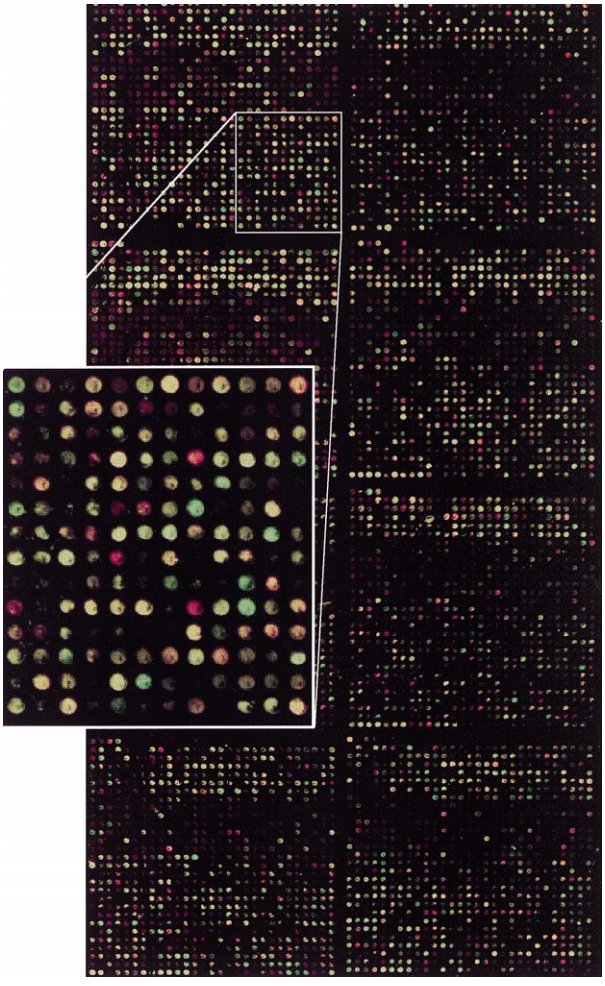


Figure 4. cDNA microarray hybridisation. Data from a single hybridisation experiment viewed as a normalised ratio. Significantly increased (>1; red) or decreased (<1; green) ratios indicate gene expression relative to the reference sample (no change: yellow). For this particular image, an array of 6500 cDNA probes was hybridised with RNA from a sporadic breast cancer reverse transcribed in the presence of Cye3-dUTP (red channel) and a reference sample reverse transcribed with Cye5-dUTP (green channel). (Courtesy of Ingrid Hedenfalk and David Duggan, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, U.S.A.).

MICRODISSECTION AND LASER MICROSCOPY

Microdissection of histological sections and 'cell picking' have a long history and many different techniques. Sensitive analyses including PCR (capable of analysing single molecules in genotyping individual haploid cells such as spermatozoa) have motivated the development of new techniques. These broadly divide into mechanical methods using micromanipulated microtools, which have been applied down to the scale of dissecting individual chromosomal bands, and a variety of laser-assisted methods. Lasers are used in two main, contrasting ways. In so-called laser capture microdissection [42], cells or a tissue section on a glass slide are overlain by a thermoplastic film. Low-wattage, long-wavelength laser light is passed through the section in areas of tissue to be recovered, causing brief heating with softening or melting of the thermoplastic film, which adheres to the sections in these areas. Removal of the thermoplastic film strips away areas for analysis but leaves behind those not wanted. It has been shown that proteins from tissue so retrieved can be analysed by proteomic methods [43]. Success is influenced by physical properties of the tissue and as the wavelength of

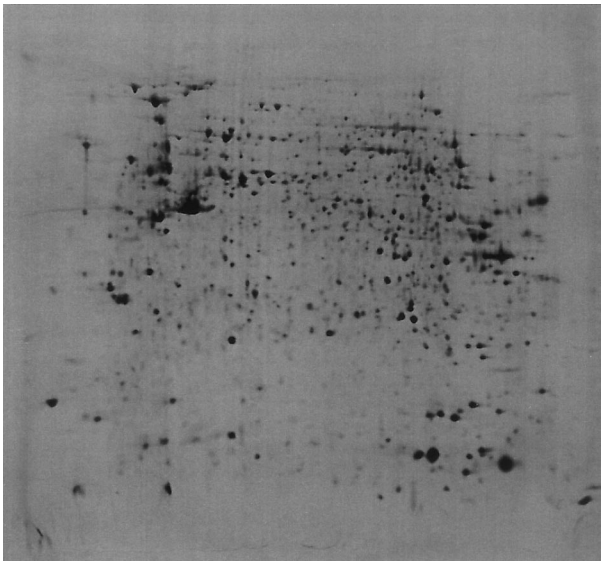


Figure 5. Two-dimensional electrophoresis of protein from a human breast carcinoma. Each spot represents a different protein or a protein modified by a process such as phosphorylation. Separation is by size vertically and by charge horizontally. Sequence of individual proteins can be established by mass spectrometry. Courtesy of Martin Page (Oxford Glycosciences, Oxfordshire, U.K.) and Mike O'Hare (Ludwig Institute, London, U.K.).

the laser light is relatively long (typically ≈7μ), precision may not be great. In contrast to this approach, laser photons may be used to destroy unwanted cells and to cut around areas of interest in sections [44]. These may be recovered mechanically or by other means. The effects of laser light on tissue sections depend on wavelength and power density.

Visible and infra-red lasers usually cause burning. An entirely different process begins to occur when short (ultraviolet) wavelengths are applied at high flux density (1–5 J/cm^2). These energetic photons cause direct photolysis of chemical bonds without heating [45], and the irradiated tissue is instantaneously converted to an atomic plasma which is ejected from the section. Such wavelengths are associated with excimer lasers and other types. Figure 6 shows photoablation of a 4μ tissue section by a demagnified rectangular mask and a KrF excimer laser ($\lambda = 248$ nm). This resolution makes it possible to envisage that a single cell type could be preserved whilst all others were ablated from a section, even cells dispersed singly in complex tissues. This might be achieved manually on a small scale, but larger throughput would require automation. Targeting cells by immunophenotype suggests itself.

KNOCKOUT AND TRANSGENIC ORGANISMS

RNA inhibition knockouts in Caenorhabditis elegans

Interference with gene function by injected RNA has usually been attributed to antisense inhibition of mRNA. It was surprising therefore that double-stranded (ds) RNA caused more potent specific interference with gene function in *C. elegans* than single-stranded RNA [46], at low concentrations of dsRNA precluding stoichiometric inhibition. It appears that dsRNA causes accelerated degradation of specific mRNA prior to translation [47]. This elegant and flexible tool for abrogation of gene function is being exploited in developmental studies in *C. elegans* and has been demonstrated in other organisms [48]. Functional insights may emerge for many of these genes from *C. elegans*, 'RNA interference' [49] and knockouts [50] and arrayed cDNAs from *C. elegans* will reveal homologies in mouse and human tissue RNA profiles.

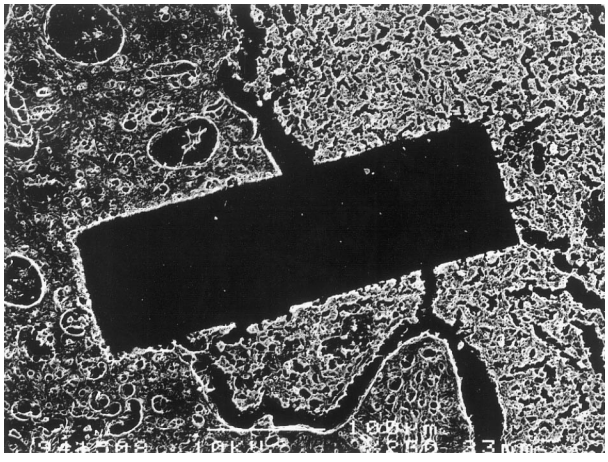


Figure 6. Scanning electron micrograph of a rectangular area ablated by KrF excimer laser from a 4μ paraffin section of metastatic adenocarcinoma (left) in human liver (right). Note the sharp demarcation between unaltered tissue and complete ablation. (Laser ablation by Dafydd Thomas, Exitech Ltd, Oxford, U.K.).

Transgenes and knockouts in higher animals

The construction of organisms with functional genes they would not normally possess (transgenes) or in which normal gene function has been deleted by gene targeting (knockouts) are firmly established techniques for exploring gene function *in vivo*. These have been extensively applied in developmental biology to demonstrate the cardinal roles of many different genes in normal embryogenesis and organogenesis, e.g. 11 genes essential for normal development of the metanephric kidney [51] and in the investigation of multigenic disorders such as hypertension by 'gene titration' approaches simulating natural variation [52]. Gene targeting by homologous recombination introduces alterations from point mutations to translocations, some with tissue-specific expression, and diverse application [53]. Defining phenotypes associated with these genetic alterations is important, and these effects may be subtle. Classical histological techniques are important in this area.

Transgenes and knockouts have been applied intensively in cancer research, e.g. in creating mouse models of syndromes of defective nucleotide excision repair like xeroderma pigmentosum and Cockayne syndrome [54], in exploring the functions of tumour suppressor genes [55] and in studying cancers at specific sites [56]. Recent advances in tissue-specific manipulation of targeted genes in mouse models have enabled molecular dissection of many diseases including cancer. For example, entirely new genetically engineered mouse models are defining the transforming potential of genes implicated in human breast cancer [57]. The demonstration that a *c-myc* transgene expressed in the mammary gland under mouse mammary tumour virus (MMTV) long terminal repeat (LTR) control induced mammary tumours [58] was a seminal event and targeted mutagenesis has been used elegantly to dissect signalling pathways [59].

Although mice with homozygous null mutations of *BRCA1* die before embryonic day 9, and heterozygotes for *BRCA1* deletion do not develop mammary tumours, time and tissue-specific CRE-LOXP mutation of the intact *BRCA1* allele in the mammary glands of mice heterozyous for *BRCA1* deletion appears to mimic the situation in human carriers of *BRCA1* mutations, by disabling the intact copy of the gene later in the life of the mouse [60]. Mammary gland-specific deletion of *BRCA1* exon 11 (encoding the Rad51 binding domain) is achieved by an MMTV-CRE or a whey acidic protein (WAP)-CRE transgene. The MMTV and WAP promoters are maximally activated during pregnancy and lactation and the CRE (or CRE-LOXP) system excises specific DNA sequences under the control of these tissue-specific promoters. The model mice thus lose *BRCA1* repair function on lactation. The resultant *BRCA1* conditional knockout appears to model the molecular mechanism of *BRCA1* involvement in human breast cancer, with the development of mammary tumours with similarities to those reported in human cancers arising in a mutant *BRCA1* background.

QUANTITATIVE AND REAL-TIME PCR

The impact of PCR needs no description, but some recent developments increase its scope. These include quantitative and real-time PCR. Many different approaches to quantitative PCR have been described including competitive PCR and real-time PCR assays which incorporate fluorogenic probes [61]. In a free probe molecule the fluorescence of a reporter molecule is quenched but as the specific PCR

product accumulates, the 5′-3′ exonuclease activity of Taq polymerase releases a fluorogenic fragment, or probe binding itself [62] removes the quenching effect and a fluorescent signal is generated proportional to specific product accumulation. This can be monitored in real time as the PCR reaction proceeds. The real-time 5′-exonuclease assay was used [63] to detect the t(14;22)(q31;q21) of follicular lymphoma with high specificity (with respect to conventional PCR) and sensitivity (picogram quantities of template), making an ideal assay for minimal residual disease and one in which the absence of post-PCR analytical stages is attractive for routine diagnostic use.

INFORMATICS AND DATA MINING

The new era of genomic studies will require new approaches to deal with data sets unprecedented in scale in the biological sciences: astronomical in the rather literal sense for astronomy is one of the sciences which has already faced the problem of dealing with terabyte datasets, and it is likely that data visualisation and 'data mining' techniques developed in the physical sciences will find application in this area, along with new statistical techniques.

These huge datasets seem daunting, and the difficulties are real, but classical diagnostic histopathology and cytopathology already require the interpretation of massive visual datasets: a 35 mm transparency scanned at the highest resolution in Kodak Photo-CD format is approximately 20 MB of data, and a microscope field is a higher resolution image. Depending on the exact task, a pathologist can view new fields every second or two, and may view several hundred fields before coming to a firm conclusion about a case; this represents gigabytes of data. A large volume of the human brain is devoted to visual data processing. Data visualisation methods may therefore allow characteristic patterns to be appreciated and are likely to be important.

Graphical display of complex genetic data employing mathematical techniques of cluster analysis have been shown to group *Saccharomyces cerevisiae* genes of similar known function, offering possible pointers to functions of new genes [64]. Another approach has been to apply neural networks in so-called 'self-organising maps' and this too has been shown to group data by expression profiles and to associate genes of similar function in *S. cerevisiae* [65]. The importance of robust statistical approaches to data evaluation has recently been emphasised [66]. The issue of 'data trawling' or 'data dredging' generates debate [67] and although the scientific purist may demand the testing only of hypotheses specified a priori, in reality data are going to be looked at repeatedly, especially if available in public-access databases, and it will be incumbent upon investigators to use statistically respectable methodologies. The collaboration of experts in diverse fields will be required [68] and attempts to integrate multiple resources are being made [69].

DIAGNOSTICS, BIOMARKERS AND DRUG DEVELOPMENT

Many of the technologies discussed above are being applied in expectation of better diagnosis, prediction of treatment response or drug resistance, and to identify biomarkers and new therapeutic targets. Large-scale comparison of RNA from tumours and normal tissues is likely to identify hundreds or thousands of potential targets, but for many the full gene sequence will be unknown and function predicted rather than established. Biologically relevant screening of putative targets will be essential, and for cancers, clinical trials are an appropriate setting: a protein is not likely to be a therapeutic target if it is not prognostic or predictive in a trial. There is a gap to bridge between a fragment of RNA sequence and clinical relevance. *In situ* hybridisation to assess expression of target mRNA in trial cases may be useful, but protein expression is well-tested and robust. Rapid production of reliable antibodies to the implied products of coding sequences will be required, but conventional production of polyclonal and monoclonal antibodies may take weeks or months. It is likely that a combination of tissue arrays from clinical trial material, and antibody production by phage display [70], will allow adequately comprehensive studies. Phage display selects antibodies against most targets within 2 weeks, and arrayed tissues from thousands of cases in a large clinical trial could be screened in an afternoon with automated reading of results directly linked to a clinical database of outcome. In theory, it could be possible to go from expression RNA array to target validation in 4 weeks, without knowing anything about the function of the molecule.

ETHICS

Powerful new technologies, by enabling previously impossible tasks, create ethical conflicts [71]. Individuals may resist screening for mutations or polymorphisms modifying their risk of common disorders such as cancer, hypertension, myocardial or brain infarction, diabetes, schizophrenia, alcoholism or drug abuse, but employers, law enforcement agencies, insurance and pension providers, actual or prospective spouses and parents may not. The prospect of multi-billion dollar sales will drive sophisticated marketing by providers using the effective tools of the global advertising industry. Business is propelled by laws as impersonal as those of thermodynamics so to avoid uses of these technologies we may find unacceptable a legal and ethical debate must ensue. Military use of technologies of the 'genome era' must also be anticipated.

PRACTICALITIES

Molecular studies need to be integrated with the effective classical methods of surgical pathology and cytopathology. Medical pathologists, educated in a tradition of primary responsibility to their patients, and as members of multidisciplinary teams responsible for patient care, are properly reluctant to modify effective working procedures. This creates difficulties of access to tissue for molecular analysis. Removal of tissue from a resection by a person not appropriately trained can make interpretation difficult or impossible; a common problem being with assessment of resection margins. Ideally, therefore, the pathologist samples tissue personally or supervises sampling. As almost every cancer resection yields more than enough tissue for diagnosis and ancillary studies, there should, on the face of it, be no problem, but routine sampling and storage of unfixed tissue is a significant commitment which needs to be undertaken critically and conscientiously, as poorly documented or stored tissue samples are of limited use.

Fixatives which cause less damage to nucleic acids and proteins than formaldehyde as a fixative in surgical pathology are sought, with the goal of making routine archival tissue blocks more suitable for molecular studies. 70% alcohol has been employed [72]. However, the advantages of optimal

formaldehyde fixation are great, and there is likely to be reluctance to abandon its use at all generally, and the limitations of formaldehyde-fixed paraffin-embedded tissue may have to be accepted in some circumstances. The Food and Drug Administration of the U.S.A. has approved an immunocytochemistry-based test of HER-2/neu (erb-B2) protein expression ('HercepTest') on fixed tissue as a basis for selecting patients more likely to benefit from specific therapy targeting this protein [73], despite uncontrolled fixation-induced degradation of HER-2/neu protein. This test has been criticised [74]. Fixation of surgical specimens may occur over 12–72 h, at temperatures between 10 and 30°C, and is influenced by quantity of fixative, mass of the specimen and prosection. Standardised antigen retrieval is not likely to compensate for these differences, but the advantages of HER-2/neu assay on unfixed tissue may be irrelevant if unavailable.

CONCLUSIONS

Appraisals of the impact of the 'new biology' on pathology have varied from the largely positive [75] to more sceptical positions [76], but pathologists need to seize the opportunities of the molecular revolution, and basic scientists who wish to address the complexities of real disease, will benefit from research partnerships with pathologists. From such partnerships will emerge real clinical progress.

1. Feulgen R, Rossenbeck H. Mikroskopic-chemischer Nachweis einen Nucleinsaure von Typus der Thymonucleinsaure und die darauf berhende elektive Färbung von Zellkernen in mikroskopischen Preparäten. *Zeitschrift Physiol Chem* 1924, **135**, 203–248.
2. Kohler G, Milstein C. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* 1975, **296**, 495–497.
3. Kallioniemi OP, Kallioniemi A, Sudar D, *et al.* Comparative genomic hybridisation—a rapid new method for detecting and mapping DNA amplifications in tumours. *Semin Cancer Biol* 1993, **4**, 41–46.
4. Reid T, Just KE, Holtgrevegrez H, *et al*. Comparative genomic hybridisation of formalin-fixed, paraffin-embedded breast tumours reveals different patterns of chromosomal gains and losses in fibroadenomas and diploid and aneuploid carcinomas. *Cancer Res* 1995, **55**, 5415–5423.
5. James LA, Mitchell ELD, Menasce L, Varley JM. Comparative genomic hybridisation of ductal carcinoma *in situ* of the breast: identification of regions of DNA amplification and deletion in common with invasive breast carcinoma. *Oncogene* 1997, **14**, 1059–1065.
6. Zitzelsberger H, Lehmann L, Werner M, Bauchinger M. Comparative genomic hybridisation for the analysis of chromosomal imbalances in solid tumours and haematological malignancies. *Histochem Cell Biol* 1997, **108**, 403–417.
7. Wells D, Sherlock JK, Handyside AH, Delhanty JDA. Detailed chromosomal and molecular genetic analysis of single cells by whole genome amplification and comparative genomic hybridisation. *Nucl Acids Res* 1999, **27**, 1214–1218.
8. Bao YP, Muller U, Che D, *et al.* A novel microchip array for detecting gene amplification in cancer cells. *Cytogenet Cell Genet* 1998, **82**, 140.
9. Ried T, Schrock E, Ning Y, Wienberg J. Chromosome painting: a useful art. *Hum Mol Genet* 1998, **7**, 1619–1626.
10. Blough RI, Heerema NA, Ulbricht TM, Smolarek TA, Roth LM, Einhorn LH. Interphase chromosome painting of paraffin-embedded tissue in the differential diagnosis of possible germ-cell tumours. *Modern Pathol* 1998, **11**, 634–641.
11. Carter NP, Ferguson-Smith MA, Perryman MT, *et al.* Reverse chromosome painting: a method for the rapid analysis of aberrant chromosomes in clinical cytogenetics. *J Med Genet* 1992, **29**, 299–307.
12. Morris JS, Carter NP, Ferguson-Smith MA, Edwards PAW. Cytogenetic analysis of three breast carcinoma cell lines using reverse chromosome painting. *Genes Chromosomes Cancer* 1997, **20**, 120–139.
13. Arkesteijn G, Jumelet E, Hagenbeek A, Smit E, Slater R, Martens A. Reverse chromosome painting for the identification of marker chromosomes and complex translocations in leukaemia. *Cytometry* 1999, **35**, 117–124.
14. Speicher MR, Ballard SG, Ward DC. Karyotyping human chromosomes by combinatorial multi-fluor FISH. *Nature Genet* 1996, **12**, 368–375.
15. Schrock E, Veldman T, Padilla-Nash H, *et al.* Spectral karyotyping refines cytogenetic diagnostics of constitutional chromosomal abnormalities. *Hum Genet* 1997, **101**, 255–262.
16. Macville M, Schrock E, Padilla-Nash H, Keck C, Ghadimi BM, Zimonjic D, Popescu N, Ried T. Comprehensive and definitive molecular cytogenetic characterisation of HeLa cells by spectral karyotyping. *Cancer Res* 1999, **59**, 141–150.
17. Fox JL. Complaints raised over restricted microarray access. *Nat Biotechnol* 1999, **17**, 325–326.
18. Duggan DJ, Bittner M, Chen Y, Meltzer P, Trent JM. Expression profiling using cDNA microarrays. *Nature Genet* 1999, **21**, S10–S14.
19. Cheung VG, Morley M, Aguilar F, Massimi A, Kucherlapati R, Childs G. Making and reading microarrays. *Nature Genet* 1999, **21**, S15–S19.
20. Lipshutz RJ, Fodor SPA, Gingeras TR, Lockhart DJ. High density synthetic oligonucleotide arrays. *Nature Genet* 1999, **21**, S20–S24.
21. Southern E, Mir K, Shchepinov M. Molecular interactions on microarrays. *Nature Genet* 1999, **21**, S5–S9.
22. Carulli JP, Artinger M, Swain PM, *et al.* High throughput analysis of differential gene expression. *J Cell Biochem Suppl* 1998, **31**, 286–296.
23. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. *Science* 1995, **270**, 484–487.
24. Datson NA, van der Perk-de Jong J, van den Berg MP, de Kloet ER, Vreugdenhil E. MicroSAGE: a modified procedure for serial analysis of gene expression in limited amounts of tissue. *Nucleic Acids Res* 1999, **27**, 1300–1307.
25. Powell J. Enhanced concatemer cloning—a modification to the SAGE (Serial Analysis of Gene Expression) technique. *Nucl Acids Res* 1998, **26**, 3445–3446.
26. Bertelsen AH, Velculescu VE. High-throughput gene expression analysis using SAGE. *Drug Discovery Today* 1998, **3**, 152–159.
27. Hibi K, Liu Q, Beaudry GA, *et al.* Serial analysis of gene expression in non-small cell lung cancer. *Cancer Res* 1998, **58**, 5690–5694.
28. Hacia JG. Resequencing and mutational analysis using oligonucleotide microarrays. *Nat Genet* 1999, **21**, 42–47.
29. Liang P, Pardee AB. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* 1992, **257**, 967–971.
30. Welsh J, Chada K, Dalal D, Cheng R, Ralph D, McLelland M. Arbitrarily primed PCR fingerprinting of RNA. *Nucl Acids Res* 1992, **20**, 4965–4970.
31. Tortola S, Capella G, Marcuello E, *et al.* Analysis of differential gene expression in human colorectal tumor tissues by RNA arbitrarily primed-PCR: a technical assessment. *Lab Invest* 1998, **78**, 309–317.
32. Lisitsyn N, Lisitsyna N, Wigler M. Cloning the differences between two complex genomes. *Science* 1993, **259**, 946–951.
33. Hubank M, Schatz DG. Identifying differences in mRNA expression by representational difference analysis of cDNA. *Nucl Acids Res* 1994, **22**, 5640–5648.
34. Packer NH, Harrison MJ. Glycobiology and proteomics: is mass spectroscopy the holy grail? *Electrophoresis* 1998, **19**, 1872–1882.
35. Smoluch MT, Sallberg M, Silberring J. Nanospray mass spectrometry for identification of peptides. *Acta Biochim Pol* 1999, **46**, 475–481.
36. Jespersen S, Chaurand P, Strien FJCv, Spengler B, Greef Jvd. Direct sequencing of neuropeptides in biological tissues by MALDI-PSD mass spectrometry. *Anal Chem* 1999, **71**, 660–666.
37. Celis JE, Celis P, Ostergaard M, *et al.* Proteomics and immunohistochemistry define some of the steps involved in the squamous differentiation of the bladder transitional epithelium: a novel strategy for identifying metaplastic lesions. *Cancer Res* 1999, **59**, 3003–3009.

38. Fu DJ, Tang K, Braun A, *et al.* Sequencing exons 5 to 8 of the p53 gene by MALDI-TOF mass spectrometry. *Nature Biotechnol* 1998, **16**, 381–384.
39. Kononen J, Bubendorf L, Kallioniemi A, *et al.* Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nature Med* 1998, **4**, 844–847.
40. Bubendorf L, Kononen J, Koivisto P, *et al.* Survey of gene amplifications during prostate cancer progression by high-throughout fluorescence *in situ* hybridization on tissue microarrays. *Cancer Res* 1999, **59**, 803–806. (Published erratum appears in *Cancer Res* 1999, **59**(6), 1388).
41. Moch H, Schraml P, Bubendorf L, *et al.* High-throughput tissue microarray analysis to evaluate genes uncovered by cDNA microarray screening in renal cell carcinoma. *Am J Pathol* 1999, **154**, 981–986.
42. Emmert-Buck MR, Bonner RF, Smith PD, *et al.* Laser capture microdissection. *Science* 1996, **274**, 998–1001.
43. Banks RE, Dunn MJ, Forbes MA, *et al.* The potential use of laser capture microdissection to selectively obtain distinct populations of cells for proteomic analysis—preliminary findings. *Electrophoresis* 1999, **20**, 689–700.
44. Bohm M, Wieland I, Schutze K, Rubben H, Microbeam MO. MeNT: non-contact laser microdissection of membrane-mounted native tissue. *Am J Pathol* 1997, **151**, 63–67.
45. Srinivasan R. Ablation of polymers and biological tissue by ultraviolet light. *Science* 1986, **234**, 559–565.
46. Fire A, Xu SQ, Montgomery MK, Kostas SA, Driver SE, Mello CC. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* 1998, **391**, 806–811.
47. Montgomery MK, Xu SQ, Fire A. RNA as a target of double-stranded RNA mediated genetic interference in Caenorhabditis elegans. *Proc Natl Acad Sci USA* 1998, **95**, 15502–15507.
48. Ngo H, Tschudi C, Gull K, Ullu E. Double-stranded RNA induces mRNA degradation in Trypanosoma brucei. *Proc Natl Acad Sci USA* 1998, **95**, 14687–14692.
49. Tabara H, Grishok A, Mello CC. RNA in C. elegans: soaking in the genome sequence. *Science* 1998, **282**, 430–431.
50. Hodgkin J, Herman RK. Changing styles in C. elegans genetics. *Trends Genet* 1998, **14**, 352–357.
51. Lipschutz JH. Molecular development of the kidney: a review of the results of gene disruption studies. *Am J Kidney Dis* 1998, **31**, 383–397.
52. Takahashi N, Smithies O. Gene targeting approaches to analysing hypertension. *J Am Soc Nephrol* 1999, **10**, 1598–1605.
53. Muller U. Ten years of gene targeting: targeted mouse mutants, from vector design to phenotype analysis. *Mechanisms Dev* 1999, **82**, 3–21.
54. Boer Jd, Hoeijmakers J. Cancer from the outside, aging from the inside: mouse models to study the consequences of defective nucleotide excision repair. *Biochimie* 1999, **81**, 127–137.
55. Ghebranious N, Donehower LA. Mouse models in tumour suppression. *Oncogene* 1998, **17**, 3385–3400.
56. Banerjee A, Quirke P. Experimental models of colorectal cancer. *Dis Colon Rectum* 1998, **41**, 490–505.
57. Cardiff RD, Munn RJ. The histopathology of transgenes and knockouts in the mammary gland. *Adv Oncobiol* 1998, **2**, 177–202.
58. Stewart TA, Pattengale PK, Leder P. Spontaneous mammary adenocarcinomas in transgenic mice that carry and express MTV/myc fusion genes. *Cell* 1984, **38**, 627–637.
59. Webster MA, Hutchinson JN, Rauh MJ, *et al.* Requirement for both Shc and phosphatidylinositol 3′ kinase signaling pathways in polyomavirus middle T-mediated mammary tumorigenesis. *Mol Cell Biol* 1998, **18**, 2344–2359.
60. Kuhn R, Schwenk F, Aguet M, Rajewsky K. Inducible gene targeting in mice. *Science* 1995, **269**, 1427–1429.
61. Orlando C, Pinzani P, Pazzagli M. Developments in quantitative PCR. *Clin Chem and Lab Med* 1998, **36**, 255–269.
62. Lie YS, Petropoulos CJ. Advances in quantitative PCR technology: 5′ nuclease assays. *Curr Opinion Biotechnol* 1998, **9**, 43–48.
63. Luthra R, McBride JA, Cabanillas F, Sarris A. Novel 5′ exonuclease-based real-time PCR assay for the detection of t(14;18)(q32;q21) in patients with follicular lymphoma. *Am J Pathol* 1998, **153**, 63–68.
64. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 1998, **95**, 14863–14868.
65. Törönen P, Kolehmainen M, Wong G, Castren E. Analysis of gene expression data using self-organizing maps. *FEBS Lett* 1999, **451**, 142–146.
66. Wittes J, Friedman HP. Searching for evidence of altered gene expression: a comment on statistical analysis of microarray data. *J Natl Cancer Inst* 1999, **91**, 400–401.
67. Michels KB, Rosner BA. Data trawling: to fish or not to fish. *Lancet* 1996, **348**, 1152–1153.
68. Bassett DE, Eisen MB, Boguski MS. Gene expression informatics—it's all in your mine. *Nature Genetics* 1999, **21s**, 51–55.
69. Ermolaeva O, Rastogi M, Pruitt KD, *et al.* Data management and analysis for gene expression arrays. *Nat Genet* 1998, **20**, 19–23.
70. Griffiths AD, Duncan AR. Strategies for selection of antibodies by phage display. *Curr Opin Biotechnol* 1998, **9**, 102–108.
71. Henn W. Genetic screening with the DNA chip: a new Pandora's box? *J Med Ethics* 1999, **25**, 200–203.
72. Cole KA, Krizman DB, Emmert-Buck MR. The genetics of cancer—a 3D model. *Nature Genet* 1999, **21**, 38–41.
73. Goldenberg MM. Trastuzumab, a recombinant DNA derived humanized monoclonal antibody, a novel agent for the treatment of metastatic breast cancer. *Clin Therw* 1999, **21**, 309–318.
74. Jacobs TW, Gown AM, Yaziji H, Barnes MJ, Schnitt SJ. Specificity of HercepTest in determining HER-2/new status of breast cancers using the United States Food and Drug Administration approved scoring system. *J Clin Oncol* 1999, **17**, 1983–1987.
75. Quirke P, Mapstone N. The new biology: histopathology. *Lancet* 1999, **354**, S26–S31.
76. Jones D, Fletcher CDM. How shall we apply the new biology to diagnostics in surgical pathology? *J Pathol* 1999, **187**, 147–154.