Gene Expression in Inherited Breast Cancer

Ingrid A. Hedenfalk,*,† Markus Ringnér,* Jeffrey M. Trent,* and Åke Borg†

**Cancer Genetics Branch*
National Human Genome Research Institute
National Institutes of Health
Bethesda, Maryland 20892
†Department of Oncology
Lund University
SE-221 85 Lund, Sweden

Large proportions of hereditary breast cancers are due to mutations in the two breast cancer susceptibility genes *BRCA1* and *BRCA2*. Considerable effort has gone into studying the function(s) of these tumor suppressor genes, both in attempts to better understand why individuals with these inherited mutations acquire breast (and ovarian) cancer and

Advances in CANCER RESEARCH
0065-230X/02 $35.00

to potentially develop better treatment strategies. The advent of tools such as cDNA microarrays has enabled researchers to study global gene expression patterns in, for example, primary tumors, thus providing more comprehensive overviews of tumor development and progression. Our recent study (Hedenfalk *et al.,* 2001) strongly supports the principle that genomic approaches to classification of hereditary breast cancers are possible, and that further studies will likely identify the most significant genes that discriminate between subgroups and may influence prognosis and treatment. A large number of hereditary breast cancer cases cannot be accounted for by mutations in these two genes and are believed to be due to as yet unidentified breast cancer predisposition genes (*BRCAx*). Subclassification of these non-*BRCA1*/2 breast cancers using cDNA microarray-based gene expression profiling, followed by linkage analysis and/or investigation of genomic alterations, may help in the recognition of novel breast cancer predisposition loci. To summarize, gene expression-based analysis of hereditary breast cancer can potentially be used for classification purposes, as well as to expand upon our knowledge of differences between different forms of hereditary breast cancer. Initial studies indicate that a patient's genotype does in fact leave an identifiable trace on her/his cancer's gene expression profile.

I. INTRODUCTION

Breast cancer is one of the most common malignancies affecting women in the Western world today, the lifetime risk being approximately 10% (Casey, 1997). Breast cancer is both genetically and histopathologically heterogeneous, and the mechanism(s) underlying breast cancer development remains largely unknown. Approximately 5–10% of breast cancers are of hereditary origin, and two major breast cancer susceptibility genes have been identified to date, *BRCA1* and *BRCA2*. These two genes were initially proposed to be responsible for the majority of inherited breast cancer (Easton, 1999; Miki *et al.,* 1994; Wooster *et al.,* 1995), but more recent population-based studies suggest that they account for a far smaller portion of familial breast cancer, with considerable variation between different populations (Szabo and King, 1997). Presumably, additional *BRCA* genes with high-penetrance alleles may exist (Kainu *et al.,* 2000), but familial breast cancer may also be due to low-penetrance or recessively acting susceptibility alleles (Cui *et al.,* 2001).

Large-scale gene expression studies using microarrays have come to play an important role in our effort to better understand diseases such as cancer. First, microarrays can be used to subclassify tumors into homogeneous entities based on gene expression profiles. Second, genome-wide expression data can help us further characterize the biology of these "new" subgroups. Finally, microarray experiments can aid in the search for new therapeutic targets and in the identification of novel diagnostic markers.

II. EPIDEMIOLOGY OF FAMILIAL BREAST CANCER

Germline mutations in *BRCA1* have been identified in 15–45% of women with a strong family history of breast cancer and in 60–80% of women with a family history of both breast and ovarian cancer (Couch *et al.,* 1997; Narod *et al.,* 1995b; Peto *et al.,* 1999). The lifetime risk for breast cancer is 60–80% in females carrying a *BRCA1* mutation, although penetrance estimates vary depending on the study population (Easton *et al.,* 1993; Struewing *et al.,* 1996). Lifetime ovarian cancer risks are approximately 20–40% (Easton *et al.,* 1995; Struewing *et al.,* 1997) in *BRCA1* mutation carriers, and, to a much lesser extent, males have an increased risk of prostate cancer (Ford *et al.,* 1994). A correlation between early onset of disease and bilateral breast cancer, and family history has been shown. The lifetime breast cancer risk for *BRCA2* mutation carriers is estimated to be 60–85%, and the lifetime ovarian cancer risk is approximately 10–20% (Easton *et al.,* 1997; Ford *et al.,* 1998). Male *BRCA2* mutation carriers are also at increased risk of developing breast cancer, with a lifetime risk of 6% (Breast Cancer Linkage Consortium (BCLC), 1999). Moreover, *BRCA2* mutations may also be associated with an increased risk for prostate, pancreas, colon, gall bladder, bile duct, and stomach cancers, as well as malignant melanoma (BCLC, 1999).

III. THE BRCA1 AND BRCA2 GENES

In 1990 Hall and colleagues reported linkage of early-onset breast cancer families to chromosome 17q12 (Hall *et al.,* 1990). In 1994, *BRCA1* was cloned by Miki and colleagues, as they constructed a transcriptional map of a 600 kb region at 17q12, finding mutations that segregated with 17q-linked susceptibility for breast and ovarian cancer (Miki *et al.,* 1994). The search for breast cancer susceptibility genes continued, as it was realized that only approximately 45% of families with multiple cases of early-onset breast cancer showed evidence of linkage to *BRCA1*. In 1994 Wooster and colleagues performed genetic linkage analysis on families with multiple cases of early-onset breast cancer, but without evidence of linkage to *BRCA1* (Wooster *et al.,* 1994). Cosegregation of disease with chromosome 13q markers was found, and in 1995 *BRCA2* was identified (Wooster *et al.,* 1995). The *BRCA1* and *BRCA2* genes are thought to account for the majority of breast and ovarian cancer families (Narod *et al.,* 1995a,b). Nonetheless, despite the considerable variation in the contribution to breast cancer from *BRCA1* and *BRCA2* in different populations, it remains evident that additional breast cancer susceptibility genes are still to be identified.

A. Gene Structures and Mutation Spectra

BRCA1 is a large gene spread over 80 kb of genomic DNA composed of 22 coding exons that are transcribed into a 7.8-kb mRNA encoding a protein containing 1863 amino acids (Miki *et al.,* 1994; Smith *et al.,* 1996) (Fig. 1; see color insert). The approximate molecular mass of the BRCA1 protein is 220 kDa. The *BRCA1* gene bears no homology with other genes, with the exception of a RING finger motif at the amino-terminal end. In other proteins, such a motif has been shown to interact with nucleic acids and to form protein complexes, suggesting a role for *BRCA1* in transcription. In addition, there is a nuclear localization sequence (NLS) in exon 11, and a conserved acidic carboxy terminus, the BRCT (BRCA1 carboxyl-terminal) domain. To date, more than 600 different mutations in the *BRCA1* gene have been reported (Breast Cancer Information Core: http://www.nhgri.nih.gov/Intramural_research/Lab_transfer/Bic/). The majority of these are frameshift or nonsense mutations located throughout the gene and typically result in premature translation termination; in the most subtle form merely the last 11 residues of the protein are lost (Breast Cancer Information Core, see above). Many tumor-associated point mutations in *BRCA1* are found in conserved domains such as the RING finger and the BRCT domain (Irminger-Finger *et al.,* 1999). However, the majority of missense variants reported cannot readily be distinguished as either disease-associated mutations or benign polymorphisms, posing a very relevant problem in genetic counseling. Several *BRCA1* founder mutations have been identified; the two most common are 185delAG and 5382insC, which account for approximately 10% of all the mutations seen in *BRCA1* (Couch and Weber, 1996). Nevertheless, mutations span the whole *BRCA1* gene, a large proportion of which appear in exon 11, which constitutes 60% of the gene. Because of this fact and the size of the gene, mutation screening is both time-consuming and laborious. Traditionally, mutation screening entails the use of the protein truncation test (PTT), single-strand conformational polymorphism (SSCP) analysis, and direct sequencing of the coding region to pinpoint the mutation. One shortcoming of this traditional approach is that large rearrangements may not be detected, and it has been suggested that as many as 30% of mutations in the *BRCA1* gene are undetected by standard mutation detection methods (Unger *et al.,* 2000). Consequently, a proportion of the families that initially test negative for *BRCA1* and *BRCA2* mutations may do so because of cryptic mutations in these genes that are not detectable by conventional PCR-based methods.

BRCA2, like *BRCA1,* is a large gene, consisting of 27 exons that encode a transcript of approximately 12 kb, contained within 70 kb of genomic sequence (Fig. 1). The BRCA2 protein consists of 3418 amino acids, with an estimated molecular mass of 384 kDa (Wooster *et al.,* 1995). Also in common

with BRCA1, BRCA2 shows no homology to other known proteins and contains no previously defined functional domains. There are eight copies of a 30- to 80-amino-acid repeat (BRC repeats) that are present within exon 11 of the gene (Bignell *et al.,* 1997). *BRCA2* mutations span the whole coding region of the gene, and most of these mutations cause premature protein truncation, leading to loss of protein function (Tavtigian *et al.,* 1996). To date, more than 250 mutations have been found (Breast Cancer Information Core, see above). No mutation hotspots have been identified so far.

B. Normal Regulation and Expression of BRCA1 and BRCA2

Both *BRCA1* and *BRCA2* are ubiquitously expressed, with the highest levels found in thymus and testis (Miki *et al.,* 1994). *BRCA1* and *BRCA2* are required for proliferation in early embryogenesis and are up-regulated with the proliferation of breast epithelial cells during puberty and pregnancy (Rajan *et al.,* 1997). Estrogen levels are high in both breast and ovarian tissue during these phases, suggesting that estrogen might stimulate this expression. In normal cells BRCA1 and BRCA2 are nuclear proteins (Bertwistle *et al.,* 1997; Scully *et al.,* 1996). The expression of *BRCA1* and *BRCA2* increases in late G_1 phase of the cell cycle (Gudas *et al.,* 1996; Wang *et al.,* 1997). In mitotic cells, BRCA1, BRCA2, and RAD51 interact and colocalize in a punctate pattern in the nucleus during the S phase of the cell cycle (Chen *et al.,* 1998; Scully *et al.,* 1997). BRCA1 function is regulated by phosphorylation; it is hyperphosphorylated during late G_1 and S phase, and dephosphorylated in M phase (Ruffner *et al.,* 1999). *Id4* (inhibitor of DNA binding 4) has been shown to negatively regulate *BRCA1* (Beger *et al.,* 2001). Overexpression of *Id4* and concomitant reduction of *BRCA1* expression are associated with anchorage-independent growth. Interestingly, estrogen reduces *Id4* expression, hence increasing the expression of *BRCA1*. Conversely, estrogen receptor (ER) negative cells may overexpress *Id4,* with consequent reduction in *BRCA1* expression. In addition, NF-κB has been shown to up-regulate the expression of *BRCA2* by binding the *BRCA2* promoter (Wu *et al.,* 2000). Studies of mammalian cells deficient in *BRCA1* have suggested that it is involved in DNA double-strand break repair, transcription-coupled repair, and cell cycle control, all of which are important for maintaining genomic stability (reviewed by Deng and Scott, 2000).

C. Functions of the BRCA1 and BRCA2 Genes

Both *BRCA1* and *BRCA2* encode large, multifunctional proteins, and both function as tumor suppressor genes. BRCA1 and BRCA2 proteins are

thought to be involved in two main fundamental cellular processes—DNA damage repair and transcriptional regulation (reviewed by Monteiro, 2000; Scully and Livingston, 2000; Wang *et al.,* 2000; Welcsh *et al.,* 2000; Zheng *et al.,* 2000a). In addition, chromatin remodeling functions have been attributed to both BRCA1 and BRCA2 (reviewed by Irminger-Finger *et al.,* 1999; Welcsh *et al.,* 2000). A schematic overview of the roles of BRCA1 and BRCA2 is provided in Fig. 2 (see color insert).

1. TRANSCRIPTIONAL REGULATION AND CHROMATIN REMODELING

BRCA1 contains several functional domains that interact directly or indirectly with a variety of molecules, including tumor suppressors, oncogenes, DNA damage repair proteins, cell cycle regulators, transcriptional activators, and repressors. In support of a role for BRCA1 in transcription, the C-terminal domain of BRCA1 (BRCT) interacts with RNA polymerase II subunits hRPB2 and hRPB10α, as well as with components of the RNA polymerase holoenzyme, including RNA helicase A, CBP/p300, and the BRG1 subunit of SWI/SNF (Bochar *et al.,* 2000; Irminger-Finger *et al.,* 1999; Pao *et al.,* 2000). More N-terminal domains of BRCA1 interact with sequence specific transcription factors, such as p53, c-myc, STAT1, ERα, cAMP-dependent transcription factor-1 (ATF1), and a zinc finger/KRAB-domain protein, ZBRK1 (Fan *et al.,* 2001; Houvras *et al.,* 2000; Monteiro, 2000; Scully and Livingston, 2000; Welcsh *et al.,* 2000; Zheng *et al.,* 2000b). Thus, BRCA1 may serve as a coactivator and bridging factor to RNA polymerase II holoenzyme, thereby altering the expression of target genes, such as *p21* and *GADD45,* involved in cell cycle arrest. Via its BRCT domains, BRCA1 also binds to transcriptional repressor proteins CtIP/CtBP, pRB, and the histone deacetylases RbAp46, RbAp48, HDAC1, and HDAC2 (Chen *et al.,* 2001; Deng and Brodie, 2000; Yarden and Brody, 1999; Yu and Baer, 2000). Taken together, these data suggest that BRCA1 may have either a positive or a negative modulator effect on transcription, depending on the context. Target genes with a ZBRK1 motif, potentially being repressed by BRCA1, include Bax, TopoIIa, and TIMP-1/2 (Zheng *et al.,* 2000b). An additional putative transcriptional activation domain, AD1, containing a coiled-coil motif, has been mapped to a region close to the BRCT domains of BRCA1 (Hu *et al.,* 2000). *In vitro* transactivation assays suggest a role for an N-terminal region of BRCA2 in transcriptional regulation, and another N-terminal region of BRCA2 interacts with P/CAF (Scully and Livingston, 2000; Welcsh *et al.,* 2000). BRCA2 may regulate transcription through recruitment of the histone-acetyltransferase activity of the P/CAF coactivator. Cancer-predisposing mutations in *BRCA1* and *BRCA2* that abolish

transcriptional activation also prevent chromatin remodeling, presumably with a direct effect on DNA replication and repair processes.

2. DNA REPAIR

BRCA1 binds to BRCA2, p53, RAD51, and many other proteins involved in cell cycling and DNA damage response (Scully *et al.,* 1997; Scully and Livingston, 2000). BRCA1 becomes phosphorylated at critical serine/threonine residues by the ATM (ataxia telangiectasia mutated) and ATM-related kinase ATR proteins in response to DNA damage (Cortez *et al.,* 1999). BRCA1 phosphorylation in response to double-strand breaks induced by ionizing radiation (IR) may also be controlled by ATM via CHK2 (checkpoint kinase 2) (Lee *et al.,* 2000). ATM also phosphorylates CtIP, which dissociates from BRCA1, allowing activation of GADD45 to take place (Li *et al.,* 2000; Scully and Livingston, 2000; Welcsh *et al.,* 2000). The activity of ATR is ATM-independent and is also induced by DNA damage caused by UV light and hydroxyurea, suggesting that ATM- and ATR-induced BRCA1 activation is triggered by different types of DNA damage (Scully and Livingston, 2000). In undamaged cells BRCA1 and BRCA2 colocalize with RAD51 in nuclear foci during the S and G_2 phases of the cell cycle. The interaction between RAD51 and BRCA2 is directly mediated by six of the eight central BRC repeats of BRCA2, whereas the RAD51–BRCA1 interaction may be more indirect and possibly is mediated by BRCA2. RAD51 has a known role in double-strand break (DSB) repair in both pro- and eukaryotic cells by promoting joint molecule formation and strand exchange between homologous duplex DNA. RAD51 also forms the synaptonemal complexes during meiotic homologous recombination (HR). BRCA1 and BRCA2 have been shown to participate in both of these processes (Monteiro, 2000; Scully and Livingston, 2000; Welcsh *et al.,* 2000), indicating involvement in recombination-mediated repair of double-stranded breaks and the maintenance of chromosome integrity (Chen *et al.,* 1999).

Upon DNA damage hyperphosphorylated BRCA1 relocalizes, together with RAD51, BRCA2, and additional components, to sites of DNA synthesis (as shown by PCNA staining), and presumably at stalled replication forks. BRCA1 is also part of the RAD50–MRE11–NBS1/p95 complex, an essential component of recombination-mediated repair of DNA double-strand breaks (Zhong *et al.,* 1999). NBS1/p95 is phosphorylated by ATM in response to IR, and the RAD50-MRE11-NBS1/p95 complex is responsible for the end processing of DNA-DSBs that precedes both nonhomologous end joining (NHEJ) and HR. It is possible that BRCA1 may couple the RAD50–MRE11–NBS1/p95 associated end processing and Rad51-induced strand exchange during HR. It has been shown that BRCA1 binds directly to DNA,

thereby inhibiting the nucleolytic activity of the RAD50–MRE11–NBS1/p95 complex, an enzyme implicated in numerous aspects of double-strand break repair (Paull *et al.,* 2001). BRCA1 may also function as a coordinator of an even larger BRCA1-associated genome surveillance complex (BASC), including additional tumor suppressor and DNA repair proteins such as MSH2, MSH6, MLH1, as well as the Bloom syndrome gene (BS), RecQ type DNA helicase (Monteiro, 2000; Scully and Livingston, 2000; Welcsh *et al.,* 2000). BRCA1 has also been found to interact with one of the Fanconi anemia genes, *FANCD2* (Garcia-Higuera *et al.,* 2001). BRCA1 may play an important general role in the maintenance of genome integrity during DNA synthesis, acting directly downstream of DNA damage sensors and checkpoint genes in coordinating the assembly of DNA repair complexes. These repair activities may in large be HR-related, and sister chromatid recombination constitutes one process to execute DNA repair at persistent single-strand DNA tracts and stalled replication forks (Scully *et al.,* 2000). RAD51 is a key protein in HR, and BRCA2 may serve as a scaffold for regulation of RAD51-induced nucleoprotein filaments, as well as for its nuclear localization (Davies *et al.,* 2001).

An additional role for BRCA1 in transcription-coupled repair (TCR), specifically of oxidative DNA damage, has also been suggested (Gowen *et al.,* 1998). TCR requires an active RNA polymerase II, which is compatible with the interaction between BRCA1 BRCT domains and several components of the core and holoenzyme (Monteiro, 2000; Scully and Livingston, 2000; Welcsh *et al.,* 2000). Removal of oxidative DNA damage requires excision repair proteins CSA and CSB of the Cockayne syndrome, the XPG product of the xeroderma pigmentosum syndrome (XP), and the mismatch repair protein MSH2 (Gowen *et al.,* 1998).

BARD1, a RING finger and BRCT domain protein that interacts with the RING finger of BRCA1, is also present in the BRCA1–BRCA2–RAD51 nuclear complex (Monteiro, 2000; Scully and Livingston, 2000; Welcsh *et al.,* 2000). The BRCA1–BARD1 interaction is mediated by regions adjacent to the RING fingers, thus forming a RING finger heterodimer with ubiquitine activity (Hashizume *et al.,* 2001). A disease-associated *BRCA1* mutation, C61G, located in the RING finger abolishes the polyubiquitination ability of BRCA1–BARD1. The substrates targeted for degradation are still unknown, but could be nuclear proteins involved in DNA damage repair and/or transcription (Hashizume *et al.,* 2001). BAP1, another BRCA1-associated protein that binds to the RING finger, encodes a ubiquitin C-terminal hydrolase or thiol protease that catalyzes proteolytic processing of ubiquitin (Scully and Livingston, 2000; Welcsh *et al.,* 2000). Moreover, BARD1 interacts with the mRNA polyadenylation factor CstF, which may reflect a link between repression of nuclear mRNA processing, DNA repair, and tumor progression (Kleiman and Manley, 2001).

3. CELL CYCLE CHECKPOINT FUNCTIONS AND CENTROSOME REGULATION

As mentioned previously, BRCA1 and BRCA2 are expressed late in the G_1 phase of the cell cycle, possibly by estrogen-dependent stimulation (Monteiro, 2000; Scully and Livingston, 2000; Welcsh *et al.,* 2000). Cells without functional BRCA1 do not arrest at the G_2/M checkpoint after DNA damage and are deficient in TCR (Gowen *et al.,* 1998; Larson *et al.,* 1997). Mouse embryonic fibroblasts (MEFs) with a homozygous deletion of BRCA1 exon 11 display normal (p53-induced) G_1 arrest upon IR, but are defective in G_2/M arrest, resulting in extensive chromosomal abnormalities. These cells also contain multiple centrosomes, leading to unequal chromosome segregation at mitosis, and aneuploidy. BRCA1, as well as other tumor suppressor proteins such as p53 and pRb, localizes to centrosomes, possibly via γ-tubulin, during mitosis (Hsu and White, 1998; Xu *et al.,* 1999b), suggesting a role for BRCA1 in centrosome amplification and the G_2/M checkpoint (Scully and Livingston, 2000; Welcsh *et al.,* 2000). Centrosome hyperamplification is frequently seen in advanced stages of breast cancer (Carroll *et al.,* 1999). In fact, it has been suggested that BRCA1 regulates the transition between G_2 and M phases through regulation of cdc25 kinase activity (Yarden *et al.,* 2001b). In addition, it has been suggested that BRCA1 plays a role in the regulation of apoptosis (Harkin *et al.,* 1999; Shao *et al.,* 1996). Mice have been created with several different homozygous *Brca1* or *Brca2* mutations, and null mutations result in embryonic lethality for both genes (see, *e.g.,* Gowen *et al.,* 1996, and Sharan *et al.,* 1997). Of interest, elimination of one *p53* allele completely rescues this embryonic lethality in *Brca1* null mice and restores normal mammary gland development (Xu *et al.,* 2001). These recent findings may provide a link to understanding the mechanism for *BRCA1*-associated breast carcinogenesis.

BRCA2 has been found to interact with the DNA binding protein BRCA2-associated factor 35 (BRAF35) in close association with condensed chromatin (Marmorstein *et al.,* 2001). A role for the BRCA2–BRAF35 complex in resolving and packaging of entangled chromatin fibers or maintenance of chromosome integrity throughout segregation at mitosis has been suggested. In addition, a role for the complex in DNA repair and/or recombination is possible (Marmorstein *et al.,* 2001). BRCA2 has been shown to interact with and become phosphorylated by the mitotic checkpoint protein hBUBR1 in cells with microtubuli disruption, and BRCA2 or hBUBR1 deficiency could result in genomic instability (Futamura *et al.,* 2000). Finally, an interaction has also been noted between BRCA2 and DSS1, a conserved and largely uncharacterized protein of importance for proper cell cycle completion in yeast (Marston *et al.,* 1999).

IV. TUMOR PROGRESSION AND THE ROLE OF ESTROGEN IN BREAST CANCER DEVELOPMENT

Estrogen receptor levels are low in the normal breast, but vary from woman to woman, and high levels have been directly associated with an increased risk of breast cancer (Khan *et al.,* 1994). There are two types of estrogen receptors, α and β, and the α receptor has a higher affinity for estrogen than β. The relative expression of ERα to ERβ is higher in invasive tumors than in normal breast tissue (Leygue *et al.,* 1998), suggesting that the balance between these receptors is important for the relative risk of breast carcinogenesis. The level of expression of ERα is widely used as a marker of hormone responsiveness and efficacy of treatment with antiestrogenic drugs, such as tamoxifen. Levels of ER expression in tumors show an age-dependent variation; tumors from young women are often ER negative, whereas tumors from older women and male breast cancer patients often express ER (Loman *et al.,* 1998). It was recognized early on that tumors from *BRCA1* mutation carriers were often ER negative, even when the carriers were compared with age- and stage-matched controls (Johannsson *et al.,* 1997), whereas tumors from *BRCA2* mutation carriers are often ER positive (Loman *et al.,* 1998).

Different models have been proposed to account for the differences in hormone receptor status of breast cancers (Parl, 2000). One suggests that all breast cancers are ER positive initially and gradually evolve into hormone independence, resulting in loss of ER expression. An alternative model suggests that the ER status is a basic characteristic of each tumor, not a marker of tumor progression and dedifferentiation, and possibly has its origin in specific histogenesis. Alternatively, breast tumors could be of polyclonal origin and ER positive and negative cells may coexist transiently before clonal outgrowth. Moreover, it has been shown that ER positive and negative breast tumors display distinct gene expression profiles, and can be readily separated even after exclusion of ER-related genes, supporting the notion that the differences between ER positive and negative breast tumors reflect not only differences in hormone responsiveness, but also possibly differences in histogenesis (Gruvberger *et al.,* 2001).

It has been suggested that the resting adult breast epithelium is organized into myoepithelial (basal) cells and luminal epithelial cells and that the latter become multilayered upon hormonal stimulation (Osborne, 1996). Although myoepithelial cells do not express ER, the luminal component comprises both ER positive and ER negative cells. Cytokeratins 7, 8, 18, and 19 are almost exclusively expressed by luminal cells, whereas cytokeratins 5, 13, 14, and 17 are predominantly expressed by myoepithelial cells (Ronnov-Jessen *et al.,* 1996). It is conceivable that the breast epithelium includes several cell types of distinctive or successive lineage and that these can act as progenitors of

different types of breast cancer. As a consequence, the differences in phenotype between *BRCA1* and *BRCA2* breast cancers could be related to differences in stem cell origin in terms of hormone status or genetic repertoire.

BRCA1 and *BRCA2* are likely to play an important role in the regulation of growth and differentiation of the mammary gland. Mice with a conditional disruption of *Brca1* in breast epithelial cells display abnormal mammary morphogenesis, with smaller glands and ducts and endbuds that fail to branch out, most likely due to DNA-damage-induced growth arrest (Xu *et al.,* 1999a). A major factor promoting development of breast cancer seems to be estrogen stimulation of mammary epithelia. Moreover, *BRCA1* may also directly regulate proliferation of epithelial cells in the breast by modulating estrogen-dependent transcriptional pathways (Fan *et al.,* 1999). Wild-type *BRCA1* can suppress the ligand-dependent transcriptional activity (AF-2) of ERα (Fan *et al.,* 2001), and mutations in *BRCA1* can result in the loss of this ability, contributing to tumorigenesis in individuals with *BRCA1* mutations.

Perou *et al.* (2000) have suggested that breast cancers can be divided into four distinct subclasses: ER positive/luminal like, ER negative/basal-like, ER negative/Erb-B2 amplified, and ER negative/luminal-like (Erb-B2 negative, keratin 8 positive). In our investigation of gene expression profiles in hereditary breast cancers, we found that the *BRCA1* tumors included in our study showed low expression of keratin 8, Erb-B2, and ER (Hedenfalk *et al.,* 2001). This does not, however, necessarily imply a basal cell origin for *BRCA1* breast tumors, as phenotypic changes may occur during tumor evolution. Indeed, it has been demonstrated that luminal cells can give rise to myoepithelial cells (Pechoux *et al.,* 1999). Moreover, we found that a majority of *BRCA1* tumors stained positively for the luminal marker MUC-1 on tissue microarrays (see Olopade *et al.,* 2001, for a discussion).

Consequently, the distinct phenotypes of *BRCA1* and *BRCA2* tumors could be related to different stem cell origins or responsiveness to hormonal stimuli.

V. CHARACTERISTICS OF HEREDITARY BREAST CANCERS

A. Pathology and Histology of Hereditary Breast Cancer

There is emerging evidence that *BRCA1*- and *BRCA2*-associated breast cancers have distinct histopathological features. A correlation between *BRCA1* and high mitotic count, continuous pushing margins, lymphocyte infiltration, and medullary carcinoma has been shown (BCLC, 1997; Lakhani

et al., 1998). *BRCA2*-related breast cancers have also been associated with specific tumor types, although the *BRCA2* phenotype may be more heterogeneous. In the review performed by the BCLC, both *BRCA1* and *BRCA2* breast cancers were associated with a high histological grade, but *BRCA2* tumors had a high grade only because of decreased tubule formation, showing no difference from sporadic breast cancers in mitotic count or pleomorphism (BCLC, 1997).

As mentioned previously, differences in steroid receptor levels between *BRCA1* and *BRCA2* breast cancers have been found in several studies. *BRCA1* breast cancers have been found to most often be ER and progesterone receptor (PR) negative (Johannsson *et al.,* 1997; Karp *et al.,* 1997; Osin *et al.,* 1998; Verhoog *et al.,* 1998, 1999). In contrast, *BRCA2* breast cancers, although more heterogeneous in steroid receptor levels, more often tend to be ER and PR positive (Osin *et al.,* 1998; Verhoog *et al.,* 1999).

A study of the histological features of cancers in families not attributable to mutations in *BRCA1* or *BRCA2* indicated that these breast cancers differed histologically from both *BRCA1* and *BRCA2* breast cancers (Lakhani *et al.,* 2000). These non-*BRCA1*/2 breast tumors were generally of lower grade and showed less nuclear pleomorphism and lower mitotic activity than *BRCA1* and *BRCA2* breast cancers. The study also suggested that non-*BRCA1*/2 breast cancers differ from nonfamilial breast cancers (Lakhani *et al.,* 2000).

The apparently more aggressive appearance of *BRCA1* tumors could be related to their intrinsic chromosomal instability, defective DNA repair, and dysfunctional centrosome regulation. The overall high grade of *BRCA2* tumors is mainly attributable to the low degree of tubule formation, suggesting a state of dedifferentiation in these tumors. The group of non-*BRCA1*/2 cancers clearly has a less aggressive appearance than both *BRCA1* and *BRCA2* tumors, implying that the underlying susceptibility genes are of a different type than *BRCA1* and *BRCA2*.

B. Somatic Genetic Changes in Hereditary Breast Cancer

The high degree of aneuploidy associated with *BRCA1* and, to a lesser extent, *BRCA2* tumors is compatible with a role in maintaining genomic stability. Comparative genomic hybridization (CGH) (Kallioniemi *et al.,* 1992) has been used to characterize genomic alterations in these tumor types, and a high frequency of copy number alterations has been shown in *BRCA1* and *BRCA2* breast cancers, compared to sporadic cases (Tirkkonen *et al.,* 1997). Moreover, distinct profiles of copy number gains and losses have been found for both *BRCA1* and *BRCA2* breast cancers, suggesting that progression

of breast cancer traverses down distinct pathways in the different types of hereditary breast cancer (Kainu *et al.,* 2000; Wistuba *et al.,* 2000). In the study by Tirkkonen *et al.,* loss of 5q, 4q, 4p, 2q, and 12q was found to be significantly higher in *BRCA1* tumors than in sporadic tumors. In *BRCA2* tumors, loss of 13q and 6q, as well as gain of 17q23 and 20q13, was significantly more common (Tirkkonen *et al.,* 1997).

VI. OTHER CAUSES OF BREAST CANCER

A. Syndromes with Increased Incidence of Breast Cancer

Breast cancer is part of the disease spectrum in a number of multicancer syndromes of known genetic origin, such as the Li–Fraumeni syndrome, Li–Fraumeni-like syndrome, Cowden's disease, and Peutz–Jeghers syndrome where affected individuals inherit mutations in *p53*, *hCHK2, PTEN,* and *STK11/LKB1,* respectively (Bell *et al.,* 1999; Boardman *et al.,* 1998; Liaw *et al.,* 1997; Malkin *et al.,* 1990). In addition, breast cancer occurs in some women who are affected with Muir–Torre syndrome and who have been found to harbor mutations in the DNA repair genes *MSH2* and *MLH1* (Lairmore and Norton, 1997). These syndromes however, are rare in the population and account for only a small portion of hereditary breast cancer cases. Nonetheless, this suggests that additional syndromes with other combinations of malignancies may exist.

B. Other Genes Conferring Breast Cancer Susceptibility

Complete characterization of the components in the *BRCA1* and *BRCA2* signaling pathways is likely to uncover novel breast-cancer-predisposing genes. Breast cancer may be part of the AT (-like) syndromes in which *ATM, NBS1,* or *MRE11* mutations have been reported (Li *et al.,* 2000; Stewart *et al.,* 1999; Swift *et al.,* 1991). In addition, epidemiological studies of AT families suggest that heterozygote AT carriers may have an increased risk for developing breast cancer, although this observation remains controversial (Athma *et al.,* 1996; Easton, 1994). Suspected disease-associated variants of *BARD1* have been seen in patients with breast, ovarian, and uterine cancer (Thai *et al.,* 1998), and somatic *CBP/p300* mutations have been reported in breast cancer patients (Gayther *et al.,* 2000). A single-nucleotide polymorphism (SNP) in the 5′ UTR of *RAD51* has been shown to modify the penetrance of *BRCA2* mutations (Levy-Lahad *et al.,* 2001). In addition,

the location of *BRAF35* and *BRG1* on chromosome 19p coincides with a commonly deleted region in ovarian cancer, and *RAD50* is located in a region that is often deleted in *BRCA1*-associated breast cancers.

A study has reported the interaction of BRCA1 with a novel protein, termed BACH1, a function that was found to be important for DNA double-strand break repair (Cantor *et al.,* 2001). In addition, mutations in the *BACH1* gene were found in two early-onset breast cancer patients, suggesting that *BACH1* might be a novel target for germline breast-cancer-inducing mutations. Moreover, a recent study reported the association between an SNP in the prohibitin gene and breast cancer in women with a first-degree relative with the disease (Jupe *et al.,* 2001).

C. Low-Penetrance Susceptibility Genes

Low-penetrance susceptibility genes, or "modifier genes," are defined as polymorphic genes with specific alleles that are associated with an altered risk for disease susceptibility. Mutations in these low-penetrance genes might be relatively common in the general population; therefore, although each variant may be associated with only a small increased risk for breast cancer in an individual, the risk in the population as a whole might be high. Based on the apparent variability in breast cancer risk in carriers of *BRCA1* and *BRCA2* mutations, and the notion that genes that affect breast cancer risk in the general population may also presumably affect breast cancer risk in *BRCA1* and *BRCA2* mutation carriers, a number of studies have evaluated variants in candidate genes looking for modifiers of penetrance.

Modifier genes mediate a low or moderate increase in breast cancer risk. A number of candidate genetic variants have been identified in association studies to be associated with breast cancer risk. These include genes involved in steroid metabolism pathways, genes involved in metabolism of exogenous carcinogens, DNA repair pathways, and immunomodulatory pathways (Dunning *et al.,* 1999). These variant alleles are, however, only associated with risks of approximately 1.5-fold and are predicted to account for only a few percent of breast cancer incidence.

A number of polymorphisms in *CYP1A1,* a gene that encodes aryl hydrocarbon hydroxylase, which catalyzes the conversion of estradiol to hydroxylated estrogen, have been investigated in relation to their association with breast cancer risk (Crofts *et al.,* 1994). Alterations in the activity of *CYP1A1* could lead to changes in the levels of estrogen, which could affect breast cancer risk. Other members of the cytochrome P-450 family that have been investigated in association with breast cancer risk include *CYP2D6, CYP2E1,* and *CYP1* (Rebbeck, 1999).

The glutathione *S*-transferases (GSTs, *e.g., GSTM1, GSTT1*) constitute a family of genes that encode for enzymes that catalyze the conjunction of

reactive chemical intermediates to soluble glutathione conjugates to facilitate clearance. Since these enzymes metabolize environmental carcinogens, there has been interest in determining if the inability to metabolize exogenous chemicals by way of GSTs may increase breast cancer risk, but this remains to be resolved.

The *N*-acetyltransferase genes *NAT1* and *NAT2* are also important in the metabolism of carcinogens. Polymorphisms in these genes are associated with an altered rate of metabolism of carcinogens, such that variant alleles result in a slow acetylator phenotype. Findings by Rebbeck *et al.* (1997) suggest that *BRCA1* mutation carriers who smoke are at an increased risk for breast cancer if they also are slow acetylators, possibly due to an association between smoking and altered steroid hormone metabolism.

An association between age at onset of breast cancer in *BRCA1* mutation carriers and length of the polymorphic androgen receptor (AR) CAG repeat has been suggested (Rebbeck *et al.,* 1999). AR alleles containing longer CAG repeats are associated with a decreased ability to activate androgen-responsive genes. Together with the finding that BRCA1 interacts physically with and is a coactivator of the AR promoter (Park *et al.,* 2000), this provides evidence that allelic variation in AR may affect breast cancer penetrance in *BRCA1* mutation carriers. Similarly, a correlation between glutamine repeat lengths in the *AIB1* gene and breast cancer risk in women with *BRCA1* mutations has been suggested, but no evidence for such a correlation was found in a large study of breast cancer patients (Haiman *et al.,* 2000). It remains possible, however, that the *AIB1* genotype may be involved in breast cancer risk in individuals highly predisposed to breast cancer.

Other allelic variants that have been associated with an increased breast cancer risk include *ATM* (mutated in ataxia telangiectasia), *ERCC* (excision-repair cross-complementation) family members, BRCA1-associated RING domain-1 (*BARD1*), and the 17β-hydroxysteroid dehydrogenase 1 (*HSD-17B1*) (Feigelson *et al.,* 2001; Rebbeck, 1999).

D. The Search for BRCA3

Several genomic regions have been suggested as candidate loci for additional breast cancer susceptibility genes, but they remain to be confirmed in other studies. Chromosome 8p has been proposed to harbor a breast cancer predisposition gene (Seitz *et al.,* 1997), although no evidence of linkage was found in a subsequent study, suggesting that if a breast cancer susceptibility exists at this locus it accounts for only a very small proportion of familial breast cancer (Rahman *et al.,* 2000). Kainu *et al.* (2000) found a high frequency of deletions at 13q21, more distal than the location of *BRCA2* and *Rb,* in non-*BRCA1/2* inherited tumors, suggesting the presence of an additional tumor suppressor gene. However, evidence against this locus has been based

on a collection of 119 families (Thompson *et al.,* 2001), emphasizing genetic heterogeneity and the need to cluster families prior to linkage analysis.

VII. GENE EXPRESSION ANALYSIS

Until recently, the approach to understanding the molecular basis of complex diseases such as cancer was to study the behavior of individual genes one at a time. The recent development of two powerful tools, microarrays and serial analysis of gene expression (SAGE), to determine the expression patterns of thousands of genes simultaneously enables scientists to study overall gene expression patterns, thereby revealing global gene expression profiles. There are two types of microarrays, those containing arrayed cDNA clones of approximately 500–2000 kb in length (Schena *et al.,* 1995) and oligonucleotide arrays where the arrayed probes consist of 20- to 80-mer oligonucleotide fragments (Lockhart *et al.,* 1996). The principle of SAGE (Velculescu *et al.,* 1995) is based on the generation of a library of short oligonucleotide fragments (SAGE tags, typically 9–10 bp) and concatenation of these sequence tags, followed by serial sequencing of multiple tags within a clone.

One difficulty in the study of *BRCA1*- and *BRCA2*-induced breast cancer is the limited availability of *in vitro* model systems. Only one established breast cancer cell line with a disease-causing *BRCA1* mutation has been established to date, HCC1937 (Tomlinson *et al.,* 1998), and no breast cancer cell lines with *BRCA2* mutations have been described. The HCC1937 cell line has been used to study *BRCA1*-dependent global gene expression patterns upon DNA damage, with results that support the role of *BRCA1* in chromatin remodeling and DNA repair and replication (Yarden *et al.,* 2001a).

Because of the differences in histopathology, genomic alterations, and steroid receptor levels, as well as in function of the genes, it seems likely that *BRCA1* and *BRCA2* breast cancers might display identifiable differences at the gene expression level. To address this possibility, we have analyzed gene expression levels in primary breast tumors from *BRCA1* and *BRCA2* mutation carriers, as well as sporadic tumors using cDNA microarrays containing approximately 6500 cDNA clones (Hedenfalk *et al.,* 2001).

A. Microarray Technique

The use of cDNA microarrays to study gene expression patterns in cancer was first described by DeRisi *et al.* (1996). cDNA microarrays offer a systematic method for performing very extensive gene expression profiling

within single cancer specimens. The technology is based on robotic spotting of thousands of cDNA probes onto glass microscope slides in a high-density manner (Fig. 3; see color insert). Fluorescently labeled tumor cDNA is hybridized onto the array together with a differentially labeled reference cDNA. The ratio of the two fluorescence intensities at each spot indicates the relative expression of that gene within the tumor and can be compared with the expression ratios for all other tumors analyzed. cDNA microarray analysis is a high-throughput technology and makes possible molecular classification of, for example, cancers. Moreover, this technology will provide researchers with new possibilities for identifying prognostic markers as well as targets for treatment.

The advent of microarray technology has made possible genome-wide expression profiling of diseases such as cancer. It has been proposed that a distinct cancer taxonomy can be identified by thus analyzing global gene expression patterns, and, to date, classification of rhabdomyosarcoma (Khan *et al.,* 1998), sporadic breast cancer (Perou *et al.,* 1999), leukemia (Golub *et al.,* 1999), lymphoma (Alizadeh *et al.,* 2000), and melanoma (Bittner *et al.,* 2000), as well as hereditary breast cancer (Hedenfalk *et al.,* 2001) and, recently, different forms of childhood cancer (Khan *et al.,* 2001), has been performed.

B. Data Analysis

The unprecedented quantity of data on gene expression patterns that is generated by microarray experiments has led to a need for extensive computational tools for analyzing the results. A number of available computational tools have been used to analyze gene expression patterns (see, *e.g.,* Quackenbush, 2001, and Brazma and Vilo, 2000, for reviews). It is no overstatement to say that the methods used to analyze gene expression patterns can have an influence on the interpretation of the results. It should be clear that the choice of an appropriate algorithm is dependent on the biological question explored and therefore is an integral part of the experimental design. Here we will limit ourselves to the discussion of algorithms primarily in the context of gene expression analysis of cancer. The methods can be separated into two groups: unsupervised and supervised. In unsupervised methods the gene expression patterns are grouped solely according to the expression data. If one has some previous information or prejudice about which samples or genes that are expected to group together, this information can be utilized in a supervised method.

Many of the algorithms used to analyze microarray data are based on the pairwise comparison of expression patterns of either genes or samples. This is addressed by mathematically defining a distance between genes or samples

in "expression space." Clustering algorithms group samples or genes based on their separation in expression space, as given by the distance metric. It should be noted that different choices of distance metric will place different objects in different clusters (Quackenbush, 2001).

One of the most commonly used tools for displaying large data sets generated from microarray experiments is two-dimensional hierarchical clustering (Eisen *et al.,* 1998). It is an agglomerative method in which expression profiles are joined to form groups, which are further joined until completion, and a single hierarchical tree is formed. The groups are joined based on their distance in expression space, such that close samples are joined early in the process, whereas more dissimilar samples are added to more distant nodes in the tree. Genes are grouped independent of samples into a separate tree. One advantage with hierarchical clustering is that the results are simply visualized as a tree. In addition, once the genes and samples are sorted in their respective trees, a color matrix is commonly used to display the variation in gene expression across all samples and genes, allowing visual inspection of gene expression patterns. Hierarchical clustering has, for example, been used to investigate molecular profiles of human breast tumors (Perou *et al.,* 2000), to classify diffuse large B-cell lymphoma (Alizadeh *et al.,* 2000), and to distinguish colon adenocarcinomas from adenomas and paired normal colon tissues (Notterman *et al.,* 2001).

Another unsupervised approach used to find groups is self organizing maps (SOMs) (Kohonen, 2001), a neural-network-based divisive clustering approach in which samples are partitioned into a predefined number of clusters. Compared to other methods that group samples into a set of clusters, such as *k*-means clustering (Bishop, 1995), SOMs have the additional feature that the clusters are presented in a geometric configuration, typically a two-dimensional grid. The two-dimensional grid is ordered so that clusters containing samples that are similar to one another are located close on the grid. In this way, one achieves not only a partitioning of the samples but also an indication of relationships between the clusters. SOMs have been used to cluster both genes and samples. Tamayo *et al.* (1999) used SOMs to cluster genes in experiments designed to investigate hematopoietic differentiation, and Golub *et al.* (1999) used SOMs to classify leukemias.

Since the number of genes measured is very large, one cannot visualize samples in expression space directly. One way to reduce the dimensionality of the samples to allow them to be shown in a lower dimensional space, which can be more easily visualized, is multidimensional scaling (MDS) analysis. In MDS the samples are placed in a lower dimensional space with the objective of preserving the distance between samples in expression space as well as possible. Standard MDS is unsupervised and the resulting low-dimension representation is dependent only on the distances between samples. MDS has been utilized to categorize rhabdomyosarcoma (Khan *et al.,* 1998). To be

able to make use of previous knowledge about groups of samples, MDS has been extended to a supervised method, in which genes are weighted according to their contribution to separating the samples into the clusters expected from previous information (Bittner *et al.,* 2000). Based on this weighted gene analysis, the genes are ranked, and the MDS can be redone using a smaller number of top-ranked genes. This supervised MDS analysis has been applied to melanoma (Bittner *et al.,* 2000), as well as to the investigation of gene expression profiles in *BRCA1* and *BRCA2* tumors (Hedenfalk *et al.,* 2001). Another approach to visualizing samples or genes is principal component analysis (PCA) (Joliffe, 1986), which is a mathematical technique that reduces the effective dimension of gene expression space without significant loss of information and finds the view of the data in a lower dimensional space which best separates the data.

There are many supervised approaches to investigating, gene by gene, which genes individually are good discriminators for a given grouping of samples. A standard *t* test has been used to identify genes that discriminate *BRCA1* from *BRCA2* tumors (Hedenfalk *et al.,* 2001). To identify genes that discriminate acute myeloid leukemia from acute lymphoblastic leukemia, Golub *et al.* (1999) used a signal-to-noise ratio that was designed to find genes that on average were expressed differently in the two groups, but also had a small variation of expression within each group. Another approach is to find good discriminator genes using the total number of misclassifications (TNoM) score, which, based on a threshold expression value, measures the number of misclassified samples for each gene (Ben-Dor *et al.,* 2000). The TNoM score has been used to analyze melanoma (Bittner *et al.,* 2000) and to distinguish *BRCA1* from *BRCA2* tumors (Hedenfalk *et al.,* 2001). The *t* test, signal-to-noise ratio, and TNoM score have all been used to build classifiers to classify samples. The idea is to extract good discriminator genes for a given grouping of a set of samples and then use these genes to classify additional samples. These methods are all used to rank genes, but how many (if any) of the top-ranked genes are significantly associated with distinguishing the given cancer groups remains to be evaluated. The common approach to this aim is random permutation tests. In these tests, one randomly permutes which group each sample belongs to, and for each random grouping of samples the discriminatory power for each gene is calculated. This way one can assess whether the discriminatory power of a gene is significantly associated with the original grouping of samples one is interested in. An appealing feature of the TNoM score is that the significance of a score can be assessed analytically and the random permutations do not have to be explicitly performed.

Another class of supervised methods, which includes supervised artificial neural networks (ANNs) (Bishop, 1995) and support vector machines (SVMs) (Cristianini and Shawe-Taylor, 2000), is based on finding a model,

defined by a set of parameters, which maps a sample from expression space to a given phenotype. In contrast to methods that evaluate each gene independently, these methods can potentially find more complex patterns of gene expression that are associated with the phenotypes of interest. A so-called training set of samples is used to calibrate the models. The models are presented with the correct classification of the training samples and this information is used to tune the parameters of the models to optimally classify the training samples. The calibrated models are subsequently used to classify additional test samples. A cross-validation procedure, in which the training set is split into two groups, with one used for calibration and one for validation, is generally used to evaluate the performance of the models. SVMs have been used to predict functional roles for uncharacterized yeast open reading frames (Brown *et al.,* 2000), as well as to classify cancer tissue samples (Furey *et al.,* 2000). In addition, ANNs have been used for the diagnostic classification of cancer samples into four groups of childhood cancer (Khan *et al.,* 2001) and to investigate the gene expression patterns associated with ER status in sporadic breast tumors (Gruvberger *et al.,* 2001). Even though these methods can be used as "black boxes," which classify tumor samples based on their gene expression patterns, it is important to note that one can evaluate which features or genes were important for the classification. For ANNs a method for ranking the genes based on their importance to the classification has been developed (Khan *et al.,* 2001).

The supervised approaches to classifying human disease states using patterns of gene expression are very promising and they can potentially have great impact on the classification of cancer. However, the advantage of supervised methods, that is, that one can make use of previous knowledge about classes of disease states, obviously restricts their applicability to investigations where one has previous knowledge. In particular, they cannot be directly applied to finding new classes of cancer. Of note is that once new classes of cancer have been suggested by unsupervised class discovery methods they can be verified using supervised classification schemes. The field of class discovery based on gene expression patterns is still in an early stage and great activity is directed to developing methods for this application. Hierarchical clustering has been used to discover two molecularly distinct types of diffuse large B-cell lymphoma in which the patients in the two subgroups showed significant differences in overall survival (Alizadeh *et al.,* 2000), and to categorize breast cancer into new subtypes (Perou *et al.,* 2000). MDS weighted gene analysis has been used together with hierarchical clustering and a nonhierarchical clustering algorithm to separate cutaneous malignant melanoma into two classes that differ in their invasive properties (Bittner *et al.,* 2000). SOMs were used to show that the already known categorization of lymphoma into acute myeloid and acute lymphoblastic lymphoma could be discovered solely from gene expression patterns (Golub

et al., 1999). Another approach is to score candidate partitions of samples and to find the partition with the largest score. Part of the problem is to find a good scoring function, which should be based on biological criteria. One such method is designed to find the partition of samples that has an overabundance of genes separating the suggested groups (Ben-Dor *et al.,* 2001). In other words, the discovered classes are those that have the largest number of discriminatory genes separating them. Regardless of which method is used to discover classes, it may be fruitful to initially obtain an estimate of the number of classes to expect from the gene expression data. For example, from the view presented by PCA the number of classes to expect can be estimated. In addition, PCA can be a powerful technique when used in conjunction with another classification technique such as SOMs (Quackenbush, 2001). It can also be used together with supervised methods such as ANNs (Khan *et al.,* 2001).

An interesting application of class discovery methods is to find new classes within the group of non-*BRCA1/2* hereditary breast cancers, as it can be suspected that they comprise a heterogeneous entity. In this context, it may be beneficial to use hybrid methods in which some previous knowledge, such as information about families or populations, is used to restrict possible new classes.

C. Gene Expression Profiles in BRCA1/2 Breast Cancers

As mentioned previously, there are pathological as well as genomic and functional differences between different types of hereditary breast cancer that might suggest differences in gene expression patterns. Although a *BRCA1*-derived breast cancer displays certain histopathological characteristics that may aid in its characterization as a *BRCA1* tumor, these tumors do not constitute an entirely uniform group. Moreover, *BRCA2* breast cancers make up a considerably more heterogeneous group. Extended knowledge of the defect(s) causing the development of breast cancer may greatly improve both treatment schemes and intervention strategies for the affected individuals.

Indeed, using a microarray of approximately 6500 cDNA clones, we have shown that it was possible to subclassify *BRCA1* and *BRCA2* breast cancer, as well as to separate them from sporadic cases (Hedenfalk *et al.,* 2001). Further investigation of the genes that were found to distinguish *BRCA1* from *BRCA2* breast cancers suggested the involvement of distinct pathways of pathogenesis in these breast tumors. Moreover, the finding of a *BRCA1*-like tumor within the sporadic group warranted further analysis, revealing the presence of hypermethylation of the *BRCA1* promoter region, causing down-regulation of *BRCA1* gene expression in this tumor (Hedenfalk *et al.,*

2001). This implies that a somatic down-regulation of *BRCA1* expression can give rise to a phenotype similar to that of germline mutations and emphasizes the significance of *BRCA1* deficiency in tumor development. Such silencing of the *BRCA1* gene has been shown in a small fraction of sporadic breast cancers displaying a *BRCA1*-like phenotype (Esteller *et al.,* 2000), and this is especially interesting in light of the fact that *BRCA1* is rarely, if ever, found to be altered by somatic mutations in sporadic cases of breast cancer. This finding illustrates the power and sensitivity of gene expression profiling of cancers. Interestingly, it has been suggested that methylation of the *BRCA1* promoter can in some instances constitute the "second hit" in tumor development (Knudson, 1971) when loss of *BRCA1* is not due to loss of heterozygosity (LOH) (M. Esteller, personal communication).

Because of the differential expression of steroid hormone receptors between *BRCA1* and *BRCA2* breast cancers, it is likely that a certain degree of the differences in gene expression levels is attributable to this fact. However, as mentioned in our study, this does not fully account for the observed differences, since separation of *BRCA1* and *BRCA2* tumors was possible even after removal of ER- and PR-related genes from the analysis. Moreover, the sporadic tumors included in the study did not cluster with the *BRCA1* and *BRCA2* tumors based on the differences in hormone receptor expression, with the exception of the ER negative sporadic case displaying hypermethylation of *BRCA1* clustering with the hormone receptor negative *BRCA1* tumors. Aberrant methylation of *BRCA1* has been shown to be associated with ER negativity (Catteau *et al.,* 1999); however, this tumor clustered with the *BRCA1* mutated cases even upon removal of ER-related genes from the analysis. This suggests that a substantial portion of the gene expression profiles of *BRCA1* and *BRCA2* breast cancers are due to the underlying mutations in these genes.

The tumor from a male *BRCA2* mutation carrier displayed a gene expression profile similar to that of the other *BRCA2* tumors in the study, but was, upon class prediction using a small number of differentially expressed genes, classified as a non-*BRCA2* tumor, suggesting slightly different properties in breast tumors arising in males and females carrying *BRCA2* mutations (Hedenfalk *et al.,* 2001). Similarly, two *BRCA2* tumors with the most N-terminal mutation (*i.e.,* causing truncation of all the RAD51 binding BRC repeats) were also misclassified in the *BRCA2* positive or negative classification (Hedenfalk *et al.,* 2001), implying that genotype–phenotype correlations may exist. Accordingly, it has been suggested that frameshift and nonsense mutations occurring within the ovarian cancer cluster region (OCCR) of *BRCA2,* which largely coincides with the location of the BRC repeats and potentially results in truncating proteins lacking one or more RAD51 binding sites, are associated with a lower risk of breast cancer and higher risk of ovarian cancer than truncating mutations in the N- or C-terminal part of

the gene (Gayther *et al.,* 1997; Thompson and Easton, 2001). A similar correlation with disease phenotype may exist for mutations in *BRCA1,* where truncating mutations positioned 3′ of the large exon 11 have been associated with a lower risk of ovarian cancer than mutations occurring in the 5′ part of the gene (Gayther *et al.,* 1995), and where certain amino acid substitutions in the BRCT domains could behave distinctly (Vallon-Christersson *et al.,* 2001). One explanation for such a genotype–phenotype correlation could be that the mutant protein products retain varying degrees of wild-type *BRCA1* function depending on the presence of specific domains. For instance, domains in the N-terminal part of the gene, 5′ of the supposed change point, might supply important *BRCA1* function in ovarian cells but not in breast epithelium, rendering these individuals more susceptible to ovarian cancer than to breast cancer. A possible means of elucidating such a genotype–phenotype correlation is to employ cDNA microarray analysis to investigate the relationship between gene expression profiles and the locations of mutations within the gene, the hypothesis being that the mutation site will be reflected in specific and unique gene expression profiles.

Finally, one sporadic tumor was classified as *BRCA2* mutation positive in the *BRCA2* classification (Hedenfalk *et al.,* 2001). One might speculate that the *BRCA2*-like gene expression profile of this tumor is due to epigenetic silencing, as was the case with the sporadic tumor classified as *BRCA1* mutation positive. However, there has been no evidence of aberrant methylation of CpG islands within the *BRCA2* promoter region in breast cancer tissues (Collins *et al.,* 1997).

The analysis of genes separating *BRCA1* from *BRCA2* tumors revealed a number of genes with high expression in *BRCA1*-associated tumors compared to *BRCA2*-associated tumors (Hedenfalk *et al.,* 2001). Many of these genes are known to be induced by *p53* in response to DNA damage (*e.g., MSH2, MSH6, GADD34*). It is, however, known that *p53* is mutated in a majority of *BRCA1*-associated breast cancers, leading to the possibility of a *p53*-independent activation of DNA damage response pathways in *BRCA1*-deficient tumors.

Moreover, in a previous study we found that the oncogene *MYB* was amplified and overexpressed in a subset of *BRCA1* breast carcinomas (Kauraniemi *et al.,* 2000), suggesting a role for this cell cycle regulator and transcription factor in the progression of some *BRCA1* tumors. That study showed how identification of genomic alterations by, for examples CGH, followed by targeted studies (*e.g.,* with microarrays) to pinpoint the putative target gene(s) within the amplicon, can be a useful approach to recognizing potentially significant genetic alterations in breast carcinogenesis.

These studies have indicated that genomic approaches to classification of hereditary breast cancers are possible and that further studies will likely pinpoint the most significant genes to differentiate subgroups and influence

prognosis and treatment. To further validate and extend cDNA microarray findings, one can use tissue microarrays containing large numbers of paraffin-embedded tissue specimens (see Kallioniemi *et al.,* 2001, for a review). This technology allows analysis of DNA, RNA, and protein across hundreds of tissue specimens in parallel and has successfully been applied to, for example verifying genes of significance in renal carcinomas and breast cancer (Moch *et al.,* 1999; Bärlund *et al.,* 2000). In addition, it appears possible that more extended studies in this field may lead to the development of a "classification chip" containing a small number of highly differentially expressed genes that could be used in the clinical setting to rapidly screen for mutations.

D. Discovering New Classes within the BRCA*x* Breast Cancers

cDNA microarray-based gene expression profiling provides a powerful tool for the elucidation of differences in tumor phenotypes and may also be used for the discovery of novel subgroups within the group of non-*BRCA1/2* hereditary breast tumors. It is likely that the most efficient approach to identifying additional breast cancer predisposition genes is the combination of large-scale gene expression profiling with positional linkage information. As mentioned previously, the non-*BRCA1/2* subgroup of breast cancers appears to be part of a histologically heterogeneous group, indicating the presence of multiple underlying alterations. This fact, in addition to the possible presence of sporadic cases in these families and population heterogeneity, limits the power of traditional linkage analysis in the search for new breast cancer predisposition genes. We are therefore currently employing the strategy of initial gene expression profiling followed by the use of various statistical methods in an effort to subclassify these families into genetically homogeneous entities. This approach is then followed by CGH analysis to identify common regions of deletion within each subgroup and genome-wide scans to pinpoint regions of linkage. Preliminary results from genome-wide expression profiling of breast tumors that have tested negative for mutations in *BRCA1* and *BRCA2* show that the gene expression profiles of these tumors are distinct from those of *BRCA1* and *BRCA2* tumors (data not shown). This is consistent with the overall phenotype of these tumors being different from that of *BRCA1* and *BRCA2* tumors (Lakhani *et al.,* 2000). Various class discovery tools are applied to these data sets to subclassify these breast tumors, and the emerging pattern has revealed that certain tumors tend to cluster together, suggesting the possibility of a common underlying genetic defect in these individuals. Tumors from related individuals tend to cluster together, implying, again, that the cause for familial clustering is genetic and not a chance clustering of breast cancer in these families (Fig. 4).

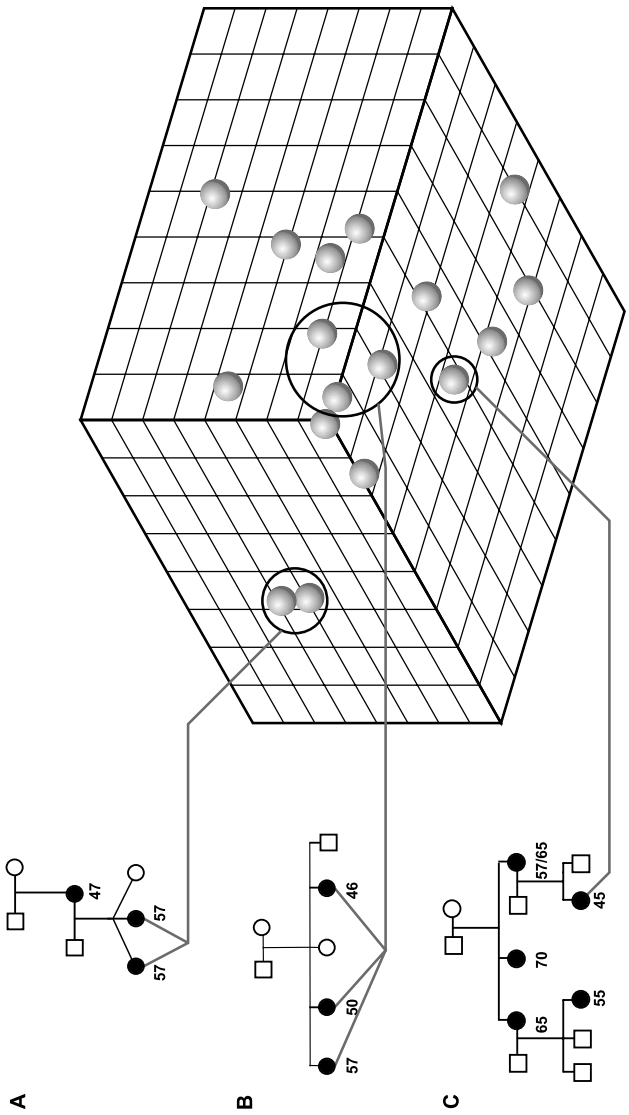

Fig. 4 Multidimensional scaling plot of *BRCAx* breast tumors. Individual tumors from non-*BRCA1/2* mutation carriers are plotted in three-dimensional space to reveal the relatedness between tumors on a gene expression level. Filled circles in the pedigrees represent individuals affected with breast cancer. The numbers below indicate age at diagnosis. Tumors from closely related individuals (pedigrees A and B) appear close to each other in this "gene expression space," indicating similarities in global gene expression patterns. One tumor from a family previously shown to be linked to chromosome 13q21 (Kainu *et al.*, 2000) is shown (pedigree C), and it is believed that other tumors with similar gene expression profiles also may be linked to this region.

E. Copy Number and Gene Expression Analysis

cDNA microarrays can also be used to analyze genomic DNA instead of RNA expression levels. As initially described by Pollack *et al.* (1999), CGH can be performed in an array format using the same cDNA microarrays. Copy number data can be compared with expression data to define candidate genes associated with either gain or loss of chromosomal regions. Moreover, targeted microarrays covering particular regions of interest can be constructed, as described by Monni *et al.* (2001), who constructed a chromosome 17q-specific microarray to analyze both copy number alterations and gene expression profiles within this commonly altered region in human breast cancer. One could envision the subclassification of non-*BRCA1/2* breast tumors into genetically homogeneous entities based on gene expression profiling and similarities in genomic alterations, followed by analysis using targeted microarrays with complete coverage of the region(s) of interest to identify the putative cancer causing gene(s). Unfortunately, the analysis of genomic alterations is currently limited to the detection of DNA amplification because of the insufficient degree of sensitivity of cDNA microarrays for detecting deletions. Moreover, based on the assumption that the underlying cause of cancer in these individuals is loss of a tumor suppressor gene, this approach is hampered by the fact that a mutation in a putative tumor suppressor gene does not necessarily result in decreased expression of the corresponding mRNA. In some instances, however, premature termination codons have been shown to initiate degradation of mutant mRNA transcripts by a mechanism termed nonsense-mediated messenger RNA decay (NMD) (Frischmeyer and Dietz, 1999). A recent report describes a strategy for identifying genes harboring nonsense mutations based on pharmacologically inhibiting the NMD pathway, resulting in stabilization of nonsense transcripts, thereby rendering detection of mutations in such genes possible (Noensie and Dietz, 2001).

VIII. CONCLUDING REMARKS

Considerable effort has focused on elucidation of the function(s) of *BRCA1* and *BRCA2* over the past two decades: from a biological point of view to increase our understanding of these tumor suppressor proteins, and from a clinical point of view in the hope of improved treatment and prophylaxis. Two important questions remain to be answered: Why the tissue specificity of tumor development in *BRCA1* and *BRCA2* mutation carriers, and why are sporadic mutations in these genes so rare? A potential explanation relates to the hormonal environment of breast and ovarian cells, and possibly to the

tissue-specific expression of as yet unidentified genes. The development of novel tools to address these questions should shed further light on the roles of *BRCA1* and *BRCA2* in normal development and in tumorigenesis. There have been many advances in the area of merging genomic research with the study of hereditary cancer, including breast cancer, and we should be prepared for surprises. Because of the power of cDNA microarray analysis, the future is likely to bring substantial changes to the molecular and pathological classification of tumors. Moreover, large-scale expression analysis will likely become increasingly useful in the search for novel therapeutic targets, as well as in the establishment of new prognostic markers for disease.

ACKNOWLEDGMENTS

We thank D. Leja for assistance with scientific illustrations. M.R. was supported by a postdoctoral fellowship from the Swedish Research Council.

REFERENCES

Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X., Powell, J. I., Yang, L., Marti, G. E., Moore, T., Hudson, J., Jr., Lu, L., Lewis, D. B., Tibshirani, R., Sherlock, G., Chan, W. C., Greiner, T. C., Weisenburger, D. D., Armitage, J. O., Warnke, R., Levy, R., Wilson, W., Grever, M. R., Byrd, J. C., Botstein, D., Brown, P. O., and Staudt, L. M. (2000). *Nature* **403,** 503–511.

Athma, P., Rappaport, R., and Swift, M. (1996). *Cancer Genet. Cytogenet.* **92,** 130–134.

Bärlund, M., Forozan, F., Kononen, J., Bubendorf, L., Chen, Y., Bittner, M. L., Torhorst, J., Haas, P., Bucher, C., Sauter, G., Kallioniemi, O. P., and Kallioniemi, A. (2000). *J. Natl. Cancer Inst.* **92,** 1252–1259.

Beger, C., Pierce, L. N., Kruger, M., Marcusson, E. G., Robbins, J. M., Welcsh, P., Welch, P. J., Welte, K., King, M. C., Barber, J. R., and Wong–Staal, F. (2001). *Proc. Natl. Acad. Sci. USA* **98,** 130–135.

Bell, D. W., Varley, J. M., Szydlo, T. E., Kang, D. H., Wahrer, D. C., Shannon, K. E., Lubratovich, M., Verselis, S. J., Isselbacher, K. J., Fraumeni, J. F., Birch, J. M., Li, F. P., Garber, J. E., and Haber, D. A. (1999). *Science* **286,** 2528–2531.

Ben-Dor, A., Bruhn, L., Friedman, N., Nachman, I., Schummer, M., and Yakhini, Z. (2000). *J. Comput. Biol.* **7,** 559–583.

Ben-Dor, A., Friedman, N., and Yakhini, Z. (2001). *In* "Proceedings of the Fifth Annual International Conference on Computational Molecular Biology (RECOMB)." ACM Press.

Bertwistle, D., Swift, S., Marston, N. J., Jackson, L. E., Crossland, S., Crompton, M. R., Marshall, C. J., and Ashworth, A. (1997). *Cancer Res.* **57,** 5485–5488.

Bignell, G., Micklem, G., Stratton, M. R., Ashworth, A., and Wooster, R. (1997). *Hum. Mol. Genet.* **6,** 53–58.

Bishop, C (1995). "Neural Networks for Pattern Recognition." Oxford University Press, Oxford.

Bittner, M., Meltzer, P., Chen, Y., Jiang, Y., Seftor, E., Hendrix, M., Radmacher, M., Simon, R., Yakhini, Z., Ben-Dor, A., Sampas, N., Dougherty, E., Wang, E., Marincola, F., Gooden, C., Lueders, J., Glatfelter, A., Pollock, P., Carpten, J., Gillanders, E., Leja, D., Dietrich, K., Beaudry, C., Berens, M., Alberts, D., Sondak, V., Hayward, N., and Trent, J. (2000). *Nature* **406,** 536–540.

Boardman, L. A., Thibodeau, S. N., Schaid, D. J., Lindor, N. M., McDonnell, S. K., Burgart, L. J., Ahlquist, D. A., Podratz, K. C., Pittelkow, M., and Hartmann, L. C. (1998). *Ann. Intern. Med.* **128,** 896–899.

Bochar, D. A., Wang, L., Beniya, H., Kinev, A., Xue, Y., Lane, W. S., Wang, W., Kashanchi, F., and Shiekhattar, R. (2000). *Cell* **102,** 257–265.

Brazma, A., and Vilo, J. (2000). *FEBS Lett.* **480,** 17–24.

Breast Cancer Linkage Consortium (1997). *Lancet* **349,** 1505–1510.

Breast Cancer Linkage Consortium (1999). *J. Natl. Cancer Inst.* **91,** 1310–1316.

Brown, M. P., Grundy, W. N., Lin, D., Cristianini, N., Sugnet, C. W., Furey, T. S., Ares, M., Jr., and Haussler, D. (2000). *Proc. Natl. Acad. Sci. USA* **97,** 262–267.

Cantor, S. B., Bell, D. W., Ganesan, S., Kass, E. M., Drapkin, R., Grossman, S., Wahrer, D. C., Sgroi, D. C., Lane, W. S., Haber, D. A., and Livingston, D. M. (2001). *Cell* **105,** 149–160.

Carroll, P. E., Okuda, M., Horn, H. F., Biddinger, P., Stambrook, P. J., Gleich, L. L., Li, Y. Q., Tarapore, P., and Fukasawa, K. (1999). *Oncogene* **18,** 1935–1944.

Casey, G. (1997). *Curr. Opin. Oncol.* **9,** 88–93.

Catteau, A., Harris, W. H., Xu, C. F., and Solomon, E. (1999). *Oncogene* **18,** 1957–1965.

Chen, G. C., Guan, L. S., Yu, J. H., Li, G. C., Choi Kim, H. R., and Wang, Z. Y. (2001). *Biochem. Biophys. Res. Commun.* **284,** 507–514.

Chen, J., Silver, D. P., Walpita, D., Cantor, S. B., Gazdar, A. F., Tomlinson, G., Couch, F. J., Weber, B. L., Ashley, T., Livingston, D. M., and Scully, R. (1998). *Mol. Cell.* **2,** 317–328.

Chen, J. J., Silver, D., Cantor, S., Livingston, D. M., and Scully, R. (1999). *Cancer Res.* **59,** 1752–1756s.

Collins, N., Wooster, R., and Stratton, M. R. (1997). *Br. J. Cancer* **76,** 1150–1156.

Cortez, D., Wang, Y., Qin, J., and Elledge, S. J. (1999). *Science* **286,** 1162–1166.

Couch, F. J., and Weber, B. L. (1996). *Hum. Mutat.* **8,** 8–18.

Couch, F. J., DeShano, M. L., Blackwood, M. A., Calzone, K., Stopfer, J., Campeau, L., Ganguly, A., Rebbeck, T., and Weber, B. L. (1997). *N. Engl. J. Med.* **336,** 1409–1415.

Cristianini, N., and Shawe-Taylor, J. (2000). "An Introduction to Support Vector Machines." Cambridge University Press, Cambridge, U.K.

Crofts, F., Taioli, E., Trachman, J., Cosma, G. N., Currie, D., Toniolo, P., and Garte, S. J. (1994). *Carcinogenesis* **15,** 2961–2963.

Cui, J., Antoniou, A. C., Dite, G. S., Southey, M. C., Venter, D. J., Easton, D. F., Giles, G. G., McCredie, M. R., and Hopper, J. L. (2001). *Am. J. Hum. Genet.* **68,** 420–431.

Davies, A. A., Masson, J. Y., McIlwraith, M. J., Stasiak, A. Z., Stasiak, A., Venkitaraman, A. R., and West, S. C. (2001). *Mol. Cell* **7,** 273–282.

Deng, C. X., and Brodie, S. G. (2000). *Bioessays* **22,** 728–737.

Deng, C. X., and Scott, F. (2000). *Oncogene* **19,** 1059–1064.

DeRisi, J., Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y., Su, Y. A., and Trent, J. M. (1996). *Nat. Genet.* **14,** 457–460.

Dunning, A. M., Healey, C. S., Pharoah, P. D., Teare, M. D., Ponder, B. A., and Easton, D. F. (1999). *Cancer. Epidemiol. Biomarkers. Prev.* **8,** 843–854.

Easton, D. F. (1994). *Int. J. Radiat. Biol.* **66,** S177–S182.

Easton, D. F. (1999). *Breast Cancer Res.* **1,** 14–17.

Easton, D. F., Bishop, D. T., Ford, D., and Crockford, G. P. (1993). *Am. J. Hum. Genet.* **52,** 678–701.

Easton, D. F., Ford, D., and Bishop, D. T. (1995). *Am. J. Hum. Genet.* **56,** 265–271.

Easton, D. F., Steele, L., Fields, P., Ormiston, W., Averill, D., Daly, P. A., McManus, R., Neuhausen, S. L., Ford, D., Wooster, R., Cannon-Albright, L. A., Stratton, M. R., and Goldgar, D. E. (1997). *Am. J. Hum. Genet.* **61,** 120–128.

Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998). *Proc. Natl. Acad. Sci. USA* **95,** 14863–14868.

Esteller, M., Silva, J. M., Dominguez, G., Bonilla, F., Matias-Guiu, X., Lerma, E., Bussaglia, E., Prat, J., Harkes, I. C., Repasky, E. A., Gabrielson, E., Schutte, M., Baylin, S. B., and Herman, J. G. (2000). *J. Natl. Cancer Inst.* **92,** 564–569.

Fan, S., Ma, Y. X., Wang, C., Yuan, R. Q., Meng, Q., Wang, J. A., Erdos, M., Goldberg, I. D., Webb, P., Kushner, P. J., Pestell, R. G., and Rosen, E. M. (2001). *Oncogene* **20,** 77–87.

Fan, S., Wang, J., Yuan, R., Ma, Y., Meng, Q., Erdos, M. R., Pestell, R. G., Yuan, F., Auborn, K. J., Goldberg, I. D., and Rosen, E. M. (1999). *Science* **284,** 1354–1356.

Feigelson, H. S., McKean-Cowdin, R., Coetzee, G. A., Stram, D. O., Kolonel, L. N., and Henderson, B. E. (2001). *Cancer Res.* **61,** 785–789.

Ford, D., Easton, D. F., Bishop, D. T., Narod, S. A., and Goldgar, D. E. (1994). *Lancet* **343,** 692–695.

Ford, D., Easton, D. F., Stratton, M., Narod, S., Goldgar, D., Devilee, P., Bishop, D. T., Weber, B., Lenoir, G., Chang-Claude, J., Sobol, H., Teare, M. D., Struewing, J., Arason, A., Scherneck, S., Peto, J., Rebbeck, T. R., Tonin, P., Neuhausen, S., Barkardottir, R., Eyfjord, J., Lynch, H., Ponder, B. A., Gayther, S. A., Birch, J. M., Lindblom, A., Stoppa-Lyonnet, D., Bignon, Y., Borg, Å., Hamann, U., Haites, N., Scott, R. J., Maugard, C. M., Vasen, H., Seitz, S., Cannon-Albright, L. A., Schofield, A., and Zelada-Hedman, M., and the Breast Cancer Linkage Consortium. (1998). *Am. J. Hum. Genet.* **62,** 676–689.

Frischmeyer, P. A., and Dietz, H. C. (1999). *Hum. Mol. Genet.* **8,** 1893–1900.

Furey, T. S., Cristianini, N., Duffy, N., Bednarski, D. W., Schummer, M., and Haussler, D. (2000). *Bioinformatics* **16,** 906–914.

Futamura, M., Arakawa, H., Matsuda, K., Katagiri, T., Saji, S., Miki, Y., and Nakamura, Y. (2000). *Cancer Res.* **60,** 1531–1535.

Garcia-Higuera, I., Taniguchi, T., Ganesan, S., Meyn, M. S., Timmers, C., Hejna, J., Grompe, M., and D'Andrea, A. D. (2001). *Mol. Cell* **7,** 249–262.

Gayther, S. A., Batley, S. J., Linger, L., Bannister, A., Thorpe, K., Chin, S. F., Daigo, Y., Russell, P., Wilson, A., Sowter, H. M., Delhanty, J. D., Ponder, B. A., Kouzarides, T., and Caldas, C. (2000). *Nat. Genet.* **24,** 300–303.

Gayther, S. A., Mangion, J., Russell, P., Seal, S., Barfoot, R., Ponder, B. A., Stratton, M. R., and Easton, D. (1997). *Nat. Genet.* **15,** 103–105.

Gayther, S. A., Warren, W., Mazoyer, S., Russell, P. A., Harrington, P. A., Chiano, M., Seal, S., Hamoudi, R., van Rensburg, E. J., Dunning, A. M., Love, R., Evans, G., Easton, D., Clayton, D., Stratton, M. R., and Ponder, B. A. J. (1995). *Nat. Genet.* **11,** 428–433.

Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D., and Lander, E. S. (1999). *Science* **286,** 531–537.

Gowen, L. C., Avrutskaya, A. V., Latour, A. M., Koller, B. H., and Leadon, S. A. (1998). *Science* **281,** 1009–1012.

Gowen, L. C., Johnson, B. L., Latour, A. M., Sulik, K. K., and Koller, B. H. (1996). *Nat. Genet.* **12,** 191–194.

Gruvberger, S., Ringnér, M., Chen, Y., Panavally, S., Saal, L. H., Borg, Å., Fernö, M., Peterson, C., and Meltzer, P. S. (2001). *Cancer Res.* **61,** 5979–5984.

Gudas, J. M., Li, T., Nguyen, H., Jensen, D., Rauscher III, F. J., and Cowan, K. H. (1996). *Cell Growth Differ.* **7,** 717–723.

Haiman, C. A., Hankinson, S. E., Spiegelman, D., Colditz, G. A., Willett, W. C., Speizer, F. E., Brown, M., and Hunter, D. J. (2000). *Breast Cancer Res.* **2,** 378–385.

Hall, J. M., Lee, M. K., Newman, B., Morrow, J. E., Anderson, L. A., Huey, B., and King, M. C. (1990). *Science* **250,** 1684–1689.

Harkin, D. P., Bean, J. M., Miklos, D., Song, Y. H., Truong, V. B., Englert, C., Christians, F. C., Ellisen, L. W., Maheswaran, S., Oliner, J. D., and Haber, D. A. (1999). *Cell* **97,** 575–586.

Hashizume, R., Fukuda, M., Maeda, I., Nishikawa, H., Oyake, D., Yabuki, Y., Ogata, H., and Ohta, T. (2001). *J. Biol. Chem.* **276,** 14537–14540.

Hedenfalk, I., Duggan, D., Chen, Y., Radmacher, M., Bittner, M., Simon, R., Meltzer, P., Gusterson, B., Esteller, M., Kallioniemi, O. P., Wilfond, B., Borg, Å., and Trent, J. (2001). *N. Engl. J. Med.* **344,** 539–548.

Houvras, Y., Benezra, M., Zhang, H., Manfredi, J. J., Weber, B. L., and Licht, J. D. (2000). *J. Biol. Chem.* **275,** 36230–36237.

Hsu, L.-C., and White, R. L. (1998). *Proc. Natl. Acad. Sci. USA* **95,** 12983–12988.

Hu, Y. F., Miyake, T., Ye, Q., and Li, R. (2000). *J. Biol. Chem.* **275,** 40910–40915.

Irminger-Finger, I., Siegel, B. D., and Leung, W. C. (1999). *Biol. Chem.* **380,** 117–128.

Johannsson, O. T., Idvall, I., Anderson, C., Borg, Å., Barkardottir, R. B., Egilsson, V., and Olsson, H. (1997). *Eur. J. Cancer* **33,** 362–371.

Joliffe, I. T. (1986). "Principal Component Analysis." Springer Verlag, New York.

Jupe, E. R., Badgett, A. A., Neas, B. R., Craft, M. A., Mitchell, D. S., Resta, R., Mulvihill, J. J., Aston, C. E., and Thompson, L. F. (2001). *Lancet* **357,** 1588–1589.

Kainu, T., Juo, S. H., Desper, R., Schaffer, A. A., Gillanders, E., Rozenblum, E., Freas-Lutz, D., Weaver, D., Stephan, D., Bailey-Wilson, J., Kallioniemi, O. P., Tirkkonen, M., Syrjakoski, K., Kuukasjarvi, T., Koivisto, P., Karhu, R., Holli, K., Arason, A., Johannesdottir, G., Bergthorsson, J. T., Johannsdottir, H., Egilsson, V., Barkardottir, R. B., Johannsson, O., Haraldsson, K., Sandberg, T., Holmberg, E., Gronberg, H., Olsson, H., Borg, Å., Vehmanen, P., Eerola, H., Heikkila, P., Pyrhonen, S., and Nevanlinna, H. (2000). *Proc. Natl. Acad. Sci. USA* **97,** 9603–9608.

Kallioniemi, A., Kallioniemi, O. P., Sudar, D., Rutovitz, D., Gray, J. W., Waldman, F., and Pinkel, D. (1992). *Science* **258,** 818–821.

Kallioniemi, O. P., Wagner, U., Kononen, J., and Sauter, G. (2001). *Hum. Mol. Genet.* **10,** 657–662.

Karp, S. E., Tonin, P. N., Begin, L. R., Martinez, J. J., Zhang, J. C., Pollak, M. N., and Foulkes, W. D. (1997). *Cancer* **80,** 435–441.

Kauraniemi, P., Hedenfalk, I., Persson, K., Duggan, D. J., Tanner, M., Johannsson, O., Olsson, H., Trent, J. M., Isola, J., and Borg, Å. (2000). *Cancer Res.* **60,** 5323–5328.

Khan, J., Simon, R., Bittner, M., Chen, Y., Leighton, S. B., Pohida, T., Smith, P. D., Jiang, Y., Gooden, G. C., Trent, J. M., and Meltzer, P. S. (1998). *Cancer. Res.* **58,** 5009–5013.

Khan, J., Wei, J. S., Ringnér, M., Saal, L. H., Ladanyi, M., Westermann, F., Berthold, F., Schwab, M., Antonescu, C. R., Peterson, C., and Meltzer, P. S. (2001). *Nat. Med.* **7,** 673–679.

Khan, S. A., Rogers, M. A., Obando, J. A., and Tamsen, A. (1994). *Cancer. Res.* **54,** 993–997.

Kleiman, F. E., and Manley, J. L. (2001). *Cell* **104,** 743–753.

Knudson, A. G., Jr. (1971). *Proc. Natl. Acad. Sci. USA* **68,** 820–823.

Kohonen, T. (2001). "Self-Organizing Maps." Springer Verlag, New York.

Lairmore, T. C., and Norton, J. A. (1997). *Am. J. Surg* **173,** 37–41.

Lakhani, S. R., Gusterson, B. A., Jacquemier, J., Sloane, J. P., Anderson, T. J., van de Vijver, M. J., Venter, D., Freeman, A., Antoniou, A., McGuffog, L., Smyth, E., Steel, C. M., Haites, N., Scott, R. J., Goldgar, D., Neuhausen, S., Daly, P. A., Ormiston, W., McManus, R., Scherneck, S., Ponder, B. A., Futreal, P. A., Peto, J., Stoppa-Lyonnet, D., Bignon, Y. J., and Stratton, M. R. (2000). *Clin. Cancer Res.* **6,** 782–789.

Lakhani, S. R., Jacquemier, J., Sloane, J. P., Gusterson, B. A., Anderson, T. J., van de Vijver, M. J., Farid, L. M., Venter, D., Antoniou, A., Storfer-Isser, A., Smyth, E., Steel, C. M., Haites, N., Scott, R. J., Goldgar, D., Neuhausen, S., Daly, P. A., Ormiston, W., McManus, R., Scherneck, S., Ponder, B. A. J., Ford, D., Peto, J., Stoppa-Lyonnet, D., Bignon, Y.-J.,

Struewing, J. P., Spurr, N. K., Bishop, D. T., Klijn, J. G. M., Devilee, P., Cornelisse, C. J., Lasset, C., Lenoir, G., Barkardottir, R. B., Egilsson, V., Hamann, U., Chang-Claude, J., Sobol, H., Weber, B., Stratton, M. R., and Easton, D. F. (1998). *J. Natl. Cancer Inst.* **90,** 1138–1145.

Larson, J. S., Tonkinson, J. L., and Lai, M. T. (1997). *Cancer. Res.* **57,** 3351–3355.

Lee, J.-S., Collins, K. M., Brown, A. L., Lee, C.-H., and Chung, J. H. (2000). *Nature* **404,** 201–204.

Levy-Lahad, E., Lahad, A., Eisenberg, S., Dagan, E., Paperna, T., Kasinetz, L., Catane, R., Kaufman, B., Beller, U., Renbaum, P., and Gershoni-Baruch, R. (2001). *Proc. Natl. Acad. Sci. USA* **98,** 3232–3236.

Leygue, E., Dotzlaw, H., Watson, P. H., and Murphy, L. C. (1998). *Cancer. Res.* **58,** 3197–3201.

Li, S., Ting, N. S., Zheng, L., Chen, P. L., Ziv, Y., Shiloh, Y., Lee, E. Y., and Lee, W. H. (2000). *Nature* **406,** 210–215.

Liaw, D., Marsh, D. J., Li, J., Dahia, P. L., Wang, S. I., Zheng, Z., Bose, S., Call, K. M., Tsou, H. C., Peacocke, M., Eng, C., and Parsons, R. (1997). *Nat. Genet.* **16,** 64–67.

Lockhart, D. J., Dong, H., Byrne, M. C., Follettie, M. T., Gallo, M. V., Chee, M. S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H., and Brown, E. L. (1996). *Nat. Biotechnol.* **14,** 1675–1680.

Loman, N., Johannsson, O., Bendahl, P. O., Borg, Å., Fernö, M., and Olsson, H. (1998). *Cancer* **83,** 310–319.

Malkin, D., Li, F. P., Strong, L. C., Fraumeni, J. F., Jr., Nelson, C. E., Kim, D. H., Kassel, J., Gryka, M. A., Bischoff, F. Z., Tainsky, M. A., and Friend, S. H. (1990). *Science* **250,** 1233–1238.

Marmorstein, L. Y., Kinev, A. V., Chan, G. K., Bochar, D. A., Beniya, H., Epstein, J. A., Yen, T. J., and Shiekhattar, R. (2001). *Cell* **104,** 247–257.

Marston, N. J., Richards, W. J., Hughes, D., Bertwistle, D., Marshall, C. J., and Ashworth, A. (1999). *Mol. Cell. Biol.* **19,** 4633–4642.

Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P. A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L. M., Ding, W., Bell, R., Rosenthal, J., Hussey, C., Tran, T., McClure, M., Frye, C., Hattier, T., Phelps, R., Haugen-Strano, A., Katcher, J., Yakumo, K., Gholalmi, Z., Shaffer, D., Stone, S., Bayer, S., Wray, C., Bogden, R., Dayananth, P., Ward, J., Tonin, P., Narod, S., Bristow, P. K., Norris, F. J., Helvering, L., Morrison, P., Rostek, P., Lai, M., Barrett, J. C., Lewis, C., Neuhausen, S., Cannon-Albright, L., Goldgar, D., Wiseman, R., Kamb, A., and Skolnick, M. H. (1994). *Science* **266,** 66–71.

Moch, H., Schraml, P., Bubendorf, L., Mirlacher, M., Kononen, J., Gasser, T., Mihatsch, M. J., Kallioniemi, O. P., and Sauter, G. (1999). *Am. J. Pathol.* **154,** 981–986.

Monni, O., Bärlund, M., Mousses, S., Kononen, J., Sauter, G., Heiskanen, M., Paavola, P., Avela, K., Chen, Y., Bittner, M. L., and Kallioniemi, A. (2001). *Proc. Natl. Acad. Sci. USA* **98,** 5711–5716.

Monteiro, A. N. (2000). *Trends Biochem. Sci.* **25,** 469–474.

Narod, S., Ford, D., Devilee, P., Barkardottir, R. B., Eyfjord, J., Lenoir, G., Serova, O., Easton, D., and Goldgar, D. (1995a). *Am. J. Hum. Genet.* **57,** 957–958.

Narod, S. A., Ford, D., Devilee, P., Barkardottir, R. B., Lynch, H. T., Smith, S. A., Ponder, B. A., Weber, B. L., Garber, J. E., Birch, J. M., Cornelis, R. S., Kelsell, D. P., Spurr, N. K., Smyth, E., Haites, N., Sobol, H., Bignon, Y. J., Chang-Claude, J., Hamann, U., Lindblom, A., Borg, Å., Piver, M. S., Gallion, H. H., Struewing, J. P., Whittemore, A., Tonin, P., Goldgar, D. E., and Easton, D. F., and the Breast Cancer Linkage Consortium. (1995b). *Am. J. Hum. Genet.* **56,** 254–264.

Noensie, E. N., and Dietz, H. C. (2001). *Nat. Biotechnol.* **19,** 434–439.

Notterman, D. A., Alon, U., Sierk, A. J., and Levine, A. J. (2001). *Cancer Res.* **61,** 3124–3130.

Olopade, O. I., Grushko, T., Sudbo, J., Reith, A., Lingjaerde, O. C., Hedenfalk, I., Simon, R., and Trent, J. (2001). *N. Engl. J. Med.* **344,** 2028–2029.

Osborne, M. P. (1996). "Breast Development and Anatomy." Lippincott-Raven, Philadelphia.
Osin, P., Crook, T., Powles, T., Peto, J., and Gusterson, B. (1998). *Lancet* **351,** 1487.
Pao, G. M., Janknecht, R., Ruffner, H., Hunter, T., and Verma, I. M. (2000). *Proc. Natl. Acad. Sci. USA* **97,** 1020–1025.
Park, J. J., Irvine, R. A., Buchanan, G., Koh, S. S., Park, J. M., Tilley, W. D., Stallcup, M. R., Press, M. F., and Coetzee, G. A. (2000). *Cancer Res.* **60,** 5946–5949.
Parl, F. F. (2000). *In* "Estrogens, Estrogen Receptor and Breast Cancer." IOS Press, Amsterdam.
Paull, T. T., Cortez, D., Bowers, B., Elledge, S. J., and Gellert, M. (2001). *Proc. Natl. Acad. Sci. USA* **98,** 6086–6091.
Pechoux, C., Gudjonsson, T., Ronnov-Jessen, L., Bissell, M. J., and Petersen, O. W. (1999). *Dev. Biol.* **206,** 88–99.
Perou, C. M., Jeffrey, S. S., van de Rijn, M., Rees, C. A., Eisen, M. B., Ross, D. T., Pergamenschikov, A., Williams, C. F., Zhu, S. X., Lee, J. C. F., Lashkari, D., Shalon, D., Brown, P. O., and Botstein, D. (1999). *Proc. Natl. Acad. Sci. USA* **96,** 9212–9217.
Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S. X., Lonning, P. E., Borresen-Dale, A. L., Brown, P. O., and Botstein, D. (2000). *Nature* **406,** 747–752.
Peto, J., Collins, N., Barfoot, R., Seal, S., Warren, W., Rahman, N., Easton, D. F., Evans, C., Deacon, J., and Stratton, M. R. (1999). *J. Natl. Cancer Inst.* **91,** 943–949.
Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D., and Brown, P. O. (1999). *Nat. Genet.* **23,** 41–46.
Quackenbush, J. (2001). *Nat. Rev. Genet.* **2,** 418–427.
Rahman, N., Teare, M. D., Seal, S., Renard, H., Mangion, J., Cour, C., Thompson, D., Shugart, Y., Eccles, D., Devilee, P., Meijers, H., Nathanson, K. L., Neuhausen, S. L., Weber, B., Chang-Claude, J., Easton, D. F., Goldgar, D., and Stratton, M. R. (2000). *Oncogene* **19,** 4170–4173.
Rajan, J. V., Marquis, S. T., Gardner, H. P., and Chodosh, L. A. (1997). *Dev. Biol.* **184,** 385–401.
Rebbeck, T. R. (1999). *Cancer* **86,** 2493–2501.
Rebbeck, T. R., Kantoff, P. W., Krithivas, K., Neuhausen, S., Blackwood, M. A., Godwin, A. K., Daly, M. B., Narod, S. A., Garber, J. E., Lynch, H. T., Weber, B. L., and Brown, M. (1999). *Am. J. Hum. Genet.* **64,** 1371–1377.
Rebbeck, T. R., Rogatko, A., and Viana, M. A. (1997). *Genet. Test.* **1,** 157–164.
Ronnov-Jessen, L., Petersen, O. W., and Bissell, M. J. (1996). *Physiol. Rev.* **76,** 69–125.
Ruffner, H., Jiang, W., Craig, A. G., Hunter, T., and Verma, I. M. (1999). *Mol. Cell Biol.* **19,** 4843–4854.
Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. (1995). *Science* **270,** 467–470.
Scully, R., and Livingston, D. M. (2000). *Nature* **408,** 429–432.
Scully, R., Chen, J., Plug, A., Xiao, Y., Weaver, D., Feunteun, J., Ashley, T., and Livingston, D. M. (1997). *Cell* **88,** 265–275.
Scully, R., Ganesan, S., Brown, M., De Caprio, J. A., Cannistra, S. A., Feunteun, J., Schnitt, S., and Livingston, D. M. (1996). *Science* **272,** 123–126.
Scully, R., Puget, N., and Vlasakova, K. (2000). *Oncogene* **19,** 6176–6183.
Seitz, S., Rohde, K., Bender, E., Nothnagel, A., Kolble, K., Schlag, P. M., and Scherneck, S. (1997). *Oncogene* **14,** 741–743.
Shao, N., Chai, Y. L., Shyam, E., Reddy, P., and Rao, V. N. (1996). *Oncogene* **13,** 1–7.
Sharan, S. K., Morimatsu, M., Albrecht, U., Lim, D. S., Regel, E., Dinh, C., Sands, A., Eichele, G., Hasty, P., and Bradley, A. (1997). *Nature* **386,** 804–810.
Smith, T. M., Lee, M. K., Szabo, C. I., Jerome, N., McEuen, M., Taylor, M., Hood, L., and King, M. C. (1996). *Genome Res.* **6,** 1029–1049.
Stewart, G. S., Maser, R. S., Stankovic, T., Bressan, D. A., Kaplan, M. I., Jaspers, N. G., Raams, A., Byrd, P. J., Petrini, J. H., and Taylor, A. M. (1999). *Cell* **99,** 577–587.

Struewing, J. P., Hartge, P., Wacholder, S., Baker, S. M., Berlin, M., McAdams, M., Timmerman, M. M., Brody, L. C., and Tucker, M. A. (1997). *N. Engl. J. Med.* **336,** 1401–1408.

Struewing, J. P., Tarone, R. E., Brody, L. C., Li, F. P., and Boice, J. D., Jr. (1996). *Lancet* **347,** 1493.

Swift, M., Morrell, D., Massey, R. B., and Chase, C. L. (1991). *N. Engl. J. Med.* **325,** 1831–1836.

Szabo, C. I., and King, M. C. (1997). *Am. J. Hum. Genet.* **60,** 1013–1020.

Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E. S., and Golub, T. R. (1999). *Proc. Natl. Acad. Sci. USA* **96,** 2907–2912.

Tavtigian, S. V., Simard, J., Rommens, J., Couch, F., Shattuck-Eidens, D., Neuhausen, S., Merajver, S., Thorlacius, S., Offit, K., Stoppa-Lyonnet, D., Belanger, C., Bell, R., Berry, S., Bogden, R., Chen, Q., Davis, T., Dumont, M., Frye, C., Hattier, T., Jammulapati, S., Janecki, T., Jiang, P., Kehrer, R., Leblanc, J. F., Mitchell, J. T., McArthur-Morrison, J., Nguyen, K., Peng, Y., Samson, C., Schroeder, M., Snyder, S. C., Steele, L., Stringfellow, M., Stroup, C., Swedlund, B., Swensen, J., Teng, D., Thomas, A., Tran, T., Tran, T., Tranchant, M., Weaver-Feldhaus, J., Wong, A. K. C., Shizuya, H., Eyfjord, J. E., Cannon-Albright, L., Labrie, F., Skolnick, M. H., Weber, B., Kamb, A., and Goldgar, D. E. (1996). *Nat. Genet.* **12,** 333–337.

Thai, T. H., Du, F., Tsan, J. T., Jin, Y., Phung, A., Spillman, M. A., Massa, H. F., Muller, C. Y., Ashfaq, R., Mathis, J. M., Miller, D. S., Trask, B. J., Baer, R., and Bowcock, A. M. (1998). *Hum. Mol. Genet.* **7,** 195–202.

Thompson, D., and Easton, D. (2001). *Am. J. Hum. Genet.* **68,** 410–419.

Thompson, D., Mangion, J., Szabo, C., van Eijk, R., Odefrey, F., Seal, S., Barfoot, R., Teare, D., Weber, B., Devilee, P., Easton, D., Goldgar, D., and Stratton, M., and the International BRCA3 Linkage Consortium. (2001). *Proc. Am. Assoc. Cancer. Res.* **42** (Abstr. No. LB-53)

Tirkkonen, M., Johannsson, O., Agnarsson, B. A., Olsson, H., Ingvarsson, S., Karhu, R., Tanner, M., Isola, J., Barkardottir, R. B., Borg, Å., and Kallioniemi, O.-P. (1997). *Cancer. Res.* **57,** 1222–1227.

Tomlinson, G. E., Chen, T. T.-L., Stastny, V. A., Virmani, A. K., Spillman, M. A., Tonk, V., Blum, J. L., Schneider, N. R., Wistuba, I. I., Shay, J. W., Minna, J. D., and Gazdar, A. F. (1998). *Cancer. Res.* **58,** 3237–3242.

Unger, M. A., Nathanson, K. L., Calzone, K., Antin-Ozerkis, D., Shih, H. A., Martin, A. M., Lenoir, G. M., Mazoyer, S., and Weber, B. L. (2000). *Am. J. Hum. Genet.* **67,** 841–850.

Vallon-Christersson, J., Cayanan, C., Haraldsson, K., Loman, N., Bergthorsson, J. T., Brondum-Nielsen, K., Gerdes, A. M., Moller, P., Kristoffersson, U., Olsson, H., Borg, Å., and Monteiro, A. N. (2001). *Hum. Mol. Genet.* **10,** 353–360.

Velculescu, V. E., Zhang, L., Vogelstein, B., and Kinzler, K. W. (1995). *Science* **270,** 484–487.

Verhoog, L. C., Brekelmans, C. T., Seynaeve, C., Dahmen, G., van Geel, A. N., Bartels, C. C., Tilanus-Linthorst, M. M., Wagner, A., Devilee, P., Halley, D. J., van den Ouweland, A. M., Meijers-Heijboer, E. J., and Klijn, J. G. (1999). *J. Clin. Oncol.* **17,** 3396–3402.

Verhoog, L. C., Brekelmans, C. T., Seynaeve, C., van den Bosch, L. M., Dahmen, G., van Geel, A. N., Tilanus-Linthorst, M. M., Bartels, C. C., Wagner, A., van den Ouweland, A., Devilee, P., Meijers-Heijboer, E. J., and Klijn, J. G. (1998). *Lancet* **351,** 316–321.

Wang, Q., Zhang, H., Fishel, R., and Greene, M. I. (2000). *Oncogene* **19,** 6152–6158.

Wang, S. C., Lin, S. J., Su, L. K., and Hung, M. C. (1997). *Biochem. Biophys. Res. Commun.* **234,** 247–251.

Welcsh, P. L., Owens, K. N., and King, M.-C. (2000). *Trends. Genet.* **16,** 69–74.

Wistuba, II, Tomlinson, G. E., Behrens, C., Virmani, A., Geradts, J., Blum, J. L., Minna, J. D., and Gazdar, A. F. (2000). *Gene. Chromosome Cancer* **28,** 359–369.

Wooster, R., Bignell, G., Lancaster, J., Swift, S., Seal, S., Mangion, J., Collins, N., Gregory, S., Gumbs, C., Micklem, G., Barfoot, R., Hamoudi, R., Patel, S., Rice, C., Biggs, P., Hashim, Y.,

Smith, A., Connor, F., Arason, A., Gudmundsson, J., Ficenec, D., Kelsell, D., Ford, D., Tonin, P., Bishop, D. T., Spurr, N. K., Ponder, B. A. J., Eeles, R., Peto, J., Devilee, P., Cornelisse, C., Lynch, H., Narod, S., Lenoir, G., Egilsson, V., Barkardottir, R. B., Easton, F. E., Bentley, D. R., Futreal, P. A., Ashworth, A., and Stratton, M. R. (1995). *Nature* **378,** 789–792.

Wooster, R., Neuhausen, S. L., Mangion, J., Quirk, Y., Ford, D., Collins, N., Nguyen, K., Seal, S., Tran, T., Averill, D., Fields, P., Marshall, G., Narod, S., Lenoir, G. M., Lynch, H., Feunteun, J., Devilee, P., Cornelisse, C. J., Menko, F. H., Daly, P. A., Ormiston, W., McManus, R., Pye, C., Lewis, C. M., Cannon-Albright, L. A., Peto, J., Ponder, B. A. J., Skolnick, M. H., Easton, D. F., Goldgar, D. E., and Stratton, M. R. (1994). *Science* **265,** 2088–2090.

Wu, K., Jiang, S. W., Thangaraju, M., Wu, G., and Couch, F. J. (2000). *J. Biol. Chem.* **275,** 35548–35556.

Xu, X., Qiao, W., Linke, S. P., Cao, L., Li, W. M., Furth, P. A., Harris, C. C., and Deng, C. X. (2001). *Nat. Genet.* **28,** 266–271.

Xu, X., Wagner, K.-U., Larson, D., Weaver, Z., Li, C., Ried, T., Hennighausen, L., Wynshaw-Boris, A., and Deng, C.-X. (1999a). *Nat. Genet.* **22,** 37–43.

Xu, X., Weaver, Z., Linke, S. P., Li, C., Gotay, J., Wang, X. W., Harris, C. C., Ried, T., and Deng, C. X. (1999b). *Mol. Cell.* **3,** 389–395.

Yarden, R. I., and Brody, L. C. (1999). *Proc. Natl. Acad. Sci. USA* **96,** 4983–4988.

Yarden, R. I., Duggan, D. J., Sgagias, M. K., Cowan, K. H., and Brody, L. C. (2001a). *Proc. Am. Assoc. Cancer Res.* **42** (Abstr. No. 2296)

Yarden, R. I., Pardo-Reoyo, S., Sgagias, M. K., Cowan, K. H., and Brody, L. C. (2001b). *Proc. Am. Assoc. Cancer. Res.* **42** (Abstr. No. 721)

Yu, X., and Baer, R. (2000). *J. Biol. Chem.* **275,** 18541–18549.

Zheng, L., Li, S., Boyer, T. G., and Lee, W. H. (2000a). *Oncogene* **19,** 6159–6175.

Zheng, L., Pan, H., Li, S., Flesken-Nikitin, A., Chen, L. P., Boyer, G. T., and Lee, H. W. (2000b). *Mol. Cell* **6,** 757–768.

Zhong, Q., Chen, C.-F., Li, S., Chen, Y., Wang, C.-C., Xiao, J., Chen, P.-L., Sharp, Z. D., and Lee, W.-H. (1999). *Science* **285,** 747–750.