PERGAMON

European Journal of Cancer 39 (2003) 1199–1215

European Journal of Cancer

www.ejconline.com

Review

Genomics and proteomics in cancer

J.P.A. Baak[a,b,*], F.R.C. Path[a,b], M.A.J.A. Hermsen[c], G. Meijer[c], J. Schmidt[d], E.A.M. Janssen[a]

[a]*Department of Pathology, Central Hospital for Rogaland, Box 8001, 4068 Stavanger, Norway*
[b]*VU Free University, Amsterdam, The Netherlands*
[c]*Department of Pathology, VU Medical Center, Amsterdam, The Netherlands*
[d]*Competence Center for Bio-Molecular Therapeutics, Vienna, Austria*

Received 28 February 2003; accepted 4 March 2003

Abstract

Cancer development is driven by the accumulation of DNA changes in the approximately 40 000 chromosomal genes. In solid tumours, chromosomal numerical/structural aberrations are common. DNA repair defects may lead to genome-wide genetic instability, which can drive further cancer progression. The genes code the actual players in the cellular processes, the 100 000–10 million proteins, which in (pre)malignant cells can also be altered in a variety of ways. Over the past decade, our knowledge of the human genome and ***Genomics*** (*the study of the human genome*) in (pre)malignancies has increased enormously and ***Proteomics*** (*the analysis of the protein complement of the genome*) has taken off as well. Both will play an increasingly important role. In this article, a short description of the essential molecular biological cell processes is given. Important genomic and proteomic research methods are described and illustrated. Applications are still limited, but the evidence so far is exciting. Will genomics replace classical diagnostic or prognostic procedures? In breast cancers, the gene expression array is stronger than classical criteria, but in endometrial hyperplasia, quantitative morphological features are more cost-effective than genetic testing. It is still too early to make strong statements, the more so because it is expected that genomics and proteomics will expand rapidly. However, it is likely that they will take a central place in the understanding, diagnosis, monitoring and treatment of (pre)cancers of many different sites.

Keywords: Cancer; Genomics; Proteomics; Chromosomal instability; DNA repair; Cytogenetics; FISH; CGH; SKY; PCR; MSI

1. Introduction

Cancer development is driven by the accumulation of DNA changes in some of the approximately 40 000 chromosomal genes. In solid tumours (90% of all human malignancies), chromosomal numerical/structural aberrations are common. DNA repair defects may lead to genome-wide genetic instability, which can drive further cancer progression. The genes are the code for the actual players in the cellular processes, the 100 000–10 million proteins, which in (pre)malignant cells can also be altered in a variety of ways. Over the past decade, our knowledge of the human genome in (pre)-malignancies has increased enormously. ***Genomics*** (*the study of the human genome*) and ***proteomics*** (*the analysis of the protein complement of the genome*) play a major role in the understanding, diagnosis, prognosis and potentially also treatment of cancer. In this article, a summary of both will be given. We start with a short description of the essential molecular biological processes and then describe important genomic and proteomic research methods. Applications will be given as illustration.

2. Essential molecular biological processes

2.1. Altering the genetic message—type of events

Altering the genetic message is essential for cancer development and can be caused by (1) DNA mutations; (2) Chromosomal aberrations; (3) Epigenetic modification; (4) Protein–protein interaction.

DNA mutations can lead to increased or decreased gene function. There are oncogenes (that increase pro-

* Corresponding author. Tel.: +47-5151-9368; fax: +47-5151-9910.
E-mail address: baja@sir.no (J.P.A. Baak).

0959-8049/03/$ - see front matter © 2003 Elsevier Science Ltd. All rights reserved.
doi:10.1016/S0959-8049(03)00265-X

liferation) and suppressor genes (that decrease proliferation). Gain of an oncogene function can change the gene product; for example, it can result in an enzyme with a higher affinity for its ligand, or a lower degradation susceptibility. Alternatively, the gene regulatory properties may change and more actively express functional proteins. Such mutations are often base pair-substitutions. In contrast, inactivating tumour suppressor gene mutations often are base pair-deletions leading to a shift in the open-reading-frame of a gene. This causes a premature stop codon. Base pair-substitutions can also directly cause a stop codon. As a result, the gene is only partly transcribed and functional protein production terminated.

DNA modification, mutations and viral genome integration can give *chromosomal instability*. Chromosomal aberrations themselves can lead to altered genes or gene expression; both genetic instabilities exist and are interrelated. Tumours can be near-diploid (around 46 chromosomes), or aneuploid (often 60–90 chromosomes in solid tumours). Already in 1933, DNA ploidy could be measured by cytometry [1], a method still clinically useful in specific areas. Individual chromosomes can also show structural changes, e.g. partial deletions or duplications, inversions, unbalanced and reciprocal translocations and amplifications. Balanced, reciprocal translocations are common in haematological malignancies, while in solid tumours virtually all rearrangements are unbalanced resulting in losses/gains of chromosome parts. Numerical changes of specific chromosomal regions have been correlated with expression levels of genes in these regions. By screening many tumours, recurrent losses and gains can be identified and subsequent investigation of regions of overlap can reveal the genes involved.

Gene expression can be silenced by *epigenetic modification* (which does not change the DNA code, but the accessibility of DNA for transcription). This occurs mainly through DNA methylation, which can directly silence genes by interfering with the binding of transcription factors to the recognition elements of genes, or indirectly by changing the chromatin structure around the gene. Although the genome of cancer cells compared with normal cells is globally hypomethylated, some sequences are hypermethylated, such as CpG islands, which lie around the transcription-start sites of approximately half of all human genes. For example, p16 is hypermethylated and silenced in many cancers [2]. DNA hypomethylation correlates with chromosome instability. DNA methylation may regulate histone acetylation and methylation, which in turn changes chromatin structure [3]. Chromatin architecture remodelling is essential for gene-transcription, and thus gene expression [4], explaining the prognostic significance of quantitative chromatin changes [5–7]. *Proteins* can be directly inactivated by chemicals (e.g. cytostatics) or by complexing with other proteins. For example, Human Papilloma Virus (HPV) and Epstein–Barr Virus (EBV) oncoproteins binding with tumour suppressor *p53* or the retinoblastoma gene result in rapid loss of their key function of cell-cycle regulation and genome integrity maintenance.

2.2. The multistep tumorigenesis hypothesis—how many events are necessary?

The number of gene defects (=hits) necessary for a normal cell to become neoplastic is tumour-dependent. Leukaemia cells are relatively stable, frequently characterised by a *single*, non-random reciprocal chromosomal translocation [8] and not affected by general random genome instability often seen in solid tumours [9]. An example of juxtaposition is the t(8;14) translocation, resulting in c-myc overexpression in Burkitt's lymphoma. A gene-fusion example is the t(9;22) translocation (Philadelphia chromosome), resulting in the BCR-ABL chimeric protein in chronic myeloid leukaemia.

Knudson [10] proposed that *two* genetic events are required for tumour formation in hereditary retinoblastoma; one the inherited gene defect and the second being the inactivation of the remaining normal allele. Characterisation of the retinoblastoma and other genes involved in inherited cancer predispositions has led to the definition of tumour suppressor genes. Usually one allele is inactivated by deletion and the other by mutation. The best-known tumour suppressor gene is *p53*, involved in a variety of solid tumours. Kinzler and Vogelstein refined this model by dividing tumour suppressor genes into caretakers and gatekeepers [11]. *Caretakers* are DNA repair genes involved in genome integrity maintenance; failure leads to genome hypermutability. Chromosomal instability accelerates cancer progression and is reflected in the young age of cancer onset in patients with hereditary DNA repair genes defects. *Gatekeepers* (or checkpoints) guard the diverse processes of the cell cycle. Defects in gatekeepers are rate-limiting for tumour growth, and *two* inactivating hits are required for tumour initiation. Without a proper gatekeeper function, DNA mutations or chromosome misdivisions may pass uncorrected to the daughter cells. Defects in caretakers lead to an increase in genetic alterations, and can accelerate cancer development via accumulation of hits in oncogenes and/or gatekeepers. In the caretaker pathway, at least *three or more* hits are necessary. Two or three genetic events still do not seem to fit the general observation in sporadic solid tumours that cancer incidence is correlated with increasing age. This observation, in combination with the natural rate of mutation, predicts that *six or more* genetic hits are required for tumour formation [12]. In addition, many solid cancers are known to have pre-

cursor lesions that show increased proliferation combined with abnormal differentiation, but stay within the original tissue boundaries. This has led to the idea of a sequence of progression steps from normal tissue, via (a number of) intermediate stages to cancer and finally to metastasis: the *multistep* hypothesis. Using colorectal tumours, a sequence model of molecular genetic changes leading to adenocarcinoma was proposed [13]. In this model, *K-ras* mutation and loss of heterozygosity (LOH) on chromosome 5q (the adenomatous polyposis coli (*APC*) tumour suppressor gene locus) are early genetic changes, whereas LOH of 18q (where initially deleted in colon cancer (*DCC*) was thought to be the gene involved, now *Smad4* is the candidate gene) and 17p (*p53*) were considered late events. Such a model has also been proposed for other solid tumours. For head and neck squamous cell carcinoma, LOH on 9p21, the locus of the cell-cycle inhibitor *p16*, would be the primary event, followed by LOH on 17p13 (the *p53* locus) and 3p21, and subsequent LOH on 11q13, 13q21, and 14q32 in an undetermined order [14].

Even these more elaborate multistep models are still simplifications of reality. Further studies are needed to provide more insight into the ways in which critical cell biological functions like proliferation, differentiation, apoptosis and invasion are affected by genetic changes, and how they contribute to tumour progression.

2.3. Early or late events?

Some genetic changes occur early (tumour-initiating hits) and others late in tumour development (progression-related events). Inactivation of tumour suppressors at the protein level by viral oncoproteins is an early event. Indeed, HPV-DNA has been detected in the earliest premalignant lesions. Continuous expression of the viral oncoproteins E6 and E7 seems necessary for cervical carcinoma development [15]. However, the long latency period between viral infection and the formation of invasive cancer indicates that additional genetic events are required for malignant transformation. In addition, gene expression modification by DNA methylation is an early event. It has been suggested that hypomethylation may lead to chromosome decondensation and mitotic nondisjunction, eventually leading to gain and loss of (parts of) chromosomes [14]. In colorectal cancers, very early lesions known as Aberrant Crypt Foci show aberrant gene methylation [16]. Activating oncogenes mutations have been implicated in all stages of solid tumour development [17]. Inactivation of a tumour suppressor gene requires two hits, suggesting that these events occur later in tumorigenesis. However, in colorectal cancer, *APC* tumour suppressor inactivation is the earliest factor causing the development of adenomas [18]. Mutation of *p53* is regarded as a late event in the progression of colorectal tumours, whereas *p53* inactivation seems to be an early event in head and neck carcinomas.

Chromosomal abnormalities are common in virtually all late-stage solid tumours but, historically, it was difficult to appreciate the meaning of recurrent and tumour-specific chromosomal changes against a background of a multitude of alterations [19,20]. Some researchers believed they were merely side-effects of genetic instability driven by preceding gene mutations. This controversy was maintained for a long time, mainly because data on chromosomal aberrations in early, pre-invasive tumours were scarce. However, Comparative Genomic Hybridisation (CGH) [21], has made clear that chromosomal abnormalities also occur in early premalignant stages [18,22–26]. It has recently been demonstrated that a single copy gain of a certain gene can have a strong effect on the downstream expression levels [27] showing that chromosomal changes are not merely side-effects of tumour progression.

2.4. Mechanisms of genomic/chromosomal instability

2.4.1. Hereditary caretaker and gatekeeper defects

A number of hereditary genomic instability syndromes due to caretaker defects are known. Most are very rare. Some cause genomic instability at the DNA level, some at the chromosome level, and some both. Examples are the *Bloom's Syndrome* [28], *Xeroderma Pigmentosum* (*XP*) [29], *Fanconi Anemia* (*FA* [30]) and *Hereditary Non-Polyposis Coli Cancer* (*HNPCC* [31]). *Retinoblastoma* is an example of a hereditary genomic instability syndrome due to gatekeeper abnormality [32]. Other examples are the *Li–Fraumeni syndrome* [33], *familial melanoma* [34] and *familial breast cancer* [35].

2.4.2. Genomic instability in sporadic tumours

Chromosomal instability is a genome-wide phenomenon. According to one theory, chromosomal aberrations result from specific oncogenes mutations (like *ras*, *myc*, *p53*) [36], but it is unlikely that they would be solely responsible. Another theory suggests that a mis-division of a certain chromosome gives rise to a daughter-cell with a growth advantage and abnormal chromosomal number, which then would cause further chromosomal instability [37]. Perhaps chromosomal instability can be compared with microsatellite instability, which also causes genome-wide occurrence of genetic events [20]. As microsatellite instability is the result of a mutation in a mismatch repair gene, likewise chromosomal instability could be caused by a chromosome-stability gene mutation. It has been proposed that *APC* could be such a gene [38,39]. However, some colon cell lines with *APC* mutations have maintained a stable chromosome complement over thousands of cell divisions *in vitro* [40,41]. Moreover, around 70% of human colorectal adenomas carry *APC* mutations, whereas

only 5% of all adenomas ever progress to invasive cancer. Other candidate genes comprise those involved in the mitotic-spindle checkpoint and sister-chromatid separation pathways [20,42,43].

2.5. Molecular DNA techniques

2.5.1. LOH, MSI

One strategy for screening the genome in breast cancer used a panel of 150 *polymorphic microsatellite markers* from throughout the whole genome (with Loss of Heterozygozity (LOH)) [44]. This allowed the identification of several chromosomal regions where allelic losses occurred more often. Correlation with clinicopathological factors showed that four specific loci correlated with lymph-node metastasis (11q23-24, 13q12, 17p13.3 and 22q13). By using microsatellites, one can also check the ability of cells to repair DNA replication errors. In tumours with *microsatellite instability* (*MSI*), both genetic and epigenetic modifications of mismatch repair genes were identified. Some cases harboured mutations or polymorphisms (=natural variation between different humans in their genetic codes) in *hMSH2* and *hMLH1* mismatch repair genes, others exhibited hypermethylation in the promoter region of *hMLH1*. These results suggest that both genetic and epigenetic alterations of *hMSH2*, and especially of *hMLH1*, contribute to genomic instability and tumorigenesis in sporadic breast cancer [45] prognosis.

2.5.2. Conventional karyotyping

Already in 1890 Hansemann [46] observed *asymmetric mitoses* in tumour cells and considered the imbalance as the cause of malignant growth. In 1956, the number of 46 human chromosomes was established [47]. The development in 1969 of chromosome banding techniques was a breakthrough as all chromosomes could be individually recognised (=*conventional karyotyping*, Fig. 1) and chromosomal rearrangements characterised. The first of these was the Philadelphia chromosome in chronic myeloid leukaemia [48].

2.5.3. *Fluorescence* in situ *hybridisation (FISH)*

FISH was developed in the 1980s, and rapidly became a powerful technique [49]. A small DNA fragment of known origin (a probe) is fluorescently labelled and hybridised to a metaphase chromosome spread or interphase nuclei. The probe binds to homologous sequences within the chromosomes, and this can be visualised by fluorescence microscopy. FISH can identify whole chromosomes, centromeres, telomeres, specific regions or genes, or aberrations in interphase tumour nuclei (Fig. 2).

2.5.4. CGH

CGH [21] provides in one experiment information on the number of copies of chromosome parts throughout the whole tumour genome. A considerable number of CGH studies have been performed on many types of solid tumours, also with clinical follow-up, enabling the identification of genetic changes that are related to prognosis [50,51]. The resolution of chromosome CGH is limited to approximately 10 million base pairs (Mbp). After washing and detection with the antibodies avidin-fluorescein isothiocyanate (FITC) (green fluorescent) and sheep-antidigoxigenin-tetramethyl rhodamine isothiocyanate (TRITC) (red fluorescent), the green to red fluorescence ratio along the chromosomes is measured by digital image analysis using dedicated software (Fig. 3). A gain or amplification in the tumour genome will be visualised by an excess of the green signal (and a fluorescence ratio >1.0), whereas a deletion or loss in the tumour genome is seen by an excess of red signal (and a fluorescence ratio <1.0).

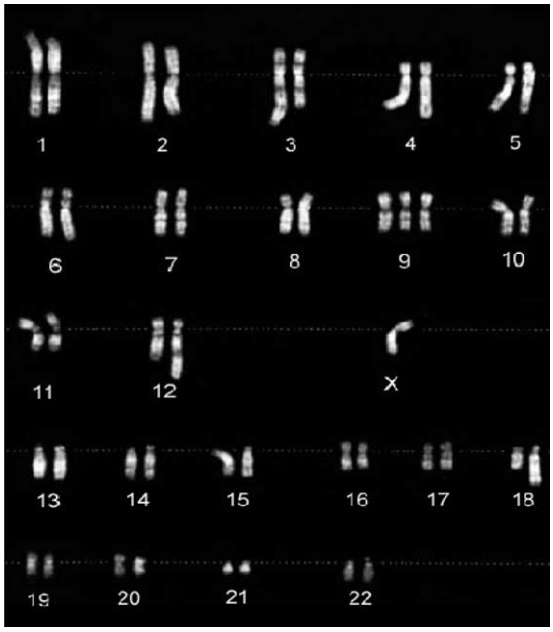

Fig. 1. Representative Q-banded karyogram of tumour cell line scc040 showing chromosomes 3, 12 and 18 with structural rearrangements, and chromosomes 9 and X with copy number abnormality (trisomy and monosomy, respectively).

2.5.5. Array CGH

Array CGH does not require karyotyping, the resolution can be 0.5 Mbp or better and it has a higher sensitivity than conventional CGH [52,53]. Rather than using metaphase spreads, an array of very small spots of genomic DNA (BAC clones, representing different genes or chromosome locations) is used as the target for hybridisation (Fig. 4). Fig. 5 shows the results of chromosome and array CGH of the same chromosome combined in one figure. Chromosome CGH detected a low level gain of chromosome arm 20q (the non-horizontal lines, arrows, indicate the mean fluorescence ratio with 95% Confidence Limits (CL)), while array

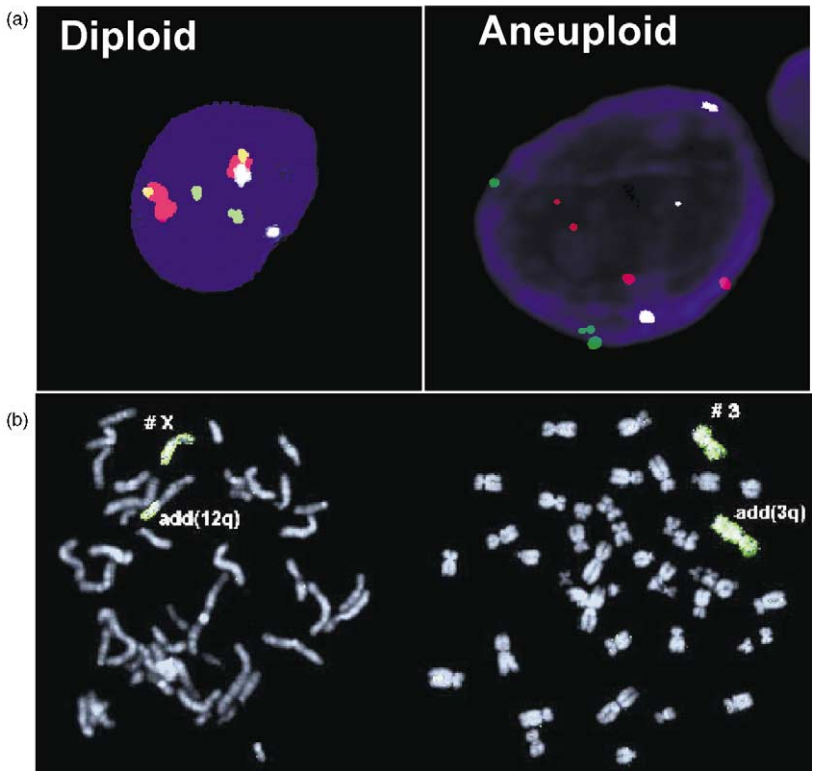

Fig. 2. Fluorescence *in situ* hybridisation (FISH) examples. Top (a), Interphase cytogenetics. Normal diploid (two spots per chromosome) and malignant (right, several spots per chromosome, indicating its aneuploidy) urothelial cells from a urinary bladder washing hybridised with centromere probes to chromosomes 3 (red), 7 (green) and 17 (white), and a locus-specific probe for *p16* at 9p21 (yellow). Bottom (b) FISH using whole chromosome X and chromosome 3 probe on tumour cell line scc040, showing that a part of chromosome X is attached to chromosome 12, and showing that the addition on chromosome 3 is in fact other chromosome 3 material.

CGH revealed a narrow high level amplification, 800 kbp wide [54]. Recently, an array CGH has been presented carrying 2400 FISH- and radiation hybrid mapped BAC-clones, covering the whole human genome with an average spacing of 0.8–1.4 Mbp [55]. Applied to tumour cell line scc040, chromosome CGH detected a gain at 3qter, the Xq material on chromosome 12, and it detected a gain of 8q, which must represent the extra material on chromosome 18. Using the genome-wide array, the same gains and losses were detected, but now the breakpoint positions of unbalanced translocations could be more accurately pinpointed at 3q24, 8q22, 18q23 and Xq21.1. Combined CGH and cDNA microarrays detected that a set of 270 genes was statistically attributable to gene amplification, and a high-resolution map of 24 amplicons was provided [56]. Characterisation of a novel amplicon at 17q21.3 implicated amplification and overexpression of the *HOXB7* gene in breast cancer including its association with poor patient outcome.

2.5.6. Spectral karyotyping (SKY)

SKY [57] gives all chromosomes a unique fluorescent colour by hybridising with chromosome-specific probes each labelled with a different combination of fluorochromes. SKY is therefore especially useful for the genome-wide detection of structural chromosomal changes (contrasting CGH, which yields numerical data only). Tumour metaphase slides are hybridised simultaneously with 24 probes, and digital images of the metaphases are captured with a fluorescent microscope using a special camera, coupled with a spectral colour analysis system that scans the spectrum from 450 to 750 nm for each pixel in the image. With dedicated image analysis software a final image is created showing all chromosomes with their unique colour, allowing easy identification of structural changes. Translocations are readily visible (Fig. 6). In contrast to analyses of structural chromosomal rearrangements by polymerase chain reaction (PCR) and FISH, SKY can screen genome-wide, and does not require prior knowledge of chromosomal breakpoints.

2.5.7. Gene expression (cDNA) arrays

The straightforward gene approaches described above have been very successful. An integrated, genome-wide picture of changing gene function has proved more elusive. This is partly due to the limited number of genes studied, even in detailed analyses, and interpretive problems in assessing the level of gene activity in isolated gene reports from multiple laboratories using variable methodologies. One solution to this problem is to increase the number of genes analysed in common tissues. This is now possible with high-density oligo-

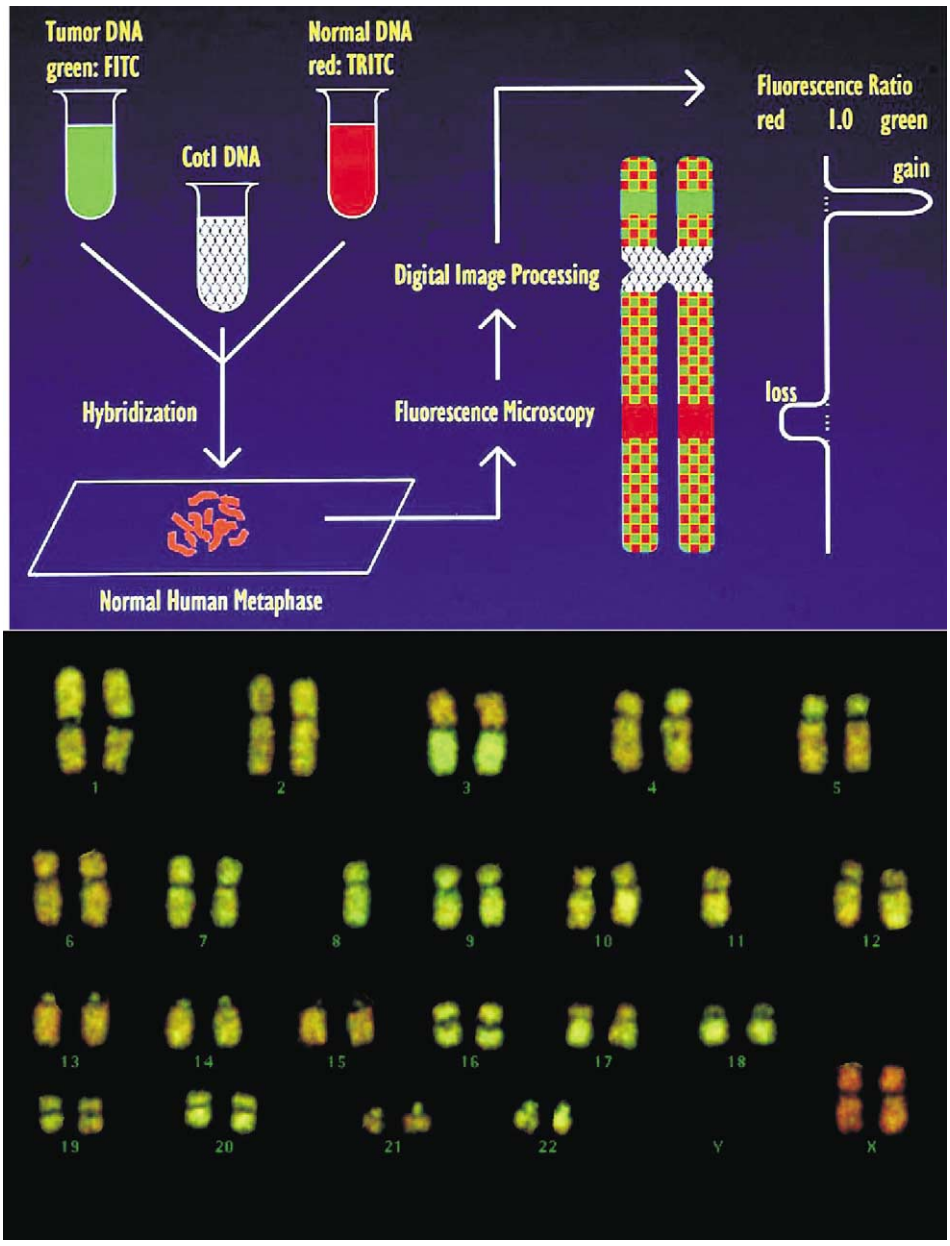

Fig. 3. Schematic view of the comparative genomic hybridisation technique (CGH) (top), with an example of a hybridisation of green-labelled neuroblastoma DNA and red-labelled normal reference DNA onto normal metaphase chromosomes (bottom). Copy number changes can be seen at chromosomes 3q, 5p, 8, 9, 16,17q, 18q, 20q, 22p and X. FITC, fluorescein isothiocyanate. TRITC, tetramethyl rhodamine isothiocyanate.

nucleotide cDNA microarrays technology, which measures in parallel many thousands of gene-specific mRNAs in a single tissue sample. The principle of Gene Expression Arrays is comparable to Array CGH (apart from using oligonucleotides). The Affymetrix GeneChip arrays are most popular [58]. Large-scale gene expression analysis has proved to be a valid strategy for developing gene expression profiles, or 'signatures', which are useful in classifying tissues according to pathological or prognostic subgroups. Although each gene may be considered or evaluated in isolation, the unique quality of these datasets resides in their ability to define coordinate expression patterns capable of representing a large number of relevant variables, and high degree of interaction. A sobering indication of the downstream effects referable to a single gene may be obtained by observing expression alterations caused by controlled changes in the gene of interest, as was done for *c-myc* in human cells [59]. cDNA arrays have been used to classify pathological subgroups of leukaemias [60] and potentially prognostic subgroups of melanomas [61,62] by relative expression levels of a panel of genes. Constructing and displaying models, which incorporate dozens or hundreds of interacting quantitative vari-

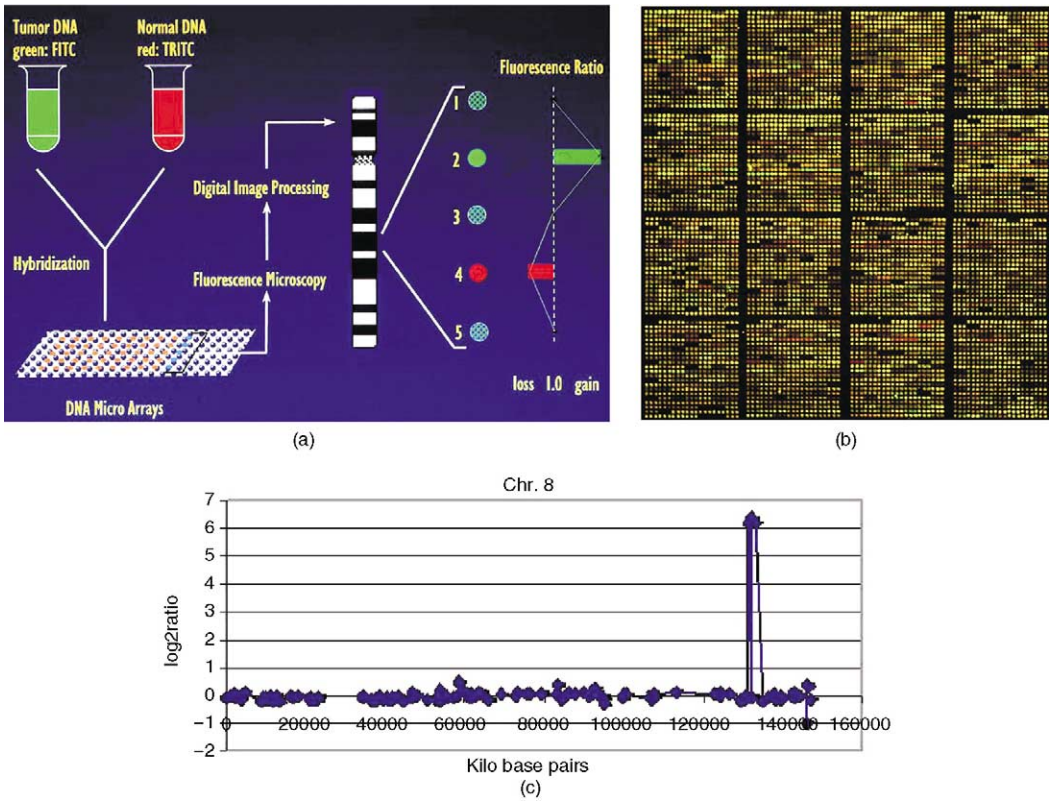

Fig. 4. Schematic view of the array CGH technique (a); example of a hybridisation of green labelled tumour DNA and red labelled normal reference DNA onto an array of 2400 BAC clones representing the whole genome with an average spacing of 1 Mbp (b). DNA copy number information per chromosome is presented horizontally; shown is a complete normal chromosome (Chr.) 8 with an amplification of the *c-myc* region (c).

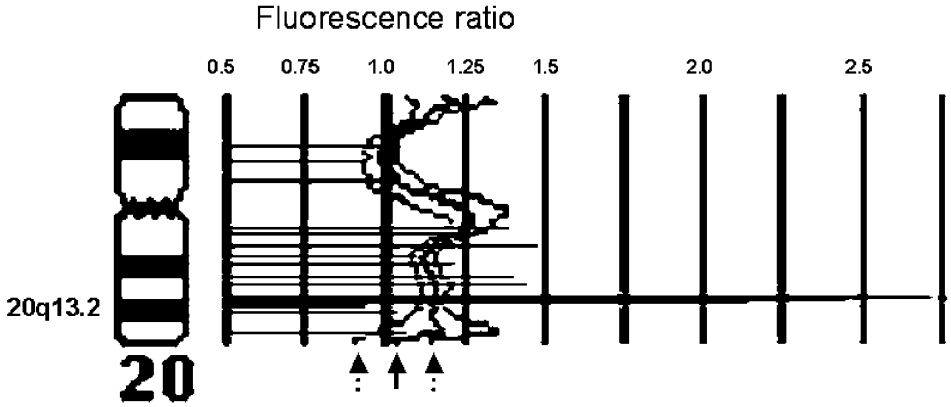

Fig. 5. Example of the higher sensitivity and resolution of array CGH over chromosome CGH: chromosome CGH detected a low level gain of a large part of chromosome arm 20q; array CGH identified a narrow high level amplification in band 20q13.2. The line with the arrow is the green, red/green. Fluorescence intensity ratio, the dotted arrows show the 95% confidence interval.

ables, is a new challenge for most biologists and requires special statistical analyses to avoid incidental significances [63]. In collaboration with Dr. G. L. Matter, Harvard Medical School, Boston, MA, USA, we have used the Permax test [64] using a program that is available free on-line [65]. The *t*-statistics have a tendency to preferentially select genes with very small intragroup variances. After determining the most significant genes from the *t*-statistics (Permax <0.5), those genes with absolute differences between means $\geqslant 100$, and ratios of means $\geqslant 3$ were identified. This selection process provides a conservative determination of genes whose expression levels are changed in different conditions. The top upregulated and downregulated *t*-test-ranked genes are identified. LocusLink ID [66] gene identifiers and synonyms can be used as search terms to retrieve citations of selected genes from PubMed [67,68]. An example is given in Table 1. Gene expression array

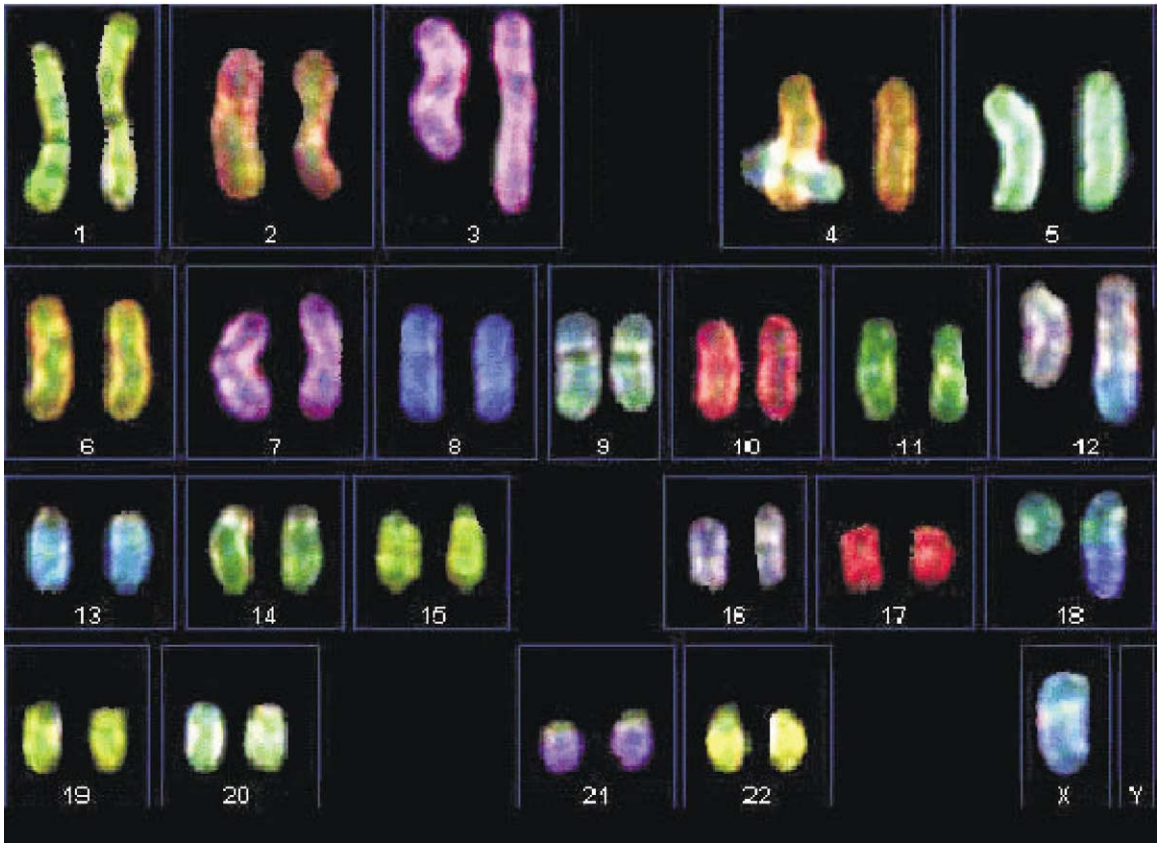

Fig. 6. Spectral karyotyping (SKY) karyogram of tumour cell line scc040. For every pixel in the image, the spectrum from 450 to 750 nm is measured, and with this spectral information the unique colour combination of each chromosome is identified. The addition on chromosome 3 is composed of chromosome 3 material, the addition on chromosome 12 comes from a part of chromosome X, and the addition on chromosome 18 consists of chromosome 8 material (the third chromosome 9 is lying in overlap with chromosome 4).

technology has been used to study neoplastic transformation in endometrial (pre)cancer [64]. Figs. 7 and 8 show the initial selection results of genes which characterise carcinoma. We found that 100 genes, which are hormonally regulated in normal tissues, are expressed in a disordered and heterogeneous fashion in cancers, with tumours resembling proliferative more than secretory endometrium. Fifty genes that distinguish normal from malignant endometrium were assigned to functional categories based on literature citations. A total of 16 620 literature citations were recovered for the 50 genes, and 233 of these mention endometrium somewhere in the abstract or title. Table 1 gives an excerpt of these 50 genes.

In breast cancer, tumours with distinct gene expression patterns were termed 'basal type' (ER+, Erb-B2−, good outcome) and 'luminal type' (ER−, Erb-B2+, worse outcome [69]). cDNAs have also been used to distinguish cancers with *BRCA1*/*BRCA2* mutations [70,71], to determine oestrogen-receptor status [72], lymph-node status [73,74] and prognostic subgroups in node-negative breast cancers [75]. So cDNA-microarrays can potentially be used in clinical practice. Unfortunately, it requires fresh tissue.

3. Proteomics

3.1. Background

Less than 2% of diseases are monogenic (caused by the mutation of a single gene), 98% of major diseases involve multiple genes (polygenic). Revealing the disease pathways depends heavily on human proteins (=proteome) analysis. Proteomics encompasses knowledge of the structure, function and expression of all proteins as a function of time, age, state and external factors [76–78]. The "Human Protein Index" was developed in the late 1970s [79] to enumerate proteins by separation on two-dimensional (2D) gels and define their genes from the protein end. Today, only a small percentage of the approximately 100 000–10 000 000 proteins have been well described. So, the task ahead to elucidate the structure of all these proteins, their function, post-translational modifications and interactions with other proteins/protein complexes/DNA/RNA/small molecules is enormous. However, it should at least be possible to understand the structure and basic function of each protein. This is not enough to understand cell and organ function, but it

Table 1
Example of the use of LocusLink and PerMax in the discrimination of normal and malignant endometrium by gene expressions arrays

GeneCode	LocusLink	ABREV	Gene class/name	Normal	Tumour	Fold diff.	Permax
Cell cycle							
x3265	894	CCND2	Cyclin D2	1766	175	10.1	0.468
x4985	1875	E2F5	E2F transcription factor 5, p130-binding	33	146	4.4	0.2448
Cell–cell interaction							
x723	3956	LGALS1	Lectin, galactoside-binding, soluble, 1 (galectin 1)	8915	1537	5.8	0.2478
x6986	6387	SDF1	Stromal cell-derived factor 1	677	49	13.8	0.3417
Cytoskeleton/adhesion							
x3120	1009	CDH11	Cadherin 11 (OB-cadherin, osteoblast)	504	57	8.8	0.5
x3108	1742	DLG4	Discs, large (Drosophila) homologue 4	278	20	13.9	0.028
Extracellular matrix							
x879	4811	NID	Nidogen (enactin)	1204	54	22.3	0.2458
Hormone response/pathway							
x671	1909	EDNRA	Endothelin receptor type A	656	58	11.3	0.446
x2670	5047	PAEP(alt2)	Progestagen-assoc. endometrial prot.: Alt. Splice 2	4998	32	156.2	0.2478
Metabolism, amino acid							
x197	3931	LCAT	Lecithin-cholesterol acyltransferase	873	20	43.7	0.035
Miscellaneous							
x2797	2072	ERCC4	Excision repair cross-complementing rodent repair deficiency	20	509	25.5	0.0959
x2719	3452	IFNA21	Interferon, alpha 21	236	30	7.9	0.2108
x6197	3660	IRF2	Interferon regulatory factor 2	208	20	10.4	0.1808
Oncogene/tumour suppressor							
x2035	8045	C11ORF13	Chromosome 11 open reading frame 13	20	918	45.9	0.2727
x5442	3479	IGF1	Insulin-like growth factor 1 (somatomedin C)	1490	35	42.6	0.1618
x6701	5156	PDGFRA	Platelet-derived growth factor receptor, alpha polypeptide	1167	65	18.0	0.3077
x4880	51352	WIT-1	Wilms' tumour-associated protein	260	20	13.0	0.038
Toxin scavenger/activator							
x1629	1545	CYP1B1	Cytochrome P450, subfamily I, polypeptide 1	179	20	9.0	0.2038
Vesicle/synapse associated							
x4516	10991	G17	G17 transporter protein	20	781	39.1	0.3247
x4495	10578	GNLY	Granulysin	3585	64	56.0	0.2218
x881	3176	HNMT	Histamine *N*-methyltransferase	205	20	10.3	0.0599

assoc., associated; prot., protein; Alt., alternative; diff., difference. This is an excerpt from Table 2 in Ref. [64], p. 182, and serves as an illustration only of the technique used.

would be a step forward and provide new molecular-biological insights, tumour markers and drugs.

3.2. Proteomics techniques

3.2.1. 2D-PAGE

2D poly-acrylamide gel electrophoresis (2D-PAGE = 2DE) [16] is a powerful technique for protein separation, first according to pH (isoelectric point) and then according to size (molecular weight). Electrophoresis gels are stained with colloidal Coomassie Brilliant Blue G, silver stain or a fluorescent stain. Commonly between 500 and 2000 spots can be detected on large gels. The digestion of spots, obtained by gel excision, directly in the gel makes it possible to further analyse proteins with Mass Spectrometry (MS). Peptide spectra obtained in this way can be used to search protein sequence databases (for example, NCBI, SWISSPROT, HGMP, Biobase/Celis database, OWL and others, see end of Reference list for Internet addresses) for identification. Since the technique has been used for over 20 years, there is potentially a vast amount of information on proteins in different databases, but exploitation is hampered by difficulties with large-scale data analysis (Fig. 9). Comparison of the results of only two experiments for a differential expression profile can be difficult and not all membrane and other hydrophobic proteins enter 2D-gels effectively. In addition, the sensitivity is never below the micromolar (10^{-6}) range, unless more material is applied (which can be clinically impossible), or enriched by Laser Capture Microdissection (which requires minimally 50 000 cells [80]). Yet, 2D gel electrophoresis is still the best option for large-scale protein separation. Gels with normal and malignant breast samples expressed growth factor protein receptors only in malignant samples, whereas Maspin expression was more intense in benign breast samples [81]. Nipple aspiration fluid from both breasts of patients with uni-

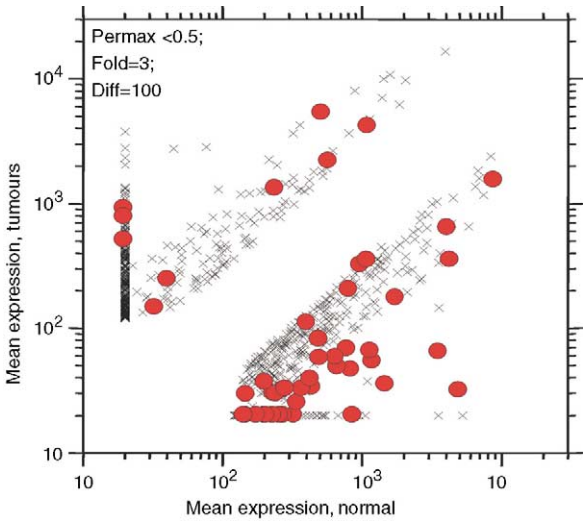

Fig. 7. Permax selected genes which discriminate normal from malignant endometrium. Scatter plot of 753 expressed genes (crosses) in normal (x axis, mean of four samples) and malignant (y axis, mean of 10 samples) endometrial tissues in which the means of normal and tumour tissues are separated by a minimum of 100 expression units and have a ratio difference = 3. t-Statistics calculated from the non-permuted dataset were compared with the cumulative frequency distribution of the permuted t-statistics to assign a probability of error (Permax). Solid circles indicate 50 selected genes with Permax values less than 0.50. From Ref. [64], p. 181. Diff, difference.

lateral breast cancer showed substantial qualitative differences in the protein expression pattern [82].

3.2.2. Other methods

For protein identification/characterisation, there is a wide scope of techniques such as MS, High Performance Liquid Chromatography (HPLC), Capillary Array (CA), Matrix Associated Laser Desorption/Ionisation (MALDI) and a growing range of modifications of these techniques (MALDI-Time Of Flight (TOF)-MS, MALDI-ion trap-TOF-MS, ESI (electro-spray ionisation) tandem MS, Quadrupole). These allow the mass fingerprinting to be immediately followed by tandem MS of peptide ions. High sensitivity (into the femtomolar = 10^{-15} range) and specificity are combined by a high flow throughput. Internal labelled markers make it possible to give semi-quantitative results, but accurate quantitation remains a problem. In addition, sample preparation is extensive and background signals from organic and inorganic compounds can be a problem.

3.2.3. Functional proteomics techniques

For functional proteomics other techniques exist. The most important are the yeast two-hybrid system, protein microarrays, Fluorescence Resonance Energy Transfer (FRET), Surface-Enhanced Laser Desorption/Ionisation and tissue microarray.

3.2.4. Yeast two-hybrid system

Since the development of the yeast two-hybrid system (Y2HS) [83], it has become one of the most popular tools in biology [84]. As its name implies, yeast cells are used as biochemical reaction 'vessels' involving two hybrid fusion proteins. Attaching the target protein to the DNA-binding domain of a yeast transcription factor (usually Gal4) creates a so-called 'bait protein'. This protein is then expressed in appropriate reporter yeast cells (lacking functional Gal4 but containing Gal4-dependent growth selection genes) together with a library of genes or gene fragments fused to the activation domain of Gal4. Interaction between the bait protein and a corresponding library protein reconstitutes a functional Gal4 transcription factor and leads to growth of this yeast cell on the selection plate. With this method, 1440 distinct proteins could be identified within 232 large multiprotein complexes, of which 231 were previously uncharacterised proteins [85]. The two-hybrid system can generate vast amounts of new data, but each and every complex has to be tested and confirmed separately afterwards. A recent Y2HS study of specific protein interaction data (Ran and cyclin-B) suggested that pKI-67 is a participant and regulator within the cell cycle, influencing reformation of the nucleolus after mitosis and anchoring of chromosomes to the inner side of the nuclear lamina [86].

3.2.5. Protein microarrays

Most protein microarrays are affinity-based. In a *sandwich immunoassay* [87], capture antibodies are immobilised on the solid support, and bound proteins are detected using a second, labelled detection antibody. In an *antigen capture assay* [88], immobilised antibodies similarly capture proteins, but the captured proteins are usually detected directly by chemically labelling the complex protein mixture before applying them to the array. In the *two-colour version*, two samples are labelled independently with distinguishable fluorophores and mixed before applying them to the array. In a *direct assay* [89], the complex mixture of proteins is itself immobilised on the solid support, and specific proteins in that mixture are visualised using labelled detection antibodies. Although powerful, these techniques have several drawbacks: lack of available antibodies, lack of purified recombinant proteins and cross reactivity with affinity agents.

3.2.6. Fluorescence resonance energy transfer (FRET)

FRET is a powerful method to monitor macromolecular interactions and to resolve the spatial and temporal dynamics of protein–protein interactions in the living cell. The basis of the method is a phenomenon that occurs between two fluorophores in close proximity (usually <10 Å, Fig. 10), if the emission curve of one of them (the energy donor) overlaps with the excitation

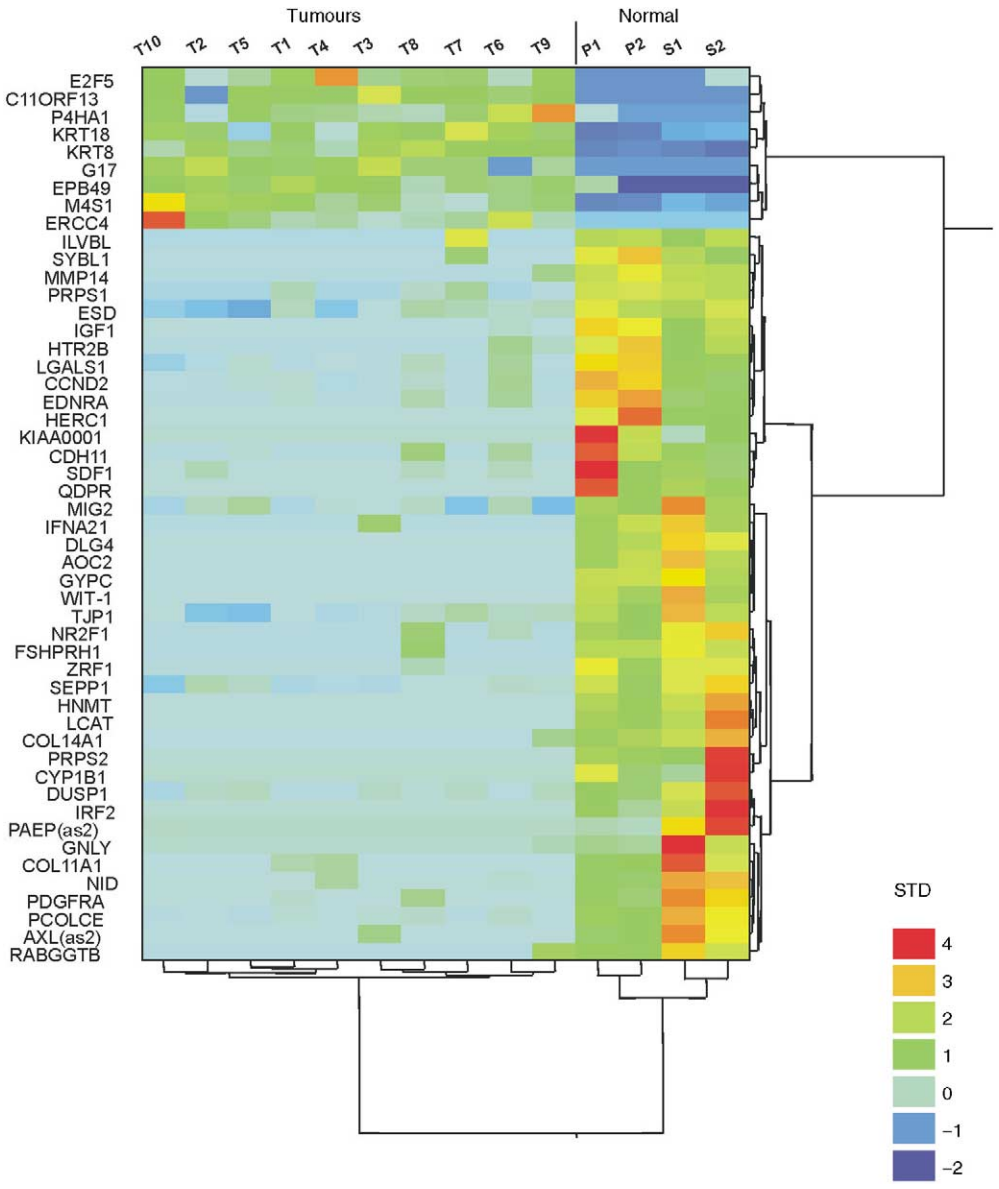

Fig. 8. Genes which distinguish normal from malignant endometrium (P, proliferative; S, secretory; T, tumours) (Permax < 0.50, 3-fold, 100 difference). Columns show individual tissues, rows represent genes. Colour scale shows standard deviation (STD) from the mean expression value for each gene. Dendrograms on the margin show agglomerative hierarchical clustering (Wards linkage, euclidean distances) of genes (right) and tissues (bottom). From Ref. [64].

spectrum of the other one (the acceptor). Illumination at the excitation wavelength of the donor results in the transfer of energy to the acceptor in a non-radiation manner. This energy transfer leads to a decrease in emission from the donor and an increase in fluorescence at the emission wavelength of the acceptor (called sensitised emission). The altered ratio of donor:acceptor fluorescence at the excitation wavelength of the donor can be measured with digital imaging and the signal crosstalk from non-FRET fluorescence can be eliminated by control samples containing just one fluorophore and calculation of correction factors (3-filter method of FRET microscopy; for review see Ref. [90]). As FRET is a non-destructive spectroscopic method, it can be performed in living cells (Fig. 11), but, in principle, it can also be applied to tissue sections of fixed material and the state and interactions of many different molecules can be analysed.

3.2.7. *Surface-enhanced laser desorption/ionisation (SELDI)*

SELDI is an affinity based mass spectometry method in which proteins (<20 kD) are selectively absorbed to a chemically modified surface, and impurities are removed by washing with buffer. The use of several chromatographic arrays and wash conditions enables high-speed, high-resolution chromatographic separations [91]. The technique has been successfully used to

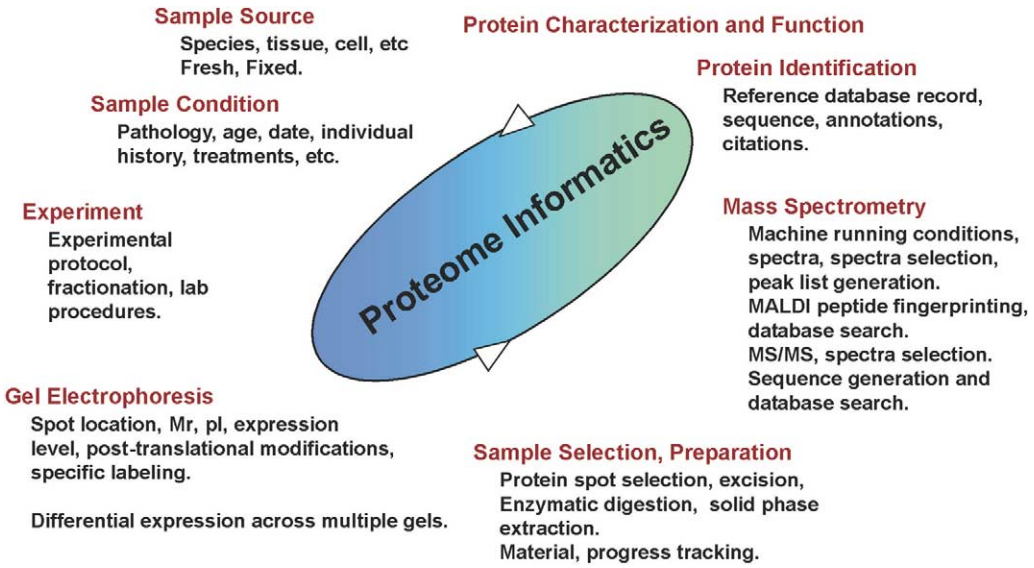

Fig. 9. Summary of variables important for input into different databases, bioinformatics will become even more important for result comparison. MS, mass spectrometry; Mr, molecular weight; PI, isoelectric point; lab, laboratory; MALDI, Matrix Associated Laser Desorption Ionisation.

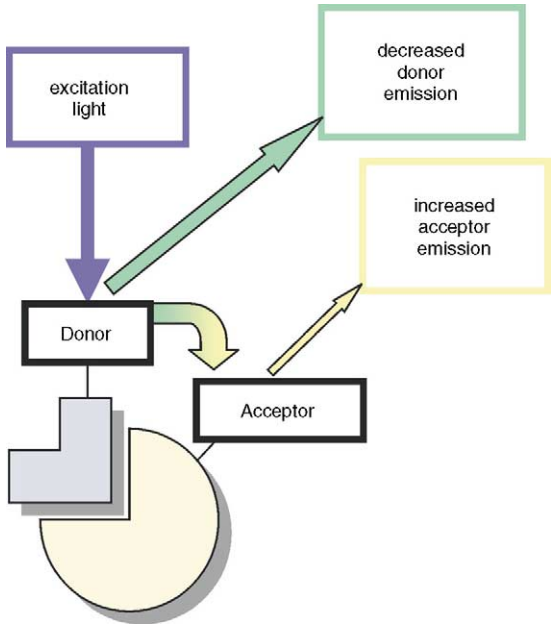

Fig. 10. Schematic illustration of fluorescence resonance energy transfer (FRET): excitation of a donor fluorophore results in energy transfer to an acceptor leading to a decrease in donor emission and an increase in acceptor fluorescence (at the donor excitation wavelength).

detect several disease-associated proteins in complex biological specimens such as cell lysates and serum [92–95].

3.2.8. Tissue microarrays (TMA)

TMAs consist of hundreds of small core (0.6–2.0 mm) tissue sections arrayed on a glass slide. The arrays can be used for Immunohistochemistry (IHC), *In Situ* Hybridisation ((ISH) DNA, RNA) and are useful for the evaluation of new antibodies and large-scale outcome studies. The disadvantage is that evaluation and scoring of all tissue-spots is rather time-consuming, but theoretically this can be automated. It has been argued that the results may not always be very reproducible, but TMAs gave a reliable impression of the whole tumour [96,97]. Staining results can be correlated with e.g. clinical outcome. By combining IHC and ISH in TMAs of 611 breast cancers, increased expression levels of the telomerase core components, hTERT and hTR appeared to be associated with a lower overall survival [98]. Recently, automated analysis of TMAs could discern subtle staining differences better and 30–50 times faster than pathologists [99].

4. Clinical genomic and proteomic applications

Cancer cytogenetics is largely restricted to haematological malignancies and soft-tissue tumours; many can be classified by typifying translocations (Table 2). The discovery of t(9;22) translocation in chronic myelogenous leukaemia (CML), their cell biological function and the characteristics of the BCR-ABL fusion protein, made it possible to develop a very specific inhibitor (STI-571), which has transformed CML therapy [100]. The number of solid tumour types that can be classified by chromosomal aberrations is limited (see Table 3, after Ref. [101]), partly because they have been much less studied. However, two examples are that *HER2/neu* amplification occurs in 15–30% of breast carcinomas and is associated with a poor prognosis, and FISH-screening could be used to select patients with an improved response to trastuzumab (Herceptin) [102]. Combined centromere probes for chromosomes 3, 7 and 17, and a locus-specific probe for *p16* at 9p21, detect

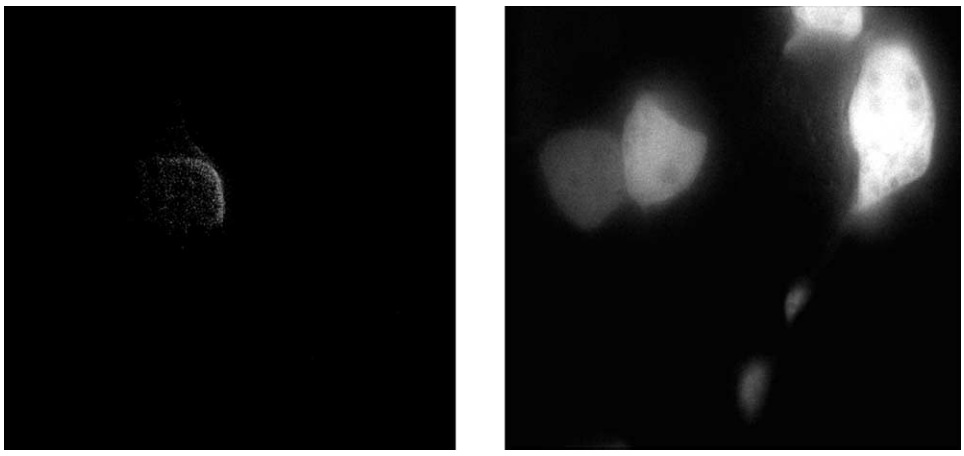

Fig. 11. Example of FRET microscopy using the three-filter method: a corrected FRET image is shown for a negative (left) and a positive (right) control sample. These corrected images were obtained after subtraction of non-FRET fluorescence from the FRET-filter channel (as reviewed in Ref. [90]).

Table 2
Characterising cytogenetic features in haematological tumours

	Chromosome change
Non-Hodgkin's lymphoma	
Lymphoplasmacytoid lymphoma	t(9;14)(p13;q32)
Follicular lymphoma	t(9;14)(p13;q32)
Mucosa-associated lymphoid tissue lymphoma	t(11;18)(q21;q21) t(1:14)(p22;q32)
Mantle cell lymphoma	t(11;14)(q13;q32)
Burkitt's lymphoma	t(8;14)(q24;q32), t(2;8)(p12;q24), t(8;22)(q24;q11)
T-cell anaplastic large cell lymphoma	t(2;5)(p23;q35)
Leukaemia	
Acute myeloid leukaemia (AML)	*MLL* gene on 11q23 with different partners: 4q21, 19p13.3, 9p22, 22q13, 1q21, *TEL* gene on 12p13 with different partners: 5q33, 9q34, 22q11
Acute lymphoblastic leukaemia (ALL)	*C-MYC* on 8q34 with different partners: 14qq32, 2p12, 22q11 *TCR-alpha/delta* on 14q11 with different partners: 8q24, 11p15, 11p13 *TCR-beta* on 7q34 with different partners: 1p32, 9q34, 19p13
Acute promyelocytic leukaemia (APL)	*RAR-alpha* on 17q21 with different partners: 15q21, 11q23
T-cell prolymphocytic leukaemia (T-PLL)	*TCR-alpha/delta* on 14q11 with different partners: 14q32.1, Xq28
Chronic myeloid leukaemia (CML)	t(9;22)(q34;q11) t(9;12)(q34;p13)
Atypical chronic myeloid leukaemia	
Myelodysplastic syndrome (MDS)	t(3;21)(q26;q22), t(12;22)(p13;q11), t(7;11)(p15;p15)

urothelial carcinomas with a higher sensitivity than with cytology, with comparable specificities [103]. These examples illustrate that a better understanding of chromosomal abnormalities in solid tumours can result in improved classification and targeted therapies. In breast cancers, a gene-expression profile array of 70 genes predicted prognosis better than grade, stage, and nodal status [75]. Recently, chromosome 3q gain by CGH predicted disease-specific survival much more accurately than classical and proliferation features [51].

What can we expect from proteomics in the near future? For detection, diagnosing and monitoring cancer, a combination of markers will be necessary. Such fingerprints will require high throughput and proteomic profiling, and sophisticated bio-informatics tools for complex data analysis and pattern recognition. Cluster analysis-identified patterns of protein expression in 2-DE maps, enabled grouping of breast cancer cell lines according to *in vitro* morphology [104]. Li and colleagues [105] used a proteomic and bioinformatic approach

Table 3
Characterising cytogenetic features in a number of solid tumours

	Chromosome change
Bone and soft-tissue tumours	
Liposarcoma, myxoid and round cell	t(12;16)(q13;p11)
Liposarcoma, well differentiated	r(12)
Synovial sarcoma	t(X;18)(p11;q11)
Rhabdomyosarcoma	t(2;13)(q35;q14) t(1;13)(p36;q14)
Extraskeletal myxoid chondrosarcoma	t(9;22)(q22;q12)
Clear cell sarcoma	t(12;22)(q13;q14)
Ewing's sarcoma/Askin's tumour/peripheral neuroepithelioma	t(11;22)(q24;q12) t(21;22)(q22;q12) t(7;22)(p22;q12)
Desmoplastic small round cell tumour	t(11;22)(p13;q12)
Epithelial tumours	
Renal cell adenoma	+7, +17, −Y
Renal cell carcinoma	del(3p), der(3p), t(X;1)(p11;q21)
Bladder carcinoma	−9
Thyroid adenoma	+5, +7, +12
Thyroid papillary carcinoma	inv(10)(q11q21)
Salivary gland:	
Pleomorphic adenoma	Rearrangements of 3p21, 8q12 and 12q15
Adenocarcinoma	del(6q)
Adenoid cystic carcinoma	t(6;9)(q21–23;p13–22)
Mucoepidermoid carcinoma	del(6q), t(11;19)(q14–21;p12–13)

to identify serum biomarkers to detect breast cancer with SELDI-TOF-MS. This led to three discriminatory biomarkers with high sensitivity (93%) and specificity (91%). Another proteoma study detected ovarian cancer [106] very accurately. Application to archival cytology material tried to identify fingerprints for diagnostic purposes [107]. Specific protein patterns could be obtained for melanoma, clear cell sarcoma, renal carcinoma and reactive effusions. Automated TMA [108] could be the ideal method for this. In breast cells with ErbB-2 overexpression, two protein samples with a differentially labelled lysine residue (Cy3 and Cy5) were mixed and analysed on one gel, thereby facilitating the detection and quantitation of differences [109].

Genomic and proteomic data may complement one another, as the following very elegant study shows [110]. Analysing several purified cell membrane preparations from breast cancer cell lines (oestrogen-positive and Epidermal Growth Factor-Receptor positive) with MALDI-TOF and tandem mass spectrometry, 501 distinct proteins could be identified. Of these, 27% were either hypothetical (proteins predicted from the transcriptome) or of unknown function and cell localisation. Three of these unknown/hypothetical 'breast cancer membrane proteins' (BCMP) were analysed further using bioinformatics to obtain the full open reading frame. After cloning and C-terminal tagging with green SuperGlo^{TM} autofluorescent protein (AFP) and immunocytochemistry, the cellular localisation was determined. A Y2HS revealed the different interacting proteins. Antibodies were then raised and used in an immunohistochemical analysis to determine the expression of the different BCMPs in normal and breast cancer tissues. Real-time quantitative reverse transcriptase (RT)-PCR was used to compare expression levels of mRNA and immunohistochemical staining. The end result of this study was a complete description, characterisation, function and localisation of three new BCMPs. The potential of such a large-scale multidimensional approach is enormous and shows that with the expansion of proteomic research, genomic research remains essential, not only for diagnostic, but also potentially for therapeutic targets. However, adequate multicentre reproducibility and validation studies are required before genomics and proteomics tests can be adopted into clinical routine [111].

Will genomics and proteomics replace classical diagnostic or prognostic procedures? The abovementioned gene expression array results in breast cancer suggest this [75]. The same holds for the prognostic chromosome 3q gains [51]. In endometrial hyperplasia, comparison of the diagnoses of four expert gynaeco-pathologists and molecular genetic clonality showed a poor correlation. However, a multivariate function called D-Score with strong prognostic value as to cancer progression was not only more reproducible than the pathologists' classification, but also correlated much more strongly with clonality. Nearly all cases with $DS \geqslant 1$ are polyclonal, whereas many (but not all) cases with $DS < 1$ are monoclonal. Some of the apparent 'polyclonal' cases with $DS < 1$ can be attributed to technical error in clon-

ality assessment, where contaminating polyclonal normal tissues obscured the monoclonal nature of the tested samples. As the measurements required for the D-Score can be done on standard tissue sections, take only 20–25 min per case and are highly cost-effective, it is currently much easier to use the quantitative image analysis prognostic criteria than genetic testing for cancer progression prediction [112].

Thus, it is simply too early to make decisive statements about the future relative clinical significance of classical pathology, quantitative molecular pathology, genomics and proteomics, as validation and reproducibility tests have yet to be carried out. Certainly, genomic and proteomics research will expand rapidly. It seems probable that they will get a central place in the research, understanding, diagnosis, monitoring and treatment of (pre)cancers of many different sites.

Acknowledgements

This work was supported by grant No. 28-2015 of the National Health Care Research Council of The Netherlands, ZonMw and Dutch Cancer Society VU2002-2018.

References

1. Casperson TO. History of the development of cytophotometry from 1935 to the present. *Anal Quant Cytol Histol* 1987, **9**, 2–5.
2. Merlo A, Herman JG, Mao L, *et al*. 5′ CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers. *Nat Med* 1995, **1**, 686–692.
3. Hendrich B, Bird A. In Jones PA, Vogt PK, eds. *Current Topics in Microbiology and Immunology, vol. 249, DNA Methylation and Cancer*. Berlin, Springer-Verlag, 2000, 170.
4. Klochendler-Yeivin A, Yaniv M. Chromatin modifiers and tumour suppression. *Biochim Biophys Acta* 2001, **1551**, M1–10.
5. Baak JPA, Vooijs GP, Brugal G, Burger G. Nuclear image cytometry: quantitation of chromatin pattern, steroid receptor content and Ki-67. In Baale JPA, Ed. *Manual of Quantitative Pathology in Cancer Diagnosis and Prognosis*. Springer-Verlag, Heidelberg, 1991, 232–243.
6. Palcic B, Garner DM, MacAulay CE. Image cytometry and chemoprevention in cervical cancer. *J Cell Biochem* 1995, **23** (Suppl.), 43–54.
7. Mommers EC, Poulin N, Sangulin J, Meijer CJ, Baak JP, van Diest PJ. Nuclear cytometric changes in breast carcinogenesis. *J Pathol* 2001, **193**, 33–39.
8. Offit K, Wong G, Filippa DA, Tao Y, Chaganti RSK. Cytogenetic analysis of 434 consecutively ascertained specimens of non-Hodgkin's lymphoma, clinical correlations. *Blood* 1991, **77**, 1508–1915.
9. Mitelman F, Johansson B, Mandahl N, Mertens F. Clinical significance of cytogenetic findings in solid tumours. *Cancer Genet Cytogenet* 1997, **95**, 1–8.
10. Knudson AG Jr. Mutation and cancer, Statistical study of retinoblastoma. *Proc Natl Acad Sci USA* 1971, **68**, 820–823.
11. Kinzler KW, Vogelstein B. Cancer-susceptibility genes. Gatekeepers and caretakers. *Nature* 1997, **386**, 761–763.
12. Cahill DP, Kinzler KW, Vogelstein B, Lengauer C. Genetic instability and darwinian selection in tumours. *Trends Cell Biol* 1999, **9**, 57–60.
13. Fearon ER, Vogel stein B. A genetic model for colorectal tumourigenesis. *Cell* 1990, **61**, 759–767.
14. Forastiere A, Koch W, Trotti A, Sidransky D. Head and neck cancer. *N Engl J Med* 2001, **345**, 1890–1900.
15. Schwarz E, Freese UK, Gissmann L, *et al*. Structure and transcription of human papillomavirus sequences in cervical carcinoma cells. *Nature* 1985, **314**, 111–114.
16. Sakurazawa N, Tanaka N, Onda M, Esumi H. Instability of X chromosome methylation in aberrant crypt foci of the human colon. *Cancer Res* 2000, **60**, 3165–3169.
17. Zarbl H, Sukumar S, Arthur AV, Martin-Zanca D, Barbacid M. Direct mutagenesis of Ha-ras-1 oncogenes by *N*-nitroso-*N*-ethylurea during initiation of mammary carcinogenesis in rats. *Nature* 1985, **315**, 382–385.
18. Stoler DL, Chen N, Basik M, *et al*. The onset and extent of genomic instability in sporadic colorectal tumour progression. *Proc Natl Acad Sci USA* 1999, **96**, 15121–15126.
19. Rowley JD. Chromosome translocations, dangerous liaisons revisited. *Nature Rev Cancer* 2001, **1**, 245–250.
20. Jallepalli PV, Lengauer C. Chromosome segregation and cancer, cutting through the mystery. *Nature Rev Cancer* 2001, **1**, 109–117.
21. Kallioniemi A, Kallioniemi OP, Sudar D, *et al*. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumours. *Science* 1992, **258**, 818–821.
22. Ried T, Knutzen R, Steinbeck R, *et al*. Comparative genomic hybridization reveals a specific pattern of chromosomal gains and losses during the genesis of colorectal tumours. *Genes, Chromosomes Cancer* 1996, **15**, 234–245.
23. Meijer GA, Hermsen MAJA, Baak JPA, *et al*. Progression from colorectal adenoma to carcinoma is associated with non-random chromosomal gains as detected by CGH. *J Clin Pathol* 1998, **51**, 901–909.
24. Bardi G, Sukhikh T, Pandis N, Fenger C, Kronborg O, Heim S. Karyotypic characterization of colorectal adenocarcinomas. *Genes Chromosomes Cancer* 1995, **12**, 97–109.
25. Shih IM, Zhou W, Goodman SN, Lengauer C, Kinzler KW, Vogelstein B. Evidence that genetic instability occurs at an early stage of colorectal tumourigenesis. *Cancer Res* 2001, **61**, 818–822.
26. Ried T, Heselmeyer-Haddad K, Blegen H, Schrock E, Auer G. Genomic changes defining the genesis, progression, and malignancy potential in solid human tumours, a phenotype/genotype correlation. *Genes Chromosomes Cancer* 1999, **25**, 195–204.
27. Redon R, Muller D, Caulee K, Wanherdrick K, Abecassis J, du Manoir S. A simple specific pattern of chromosomal aberrations at early stages of head and neck squamous cell carcinomas, PIK3CA but not p63 gene as a likely target of 3q26-qter gains. *Cancer Res* 2001, **61**, 4122–4129.
28. Ellis NA, German J. Molecular genetics of Bloom's syndrome. *Hum Mol Genet* 1996, **5**, 1457–1463.
29. Bootsma D, Kraemer KH, Cleaver JE, Hoeijmakers JHJ. Nucleotide excision repair syndromes, Xeroderma pigmentosa, Cockayne Syndrome, and trichothiodystrophy. In Vogelstein B, Kinzler KW, eds. *The Molecular Basis of Cancer*. New York, McGraw-Hill, 1998, 245–274.
30. Joenje H, Patel KJ. The emerging genetic and molecular basis of Fanconi Anaemia. *Nature* 2001, **2**, 446–457.
31. Marra G, Boland CR. Hereditary nonpolyposis colorectal cancer, the syndrome, the genes, and historical perspectives. *J Natl Cancer Inst* 1995, **87**, 1114–1125.
32. Newsham IF, Hadjistilianou T, Cavenee WK. Retinoblastoma.

In Vogelstein B, Kinzler KW, eds. *The Molecular Basis of Cancer*. New York, McGraw-Hill, 1998, 367–392.
33. Malkin D. p53 and the Li-Fraumeni syndrome. *Biochim Biophys Acta* 1994, **1198**, 197–213.
34. Kamb A, Herlyn M. Malignant melanoma. In Vogelstein B, Kinzler KW, eds. *The Molecular Basis of Cancer*. New York, McGraw-Hill, 1998, 507–518.
35. Couch FJ, Weber BL. Breast cancer. In Vogelstein B, Kinzler KW, eds. *The Molecular Basis of Cancer*. New York, McGraw-Hill, 1998, 537–564.
36. Felsher DW, Bishop JM. Transient excess of MYC activity can elicit genomic instability and tumorigenesis. *Proc Natl Acad Sci USA* 1999, **96**, 3940–3944.
37. Duesberg P, Rausch C, Rasnick D, Hehlmann R. Genetic instability of cancer cells is proportional to their degree of aneuploidy. *Proc Natl Acad Sci USA* 1998, **95**, 13692–13697.
38. Fodde R, Kuipers J, Rosenberg C, *et al*. Mutations in the APC tumour suppressor gene cause chromosomal instability. *Nat Cell Biol* 2001, **3**, 433–438.
39. Kaplan KB, Burds AA, Swedlow JR, Bekir SS, Sorger PK, Nathke IS. A role for the adenomatous polyposis coli protein in chromosome segregation. *Nat Cell Biol* 2001, **3**, 429–432.
40. Abdel-Rahman WM, Katsura K, Rens W, *et al*. Spectral karyotyping suggests additional subsets of colorectal cancers characterized by pattern of chromosome rearrangement. *Proc Natl Acad Sci USA* 2001, **98**, 2538–2543.
41. Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. *Nature* 1998, **396**, 643–649.
42. Amon A. The spindle checkpoint. *Curr Opin Genet Dev* 1999, **9**, 69–75.
43. Nigg EA. Mitotic kinases as regulators of cell division and its checkpoints. *Nat Rev Mol Cell Biol* 2001, **2**, 21–32.
44. Nagahata T, Hirano A, Utada Y, *et al*. Correlation of allelic losses and clinicopathological factors in 504 primary breast cancers. *Breast Cancer* 2002, **9**, 208–215.
45. Murata H, Khattar NH, Kang Y, Gu L, Li GM. Genetic and epigenetic modification of mismatch repair genes hMSH2 and hMLH1 in sporadic breast cancer with microsatellite instability. *Oncogene* 2002, **21**, 5696–5703.
46. Hansemann D. Uber asymmetrische Zellteilung in Epithelkrebsen und deren biologischen bedeutung. *Virchow's Arch Path Anat* 1890, **119**, 299–326.
47. Tjio JH, Levan A. The chromosome number of man. *Hereditas* 1956, **42**, 1–6.
48. Rowley JD. A new consistent chromosomal abnormality in chronic myelogenous leukemia identified by quinacrine fluorescence and Giemsa staining. *Nature* 1973, **243**, 290–293.
49. Waters JJ, Barlow AL, Gould CP. Demystified ... FISH. *Mol Pathol* 1998, **51**, 62–70.
50. Hermsen MAJA, Baak JPA, Weiss J, *et al*. Genetic analysis of 53 lymph node-negative breast carcinomas by CGH and relation to clinical, pathological, morphometric, and DNA cytometric prognostic factors. *J Pathol* 1998, **186**, 356–362.
51. Janssen EAM, Baak JPA, Guervós MA, Diest PJ van, Jiwa M, Hermsen MAJA. Lymph node-negative invasive breast carcinomas, specific chromosomal aberrations are strongly associated with high mitotic activity and predict outcome more accurately than grade, tumour diameter and oestrogen receptor. *J Pathol* (in press).
52. Solinas-Toldo S, Lampel S, Stilgenbauer S, *et al*. Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosomes Cancer* 1997, **20**, 399–407.
53. Pinkel D, Segraves R, Sudar D, *et al*. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet* 1998, **20**, 207–211.
54. Weiss MM, Snijders AM, Meijer GA, *et al*. Microarray CGH reveals a narrow gain at 20q13 in human gastric carcinomas. *Gastroenterology* 2001, **120**, A1541.
55. Snijders AM, Nowak N, Segraves R, *et al*. Assembly of microarrays for genome-wide measurement of DNA copy number. *Nat Genet* 2001, **29**, 263–264.
56. Hyman E, Kauraniemi P, Hautaniemi S, *et al*. Impact of DNA amplification on gene expression patterns in breast cancer. *Cancer Res* 2002, **62**, 6240–6245.
57. Schröck E, du Manoir S, Veldman T, *et al*. Multicolor spectral karyotyping of human chromosomes. *Science* 1996, **273**, 494–497.
58. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* 2003, **31**, E15.
59. Coller HA, Grandori C, Tamayo P, *et al*. Expression analysis with oligonucleotide microarrays reveals that MYC regulates genes involved in growth, cell cycle, signalling, and adhesion. *Proc Natl Acad Sci USA* 2000, **97**, 3260–3265.
60. Golub TR, Slonim DK, Tamayo P, *et al*. Molecular classification of cancer, class discovery and class prediction by gene expression monitoring. *Science* 1999, **286**, 531–537.
61. Su YA, Bittner ML, Chen Y, *et al*. Identification of tumor-suppressor genes using human melanoma cell lines UACC903, UACC903(+6), and SRS3 by comparison of expression profiles. *Mol Carcinog* 2000, **28**, 119–127.
62. Maniotis AJ, Folberg R, Hess A, *et al*. Vascular channel formation by human melanoma cells in vivo and in vitro, vasculogenic mimicry [see comments]. *Am J Pathol* 1999, **155**, 739–752.
63. Brasier AR. Retriever and CompareTable, two informatics tools for data analysis of high-density oligonucleotide arrays. *Biotechniques* 2002, **32**, 100–109.
64. Mutter GL, Baak JPA, Fitzgerald J, *et al*. Global expression changes of constitutive and hormonally regulated genes during endometrial neoplastic transformation. *Gynecol Oncol* 2001, **83**, 177–185.
65. Pruitt KD, Katz KS, Sicotte H, Maglott DR. Introducing RefSeq and LocusLink, curated human genome resources at the NCBI. *Trends Genet* 2000, **16**, 44–47.
66. Wheeler DL, Chappey C, Lash AE, *et al*. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 2000, **28**, 10–14.
67. National Center for Biotechnology Information (USA). Locuslink [online]. Available: http,//www ncbi nlm nih gov/LocusLink/2000.
68. Mutter GL. Endometrial precancer type collection [online]. Available: http,//www endometrium org 2000.
69. Perou CM, Sørlie T, Eisen MB, *et al*. Molecular protraits of human breast tumours. *Nature* 2000, **406**, 747–752.
70. Hedenfalk I, Duggan D, Chen Y, *et al*. Gene-expression profiles in hereditary breast cancer. *N Eng J Med* 2001, **344**, 539–548.
71. van 't Veer LJ, Dai H, van de Vijver MJ, *et al*. Expression profiling predicts outcome in breast cancer. *Breast Cancer Res* 2002, **5**, 57–58.
72. Gruvberger S, Ringner M, Chen Y, *et al*. Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. *Cancer Research* 2001, **61**, 5979–5984.
73. West M, Blanchette C, Dressman H, *et al*. Predicting the clinical status of human breast cancer by using gene expression profiles. *PNAS* 2001, **98**, 11462–11467.
74. Ahr A, Kam T, Solbach S, *et al*. Identification of high risk breast cancer by gene expression profiling. *Lancet* 2002, **359**, 131–132.
75. van de Vijver MJ, He YD, van't Veer LJ, *et al*. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002, **347**, 1999–2009.
76. Wilkins MR, Sanchez JC, Gooley AA, *et al*. Progress with proteome projects. Why all proteins expressed by a genome should

be identified and how to do it. *Biotechnol Genet Eng Rev* 1996, **13**, 19–50.
77. Wilkins MR, Sanchez JC, Williams KL, Hochstrasser DF. Current challenges and future applications for protein maps and post-translational vector maps in proteome projects. *Electrophoresis* 1996, **17**, 830–838.
78. Reynolds T. For proteomics research, a new race has begun. *J Natl Cancer Inst* 2002, **94**, 552–554.
79. Anderson NG, Anderson L. The Human Protein Index. *Clin Chem* 1982, **28**, 739–748.
80. Paweletz CP, Liotta LA, Petricoin 3rd EF. New technologies for biomarker analysis of prostate progression, laser capture microdissection and tissue proteomics. *Urology* 2001, **57**, 160–163.
81. Czerwenka KF, Manavi M, Hosmann J, *et al*. Comparative analysis of two-dimensional protein patterns in malignant and normal human breast tissue. *Cancer Detect Prev* 2001, **25**, 268–279.
82. Kuerer HM, Golgknopf IL, Fritsche H, Krishnamurthy S, Sheta EA, Hunt KK. Identification of distinct protein expression patterns in bilateral matched pair breast ductal fluid specimens from women with unilateral invasive breast carcinoma, High-throughput biomarker discovery. *Cancer* 2002, **95**, 2276–2282.
83. Fields S, Song O-K. A novel genetic system to detect protein-protein interactions. *Nature* 1989, **340**, 245–246.
84. Coates PJ, Hall PA. The yeast two hybrid system for identifying protein-protein interactions. *J Path* 2003, **199**, 4–7.
85. Gavin AC, Bosche M, Krause R, *et al*. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 2002, **15**, 141–147.
86. Schmidt MHH, Broll R, Bruch HP, Bõlger O, Duchrow M. The proliferation marker pKI-67 organizes the nucleolus during the cell cycle depending on Ran and cyclin B. *J Path* 2003, **199**, 18–27.
87. Schweitzer B. Multiplexed protein profiling on microarrays by rolling-circle amplification. *Nat Biotechnol* 2002, **20**, 359–365.
88. Haab BB, Dunham MJ, Brown PO. Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. *Genome Biol* 2001, **2**, 0004.1–0004.13.
89. Paweletz CP. Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front. *Oncogene* 2001, **20**, 1981–1989.
90. Schmid JA, Sitte HH. Fluorescence resonance energy transfer in the study of cancer pathways. *Curr Opin Oncol* 2003, **15**, 55–64.
91. Merchant M, Weinberger SR. Recent advancements in surface-enhancement laser desorption/ionisation-time of flight-mass spectrometry. *Electrophoresis* 2000, **21**, 1164–1167.
92. Wright Jr. GL, Cazares LH, Leung S-M, *et al*. Proteinchip® surface enhanced laser desorption/ionisation (SELDI) mass spectrometry, a novel protein biochip technology for detection of prostate cancer biomarkers in complex protein mixtures. *Prostate Cancer Prostate Dis* 1999, **2**, 264–276.
93. Vlahou A, Schellhammer PF, Medrinos S, *et al*. Development of a novel proteomic approach for the detection of transitional cell carcinoma of the bladder in urine. *Am J Pathol* 2001, **158**, 1491–1502.
94. Li J, Zhang Z, Rosenzweig J, Wang YY, Chan DW. Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. *Clin Chem* 2002, **48**, 1296–1304.
95. Petricoin III PF, Ardekani AM, Hitt BA, *et al*. Use of proteomic patterns in serum to identify ovarian cancer. *The Lancet* 2002, **359**, 572–577.
96. Torhorst J, Bucher C, Kononen J, *et al*. Tissue microarrays for rapid linking of molecular changes to clinical endpoints. *Am J Path* 2001, **159**, 2249–2256.
97. Zhang D, Salto-Tellez M, Choudary Putti T, Do E. Reliability of tissue microarrays in detecting protein expression and gene amplification in breast cancer. *Mod Pathol* 2003, **16**, 79–85.
98. Poremba C, Heine B, Daillo R, *et al*. Telomerase as a prognostic marker in breast cancer, high-throughput tissue microarray analysis of hTERT and hTR. *J Path* 2002, **198**, 181–189.
99. Camp RL, Chung GG, Rimm DL. Automated subcellular localization and quantification of protein expression in tissue microarrays. *Nature Medicine* 2002, **11**, 1323–1327.
100. Druker BJ. STI571 (Gleevec) as a paradigm for cancer therapy. *Trends Mol Med* 2002, **8**, S14–AS8.
101. Mitelman F, Mertens F, Johansson B. A breakpoint map of recurrent chromosomal rearrangements in human neoplasia [special issue]. *Nat Genet* 1997, 417–472.
102. Funkhouser WK, Kaiser-Rogers K. Review, Significance of, and optimal screening for, Her2/neu gene amplification and protein overexpression in breast cacinoma. *Ann Clin Lab Sci* 2001, **31**, 349–358.
103. Halling KC, King W, Sokolowa IA, *et al*. A comparison of cytology and fluorescence in situ hybridization for the detection of urothelial carcinoma. *J Urol* 2000, **164**, 1768–1775.
104. Harris RA, Yang A, Stein RC, *et al*. Cluster analysis of an extensive human breast cell line protein expression map database. *Proteomics* 2002, **2**, 212–223.
105. Li J, Zhang Z, Rosenzweig J, Wang YY, Chan DW. Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. *Clin Chem* 2002, **48**, 1296–1304.
106. Petricoin III PF, Ardekani AM, Hitt BA, *et al*. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 2002, **359**, 572–577.
107. Fetsch PA, Simone NL, Brynat-Greenwood PK, *et al*. Proteomic evaluation of archival cytologic material using SELDI affinity mass spectrometry. *Am J Clin Path* 2002, **118**, 870–876.
108. Camp RL, Chung GG, Rimm DL. Automated subcellular localization and quantification of protein expression in tissue microarrays. *Nat Med* 2002, **11**, 1323–1327.
109. Gharbi S, Gaffney P, Yang A, *et al*. Evaluation of two-dimensional differential gel electrophoresis for proteomic expression analysis of a model breast cancer cell system. *Mol Cell Prot* 2002, **1**, 91–98.
110. Adam PJ, Boyd R, Tyson KL, *et al*. Comprehensive proteomic analysis of breast cancer cell membrane reveals unique proteins with potential roles in clinical cancer. *J Biol Chem* 2003, **278**, 6482–6489.
111. Baak JPA. Perspectives Editorial: the framework of pathology, good laboratory practice by quantitative and molecular methods. *J Pathol* 2002, **198**, 277–283.
112. Baak JPA, Ørbo A, Diest PJ, *et al*. Prospective multicenter evaluation of the morphometric D-Score for prediction of the outcome of endometrial hyperplasias. *Am J Surg Path* 2001, **25**, 930–935.

Further reading

Internet addresses

Peptide spectra can be used to search protein sequence databases:
http://www.ncbi.nlm.nih.gov/
http://www.expasy.org
http://www.proteomics.cancer.dk/
http://www.hgmp.mrc.ac.uk/GenomeWeb/prot-interaction.html
http://www.bioinf.man.ac.uk/dbbrowser/OWL/