Breast Cancer Research and Treatment **28**: 121-135, 1993.

Molecular cloning of BRCA1: a gene for early onset familial breast and ovarian cancer

Anne M. Bowcock
Department of Pediatrics, University of Texas Southwestern Medical Center, Dallas TX, USA

Key words: chromosome 17q mapping, familial breast/ovarian cancer gene, gene mapping, molecular biology techniques, tumor suppressor genes, yeast artificial chromosomes (YACs)

Summary

Molecular analyses allow one to determine genetic lesions occurring early in the development of tumors. With positional cloning approaches we are searching for a gene involved in the development of early onset familial breast and ovarian cancer that maps to human chromosome 17q21 and is termed BRCA1. This involves localizing the region genetically within families with multiply affected members, capturing the region identified by genetic analyses in YACs (yeast artificial chromosomes), converting those YACs to smaller manipulable pieces (such as cosmids), and searching for genes via a variety of approaches such as direct screening of cDNA libraries with genomic clones, direct selection by hybridization, "exon trapping", and CpG island rescue. Once identified, candidate genes will be screened for mutations in affected family members in whom breast cancer segregates with the locus on 17q21. The frequency of this gene has been calculated to be 0.0033; from this the incidence of carriers, i.e. those carrying such a predisposition, is one in 150 women. The isolation of BRCA1 and the elucidation of the mutations resulting in breast and ovarian cancer predisposition will allow identification of women who have inherited germ-line mutations in BRCA1. In families known to harbor a germ-line BRCA1 mutation, diagnosis of affected members will be rapid. It is possible that one will also be able to detect alterations of the second copy of this gene early in tumor development in individuals carrying a germ-line mutation. It is not yet known how frequently somatic BRCA1 mutations predispose to breast and ovarian carcinoma in the general female population. If, as in other genetic diseases, new germ-line mutations occur in some women and thus contribute to the development of breast cancer, it may be feasible to screen women in the general population for predisposing mutations. In addition, if acquired genetic mutations of the BRCA1 gene are involved as early events in the development of non-familial forms of the disease, early detection of possible breast carcinoma may become feasible in biopsy of breast tissue.

Introduction

Although the incidence of breast cancer is estimated to be 1/9 for a woman over her lifetime, certain women appear to be at an increased risk. These women harbor germ-line mutations that predispose to breast cancer susceptibility, and in general develop the disease at an earlier age. A characteristic of familial cancers is that besides being of earlier onset than normal, the

Address for correspondence and offprints: Dr. A.M. Bowcock, Department of Pediatrics, McDermott Center for Human Growth and Development, 6000 Harry Hines Blvd., Dallas, TX 75235-8591; *Tel:* 214-648-1600; *Fax:* 214-648-1666

cancer is often bilateral. Some of the women who are at increased risk of developing breast and ovarian cancer harbor a mutation in a gene termed BRCA1 (see Table 1 for list of abbreviations used). Over their lifetime, the likelihood that these women will develop breast or ovarian cancer is approximately 90%.

Tumor development and progression is accompanied by a series of events that occur in a single clone of cells as a result of molecular lesions in a specific set of genes. Breast cancer, like other cancers, is likely to occur as a result of aberrant gene expression. Some of this is due to loss of expression of genes such as those that normally regulate or suppress cell growth (tumor suppressor genes or anti-oncogenes), while some is due to an increase in gene expression, e.g. the activation of growth-promoting factors such as proto-onco-

Table 1. Table of abbreviations

Alu-PCR:	(see IRS-PCR)
APC:	adenomatosis polyposis coli
BAC:	bacterial artificial chromosome
BRCA1:	gene for familial early onset breast and ovarian cancer
cDNA:	DNA copied off an mRNA template by reverse transcriptase
cM:	centimorgan, or 1/100th of a Morgan
COL1A1:	collagen, type I, alpha I
DCC:	deleted in colorectal carcinoma
DGGE:	denaturing gradient gel electrophoresis
D17S74:	74th single copy segment of DNA to be isolated from human chromosome 17 (term for a locus)
EDH17B2:	estradiol 17-beta dehydrogenase I
ERBB2:	avian erythroblastic leukemia viral v-erb-B2
FAP:	familial adenomatosis polyposis coli
FISH:	fluorescence in-situ hybridization
GIP:	gastric inhibitory polypeptide
kb:	kilo-base (1,000 bases)
HER2/*neu*:	See ERBB2
HOX2:	homeo box region 2
HTF islands:	HpaII tiny fragment islands
IRS-PCR:	inverse-repeated sequence PCR
LOH:	loss of heterozygosity
Mb:	mega-base (1,000,000 bases)
Morgan:	a unit of recombination (there are 33 Morgans in the human genome)
MEN1:	multiple endocrine neoplasia I
MEN2:	multiple endocrine neoplasia II
NF2:	neurofibromatosis 2 (bilateral acoustic neuroma)
NME1:	non-metastatic cells 1, expressing NM23 protein
NME2:	non-metastatic cells 2, expressing NM23 protein
PHB:	prohibitin
RARA:	retinoic acid receptor, alpha
RFLP:	restriction fragment length polymorphism
RB1:	retinoblastoma I (including osteosarcoma)
PCR:	polymerase chain reaction
SSCP:	single-strand conformation polymorphism
STS:	sequence tagged site
THRA1:	thyroid hormone receptor alpha 1 (avian erythroblastic leukemia viral (v-erbA) oncogene homolog 1, formerly ERBA1)
VNTR:	variable number of tandem repeats
YAC:	yeast artificial chromosome

genes. Chromosomal rearrangements have identified several chromosomal regions or genes likely to be involved in either the development or the progression of breast cancer. One can also localize genes predisposing to disease on the basis of co-segregation with DNA markers in multiply affected families. The gene we are searching for was identified by Mendelian genetics, and localized to human chromosome 17q21.

Current molecular genetic technology is being used to isolate this gene; termed positional cloning, it relies on localizing the disease gene genetically, capturing the region physically, searching for the genes in the region, and identifying BRCA1 among them. Although one would expect women from a large proportion of families where breast cancer is segregating as a Mendelian trait to harbor an alteration at BRCA1, it is estimated that women with no family history may also harbor an altered BRCA1 gene. It has been estimated that 1/150 to 1/500 women are at increased risk of developing breast and ovarian cancer due to an alteration of BRCA1 in their germ-line DNA. This review describes our approaches to isolating BRCA1, and outlines positional cloning approaches in general.

Segregation analysis to dissect diseases genetically

One risk factor for the development of breast cancer is a family history of the disease [26]. One can perform a "segregation analysis" to determine the best genetic model for a disease. These studies require the ascertainment of large numbers of affected individuals from a single population and provide information on the hypothetical genetic component of the disease, e.g. whether the disease is likely to be inherited in a recessive or dominant fashion, and the penetrance of the gene (percent of members with the defective gene who will develop the disease).

A large study of 4,730 histologically confirmed breast cancer cases between the ages of 20 and 54 along with 4,688 controls has provided evidence for the existence of a rare autosomal dominant allele with a frequency of 0.0033, that leads to an increased susceptibility to breast cancer [10]. The lifetime risk for a woman with such a susceptibility allele is predicted to be 92%, in contrast to the cumulative lifetime risk of non-carriers which is estimated to be approximately 10%. This study agreed with smaller, earlier ones such as that by Newman et al [35], who studied 1,579 cases of breast cancer and predicted that women with the susceptibility allele had a lifetime risk of developing breast cancer of 82%, versus 8% for the general population. The study by Newman et al suggested that 4% of cases are due to an inherited predisposition. Other studies suggest that more than one locus may predispose to familial breast cancer. It has been shown that $< 1\%$ of women, who develop breast cancer at a very young age and who often have children who develop sarcomas, carry a germ-line mutation in the tumor suppressor gene p53 [30].

17q21 linkage

Since segregation analyses suggested that some forms of breast cancer predisposition can be accounted for by a single gene, it was reasonable to attempt to map it. Gene mapping is currently performed by linkage analysis with DNA markers in multiply affected families followed by positional cloning approaches to isolate the gene subsequently.

Linkage analysis relies on the identification of a marker or markers that segregate with disease predisposition. Markers were originally protein polymorphisms, but have been replaced by DNA markers such as RFLPs (restriction fragment length polymorphisms) and VNTR (variable number of tandem repeat polymorphisms). There are many different classes of repeats that differ in the number of copies present at any one site. Some of the most useful are variable numbers of "di, tri and tetra-nucleotide repeats". Commonly called "polymorphic microsatellites", they have revolutionized linkage analysis since they are

ubiquitous, with a microsatellite occurring approximately every 40kb. Polymorphic loci containing microsatellites are highly variable since most individuals are heterozygous. This is indispensable for linkage analysis since nearly every individual is informative and one can determine which allele is inherited with the disease gene most of the time, with the result that little information is lost from the rare but important families in which breast cancer segregates as a Mendelian trait. Microsatellites can be typed with the polymerase chain reaction, which is fast and requires approximately 100 times less DNA than RFLP-based linkage analysis (e.g. 30ng instead of 3-5μg). One additional advantage of microsatellites for a disease such as breast cancer, is that since affected members have often died at the time the family is genotyped, archival tissue such as microscope slides or paraffin blocks containing the patients' normal tissue can act as an invaluable source of DNA with which to reconstruct their genotypes. This can be performed with PCR-based typing; we routinely obtain sufficient DNA from a 10μm section for approximately 100 PCR reactions. We have also obtained sufficient DNA from tumor DNA scraped off microscope slides for approximately 25 PCR reactions.

In 1990, it was shown that a VNTR marker on chromosome 17q (D17S74, or cMM86) segregated with breast cancer predisposition in seven out of 23 families (40%) where the onset of the disease occurred before the age of 46 [22]. Here the two-point lod score in the early-onset families was high enough above the threshold value of 3.0 to be strong evidence for linkage. This analysis also demonstrated that the disease is genetically heterogeneous (i.e. that breast cancer predisposition did not always segregate with this marker and may be linked to other susceptibility genes in other families). Subsequent analyses showed that a confounding influence in the "late-onset" families is the co-occurrence of the disease in relatives due to non-germ-line alterations [31]. In the case of this predisposing gene BRCA1, D17S74 was initially shown to lie at a distance of 10% recombination from it. This represents a map distance of approximately 10cM (1% recombination corresponds to a map distance of approximately 1cM). In terms of physical distance, 1cM represents 1,0000kb, on average, although this distance can vary widely (for example, 1cM may correspond to 100kb in recombination hot spots, where there is more recombination than the average, and to 10,000kb in recombination cold spots, where it is less than the average).

The results of Hall et al [22] were quickly confirmed by Narod et al [34] who studied five families where both breast and ovarian cancer was segregating and who demonstrated that breast/ovarian cancer predisposition was linked to D17S74 in three families. Here the combined lod score was 2.20 at a recombination fraction Θ of .20 (or approximately 20cM).

In an attempt to confirm the previously published linkage results, to localize the disease locus more definitively, to examine the extent of genetic heterogeneity, and to estimate the penetrance of the BRCA1 gene, a joint analysis of data from 13 groups was performed with a total of 214 families with apparent hereditary predisposition to breast and/or ovarian cancer [15]. This localized BRCA1 to an 8.3cM interval (18 cM in females) between D17S588 and D17S250, with odds of 66:1. When families with only breast cancer were considered, breast cancer predisposition was linked to this locus in approximately 45% of the families. When families with both breast and ovarian cancer were considered, cancer predisposition was linked in nearly all cases. This suggests that a gene(s) on chromosome 17q accounts for most families with both early-onset breast and ovarian cancer, but that there exist other genes predisposing to breast cancer. In the linked families, the risk associated with inheritance of the defective gene was estimated to be 59% at age 50 and 82% at age 70 [15].

An example of a hypothetical family in which breast cancer predisposition in females segregates with a highly informative DNA marker is shown in Figure 1.

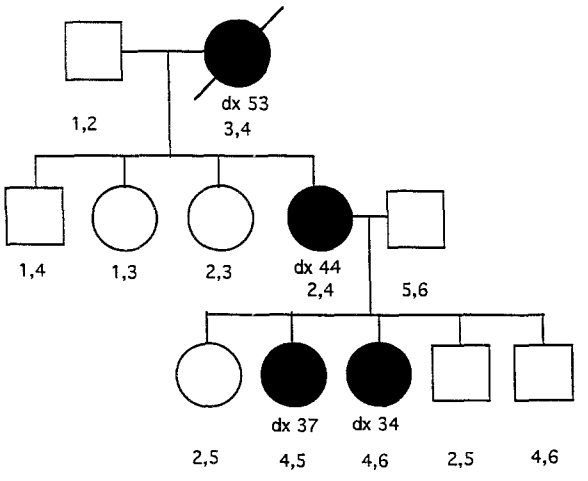

Figure 1. Example of family with early onset breast cancer in which breast cancer predisposition is segregating with allele 4 of a linked locus. Genotypes at the linked locus are shown underneath the pedigree symbols. Circles: females; squares: males; shaded circles: affected females; Dx: age at which breast cancer was diagnosed. In this pedigree, males with allele 4 are unaffected. This resembles the situation for BRCA1, where males harboring a linked allele are unaffected.

Evidence that BRCA1 is a tumor suppressor gene

There is some evidence that the BRCA1 gene is a tumor suppressor gene [54]. A hallmark of tumor suppressors is the finding of nearby allele losses, reflecting regions of chromosomal loss at the suppressor gene locus in tumor DNA [44]. In the case of a suppressor gene involved in inherited predisposition, these allele losses would be expected to occur on the chromosome containing the wild-type allele, thereby inactivating this allele (inactivation of the first allele having been inherited). When loss of heterozygosity studies are performed in tumors of affected members in multiply affected breast and ovarian cancer families shown by linkage analysis to harbor a germ-line BRCA1 mutation, it has consistently been observed by us and others [54], that the chromosome 17 which is lost is the one which carries the wild-type BRCA1 gene. The chromosome 17 retained in the tumors is the one containing the mutant BRCA1. This suggests that tumor predisposition in these cases is due to loss of a normal BRCA1 gene, and provides evidence that BRCA1 is a tumor suppressor. In the family in Figure 1, one would expect that tumors exhibiting LOH of human chromosome 17 would retain allele 4.

Familial vs sporadic forms of breast cancer

Breast cancer attributable to lesions at BRCA1 may be similar to other malignancies that occur both as a sporadic form and a familial form: e.g. renal-cell carcinomas and Von Hippel Lindau disease [51], colon tumors and familial adenomatous polyposis (FAP) and Gardner's syndrome [2, 21,23,36], and acoustic neuromas which can be found sporadically or in individuals with a genetic predisposition due to neurofibromatosis type 2 (NF2) [16].

In addition to the linkage of breast cancer predisposition to BRCA1 in some families (approximately 60% of families with three or more members with breast cancer), there is also some evidence that alterations of a gene at 17q21 occur in tumors of women with no family history. This is based primarily on studies of sporadic breast tumors. In one instance, 40.8% of premenopausal and 32.5% of postmenopausal breast carcinomas had undergone LOH at 17q21.3 [49]. Similar results have also been observed by Futreal et al [18], who describe a common region of deletion that lies between D17S250 and D17S579 at 17q11.2-a21. Other chromosomal regions implicated in the etiology of breast cancer are 3p13-14.3 (the segment where breakpoints are often seen in renal cell carcinomas [11,27,43], 11p, 13q, 16q22-q23, and 17p13 [47,48]. An association has also been demonstrated between LOH on 17p and 17q and amplification of the erbB2 oncogene [48] which has previously shown to have predictive value for recurrence of breast cancer [53]. Interestingly, tumors which had lost chromosome 17p, the locale of p53, had also lost chromosome 13q, the locale of RB1. These

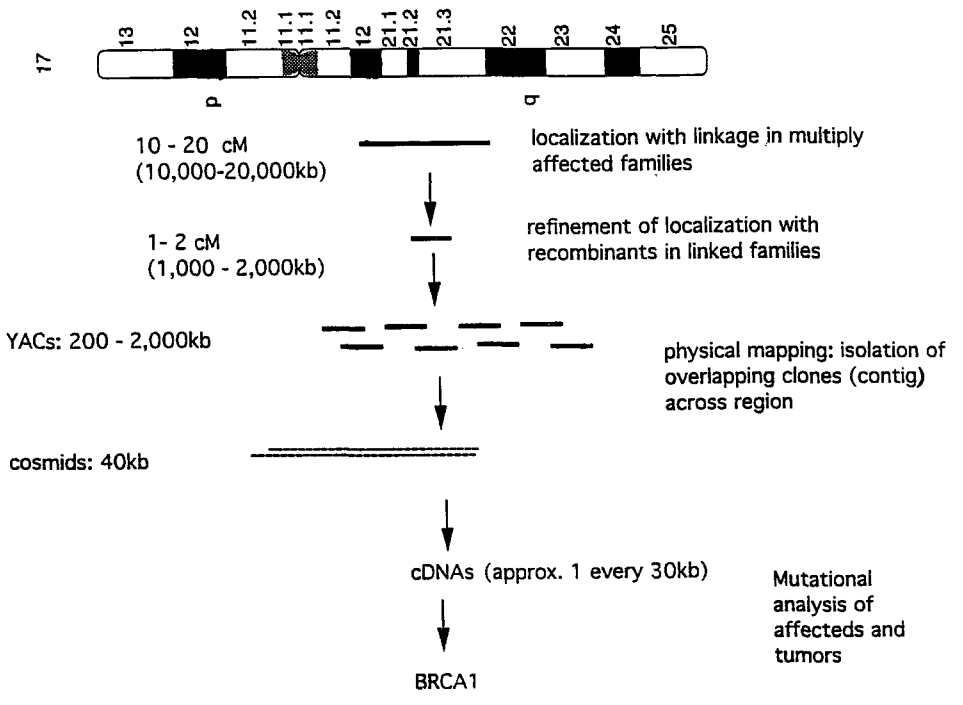

Figure 2. Outline of approaches used in positional cloning, beginning with localization of a disease locus to a chromosomal region and refinement of the region genetically, followed by cloning of the region, identification of genes in the region, and ultimately, identification of the predisposing gene on the basis that it is altered in affected members of linked families.

tumors more frequently had highly malignant histopathological features [48].

Since LOH at 17q21 has been observed in breast and ovarian carcinomas, an attempt is being made to localize BRCA1 on the basis of the smallest region that is lost. This approach facilitated the mapping of neurofibromatosis type 2 (NF2) and MEN1, as well as the cloning of the RB1 and DCC genes (see review by Ponder [8,17, 44]). However, it is not always straightforward to detect chromosomal regions on the basis of a commonly deleted region (e.g. FAP and MEN1 [44]). This may be so if the loss-of-activity mutations are not truly recessive at the cellular level, so that tumors may arise if only one allele is altered. For this reason this type of approach is being employed with reservation.

Positional cloning of BRCA1

Positional cloning has been used effectively in the isolation of several important human disease genes such as cystic fibrosis, familial adenomatous polyposis, and Huntington's disease [21,24, 46]. An outline of this approach is shown in Figure 2 and relies on first localizing a disease gene to a chromosome or chromosomal region, usually by linkage analysis, refining the region genetically by obtaining closer markers, and, once the region is small enough, cloning it in a large but manipulable form. Several different types of recombinant molecules can accommodate large segments of DNA: YACs (yeast artificial chromosomes, which accommodate segments of generally between 200kb and 2Mb), P1 phage (accommodating segments of approximately 100kb), and BACs (bacterial artificial chromosomes). These large DNA segments are usually converted to smaller overlapping ones such as cosmids (which have inserts of 40kb) in order to isolate genes. We have obtained a series of ordered, overlapping YAC clones (a YAC contig) of the region spanning BRCA1, and have identified cosmids hybri-

dizing to the YACs. The overlapping cosmid clones are now being used to screen for genes within the region that has been defined genetically. Genes that are candidates for the disease gene can then be tested for alterations in patients and in their breast and ovarian tumors.

Refinement of the gene location genetically

Once a gene has been localized by linkage to a chromosomal region, it is likely to lie several thousand base-pairs from the linked marker. The distance D17S74 was estimated to lie from the gene (10cM) corresponded to approximately 10,000kb. In order to home in on BRCA1, it was first necessary to refine the region genetically by identifying closer markers. This can be done with families in which breast cancer predisposition segregates with a locus at 17q21 (presumably BRCA1), utilizing individuals in these linked families with a chromosome carrying the altered BRCA1 allele that has undergone recombination between the disease and a closely linked marker. These recombinant individuals become crucial in the refinement of disease genes. The limit to which the region containing the disease gene can be refined genetically is dependent upon these recombinants. One caveat in the study of a common disease such as breast cancer is that one cannot discriminate between bona-fide recombinant individuals and individuals who do not carry an altered BRCA1 but develop a non-familial (sporadic) form of the disease. For this reason, refinement of the BRCA1 region genetically must be treated with some caution. With a study of recombination breakpoints in linked families, we refined BRCA1 to a region of less than 4cM, flanked by THRA1 on the centromeric side and D17S183 (SCG43) on the telomeric side [6]. This eliminated many of the genes on 17q that were candidates for being the BRCA1 gene (e.g. HER2/*neu*, THRA1, WNT3, HOX2, prohibitin [PHB], COL1A1, NME1, and NME2). The two that remained were RARA, the gene for the alpha-subunit of the retinoic acid receptor, and EDH17B2 which encodes estradiol 17β-hydroxysteroid dehydrogenase II. Subsequently these have been excluded by further genetic mapping (in the case of RARA) and sequencing of affected individuals (in the case of EDH17B2) [52]. An attempt to refine the region containing BRCA1 still further was initially hampered by an absence of additional genetic markers in this region. We have now constructed a very dense genetic map in this region which contains 33 ordered polymorphic markers (12 genes and 21 anonymous DNA segments) lying between D17S250 and D17S588. This comprises a region of approximately 8.3cM [1] with a polymorphic marker every 250kb on average.

Physical cloning of the region — YAC screens with STS

Once a region containing a disease-gene can be refined no further genetically, and no translocations or other cytogenetically visible chromosome alterations are detected in the region that may disrupt the gene in question, the only viable approach at present to obtaining the gene is to capture the linked region physically, search for all the genes within it, and see which is altered consistently in linked families and tumors as described above. Until recently, yeast artificial chromosomes (YACs) were used to clone large segments of DNA. However, YACs are not without their problems: 40 – 55% of YACs are chimeric and contain genomic sequences from two or more non-contiguous regions of the genome [9,3]. In addition, the instability of large cloned pieces of DNA results in deletions and rearrangements within some YACs. These problems need to be identified to avoid ambiguous interpretations of data. Other clones containing large inserts such as P1 bacteriophage and bacterial artificial chromosomes may not have as many associated problems as YACs, although their insert sizes are not as large and they have not been used as extensively.

We are screening YAC libraries with a PCR-

based screening method previously described [20]. We have captured almost the entire region between D17S250 and GIP in YACs with PCR assays for short, unique different sequences, termed STSs or sequence tagged sites [37]. Many of these were derived from known genes or polymorphic markers previously shown to map to this region of chromosome 17. The order of the loci in the "YAC contig" was obtained by a combination of genetic mapping and physical mapping. Physical mapping has been achieved by fluorescence in-situ hybridization" or "FISH" [56], "radiation hybrid mapping" [13], and pulsed-field gel electrophoresis [50].

YAC contig construction

YAC clones can be ordered on the basis that they share common sequences (e.g. common STSs detectable with PCR). With this type of analysis, we have linked up and ordered most of the YACs in the BRCA1 region of human chromosome 17q. Surprisingly, the chimerism rate of YACs in this region is 80%, which is higher than expected. As with most YAC contigs, there are a few gaps that still remain in the YAC contig map and some YACs have deletions and rearrangements.

YACs to cosmids

In general, it is easier to identify genes with smaller clones than YACs, such as cosmids. These have inserts of approximately 40 kb, and can be used in a variety of strategies to search for genes. In order to obtain a set of overlapping cosmid clones for the BRCA1 region, the YACs in the contig have been used to generate probes for hybridization to gridded chromosome 17-specific cosmids. The probes have been generated with "inverse repeated sequence PCR". This amplifies human DNA sequences between human-specific AluI repeats in the YAC DNA. AluI sequences are approximately 360bp long, highly conserved, and distributed throughout the human genome in 300,000 to 500,000 copies [25]. It is estimated that an AluI sequence occurs at least once every 10kb on average. Commonly called "AluI-PCR" or "IRS-PCR" [28], the approach described above will generate a variety of small fragments from the YAC inserts. Although such PCR fragments contain a small amount of AluI sequence, which is repetitive and would identify cosmids non-specifically, the AluI sequences can be blocked before hybridization by being pre-annealed with human DNA rich in repetitive sequences. An outline of this approach is shown in Figure 3. One problem with using IRS-PCR products as probes, is that regions poor in AluI sequences will be under-represented, and this may result in gaps in the resultant cosmid contig.

Identification of genes

We are screening for BRCA1 in a variety of cDNA libraries. cDNA libraries are constructed by obtaining RNA from a cell or tissue source, isolating the mRNA, converting the mRNA to cDNA (or copy DNA) with reverse transcriptase, generating double-stranded cDNA, and cloning the double stranded cDNA fragments into an appropriate vector, such a plasmid or a bacteriophage.

It is estimated that 3% of the human genome is coding sequence for functional and structural proteins [5]. Since the size of a haploid human genome is 3 X 10^9 bp, one can calculate that there are 100,000 genes. There will therefore be one gene, on average, every 30kb. Each cell type is estimated to express approximately 10,000 genes. Transcripts of these genes are expressed at varying levels of abundance ranging from one to 200,000 copies per cell. One third of all genes are expressed at low levels of approximately 1 - 10 copies per cell [19,39]. When searching for a gene such as BRCA1, one does not know at the outset, in which tissue it is expressed. This means that the choice of cDNA libraries is critical. Some libraries may be better candidates than others, however. For example, 30% of all

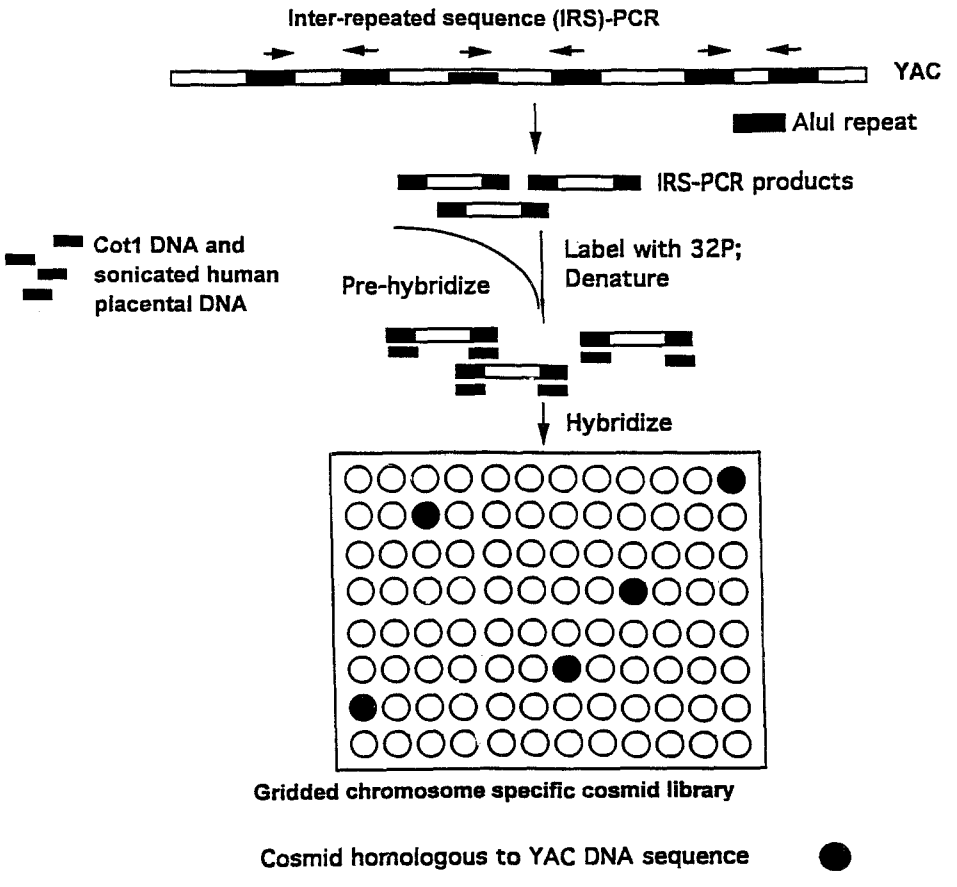

Figure 3. Conversion of a YAC to cosmids by IRS-PCR of YAC DNA, blocking of the repeats in the IRS-PCR products, and hybridization to blots of gridded colonies containing chromosome 17-specific cosmids.

genes are expressed in brain [32]. Similarly, placenta is a good source for many genes [38]. Since breast carcinomas are usually derived from epithelial cells, we are also screening a variety of epithelial cDNA libraries from nasal and tracheal epithelia, normal ovary, and ovarian carcinomas, besides a variety of libraries derived from human breast carcinoma cell lines, placenta, fibroblast, and brain.

One other issue in screening libraries, is the number of cDNAs to be screened to be fairly certain that a low abundance cDNA will be identified. For this reason, one usually screens 10^6 cDNAs derived from a specific tissue source. One way of limiting the number of cDNAs to be screened, is to begin with a "normalized" or partially normalized library, where cDNAs are equally represented in number. However, the construction of normalized libraries is still problematic, and few such libraries are available. The "direct selection by hybridization" strategy, described below, combats the problem of normalization to some degree. One final issue in screening for genes is that if the gene one is searching for is transiently expressed, the stage at which the RNA is extracted from a particular tissue is critical.

Figure 4 outlines the different approaches that are currently used in a search for cDNAs. One of the most straightforward means of identifying clones of interest in cDNA libraries is by direct hybridization of genomic probes from the region in question. Probes can be cosmids, phage or YACs. In order to eliminate non-specific detec-

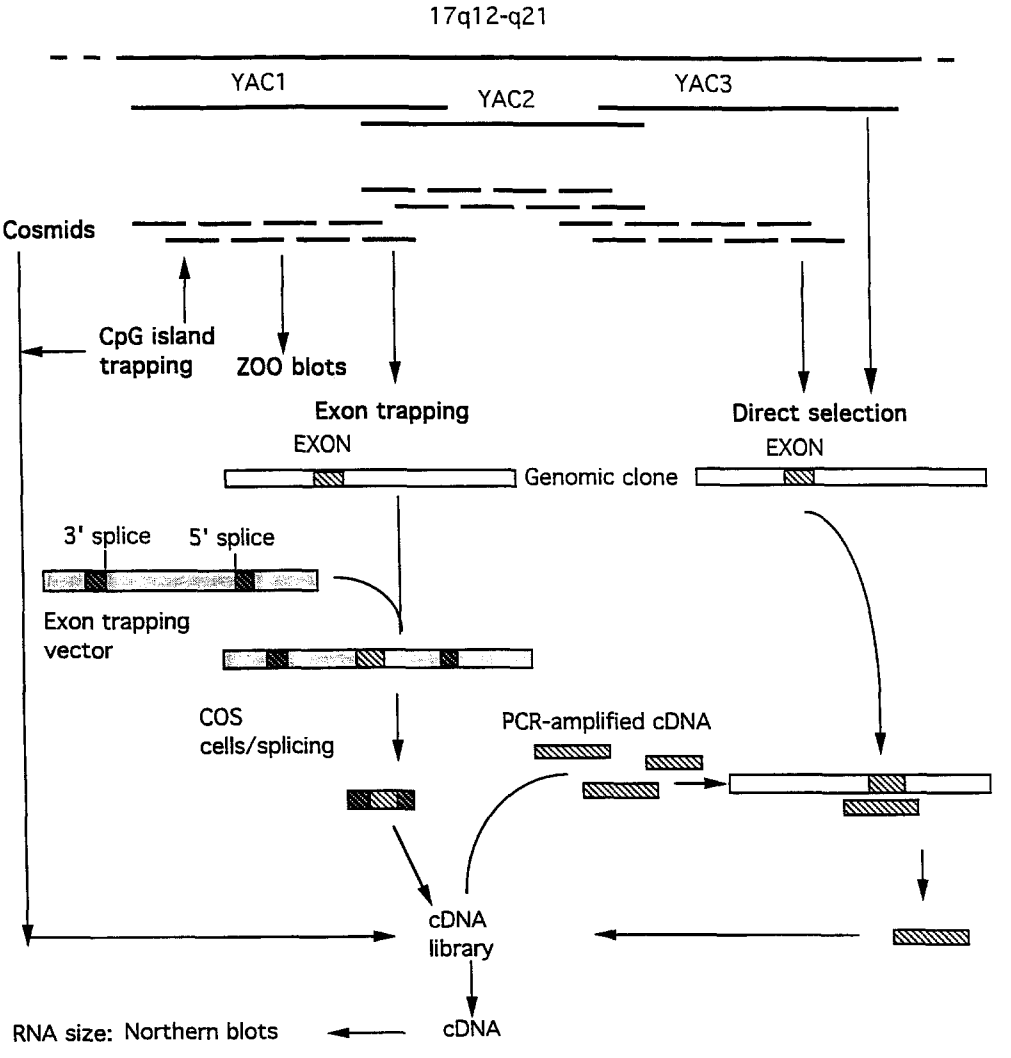

Figure 4. Strategies for the identification of genes with cloned genomic DNA from a defined region of the genome.

tion of cDNAs as a result of hybridization to human repetitive sequences present in genomic clones (approximately 1/3 of the cDNA clones in a cDNA library have AluI sequences at their 3' ends), repetitive sequences in the probe are first "blocked" by pre-hybridization with human repetitive DNA.

Cosmids are the preferred reagent in the "direct selection by hybridization" strategy. This approach, pioneered by Lovett et al [29] and Parimoo et al [42], utilizes a genomic substrate such as cosmids, phage, YACs, or even whole chromosomal DNA. After blocking the human repeats in such a substrate, it is hybridized in solution to PCR-amplified cDNA. The template/cDNA complex is then captured on magnetic beads. The complex is washed to eliminate unbound material and the bound cDNA is then eluted and detected after PCR amplification. After a second round of hybridization-selection, the cDNA species are again amplified and cloned. The resultant cDNAs are greatly enriched for those encoded by the genomic substrate.

A third approach is termed "exon trapping" and relies on capturing coding sequences from genomic DNA. The approach uses splicing

signals as the primary identifier of a potential gene. This approach was first described by Duyk et al [14] and Buckler et al [7], and relies on cloning DNA with potential coding sequences into a vector which contains an "exon trap" cassette that is spanned by a functional splice donor and/or a splice acceptor complex. Recombinant plasmids are isolated from bacteria and transfected into tissue culture cells, where the exon trap vector and insert is transcribed. If the inserted fragment contains an intron/exon boundary, splicing will occur. This can be detected by extracting cytoplasmic RNA from the transfected cells and using it as a template for cDNA synthesis and cloning. PCR based-assays can then be used to detect a novel spliced product.

One can also identify some genes in genomic DNA on the basis of regions rich in un-methylated CpG sequences [4]. These occur frequently at the 5' end of housekeeping genes and are termed HTF (HpaII tiny fragment) islands. Due to the large number of unmethylated C and G nucleotides, CpG islands are frequently recognition sites for rare-cutter enzymes such as NotI and SacI, and can be recognized as such.

Zoo blots contain DNA from a variety of animals, digested with a restriction enzyme, subjected to agarose gel electrophoresis, and Southern blotted. The rationale for hybridizing a human sub-clone to a zoo blot, is that important sequences (e.g. genes) are frequently conserved across species. This approach identifies human sub-clones that contain genes on the basis of hybridization (at reduced stringency) to DNA from other animals. These sub-clones can then be used to screen cDNA libraries.

Search for mutations

Once a set of cDNAs within the region likely to contain BRCA1 have been identified, they will be studied to see if they detect alterations in DNAs from a variety of sources: affected members in linked families, tumors from linked-family members, and sporadic tumors demonstrating loss of heterozygosity at BRCA1. This can be achieved most simply with Southern blotting and hybridization with a cDNA probe. Fairly large DNA rearrangements of greater than 500 bp may be detected in this manner. However, it may be that the mutations within BRCA1 resulting in breast and ovarian cancer predisposition are too small to detect by Southern blotting. This would be the case if they are all due to point mutations or to small insertions, deletions or other rearrangements.

One sure method of detecting mutations is by DNA sequencing. However, this can be a laborious task, particularly if the gene is very large. Sequencing can be performed on a genomic DNA template or a cDNA template prepared from RNA by reverse transcriptase. PCR allows one to sequence rapidly and directly from a complex template, but problems still remain. If one wishes to sequence a cDNA, one needs to obtain the corresponding cDNA species from RNA derived from members of linked families or from tumors. The isolation of RNA, which is particularly labile, is not always easy; in addition, the gene may not be expressed in a tissue that is easily obtained from a living individual. Although tumor tissue may be available from linked family members, BRCA1 may not be expressed in tumors if it is a tumor suppressor gene. If BRCA1 is encoded by a large gene, it may not be straightforward to identify mutations. In addition, there may be many different mutations. Precedence already exists for this in two other genes predisposing to familial cancer: RB1 and APC. BRCA1 may also resemble RB1 in having a very high spontaneous mutation rate; in the case of RB1, 30% of cases are attributed to germ-line mutations.

Indirect and rapid detection of mutations

In an attempt to detect mutations rapidly, several methods have been described: chemical cleavage [12,45], denaturing gradient electrophoresis (DGGE) and ribonuclease cleavage [33], and SSCP (single strand conformation polymorphism

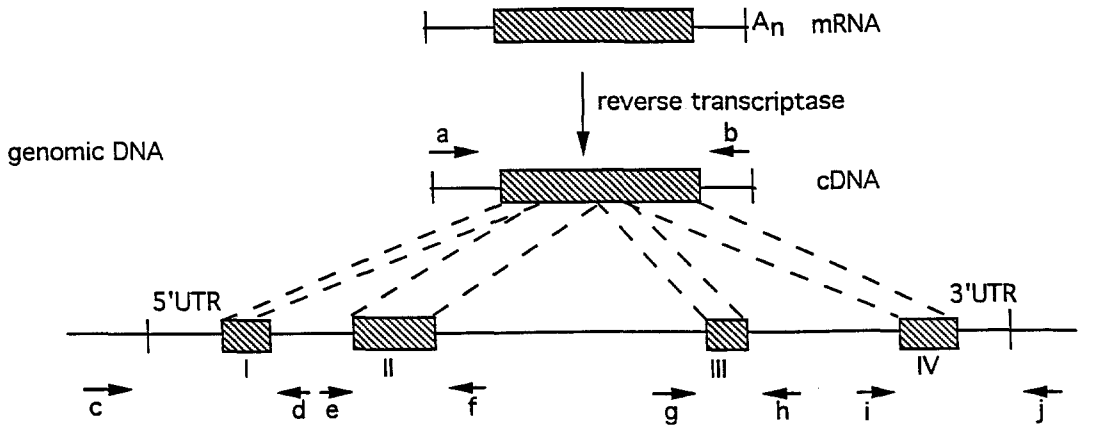

Figure 5. Comparison of the cDNA and genomic structures of a hypothetical candidate gene. In the case of the cDNA, the sequence can be analyzed after PCR amplification of the reverse transcribed product of the mRNA with primers a and b at the 5' and 3' ends of the gene. In the case of genomic DNA, a series of oligonucleotides spanning the exons would need to be designed.

[40,41]. None detect all mutations, but all have a relatively high success rate. All these methods are currently performed most successfully after PCR amplification of the region under study. Mutation scanning may be performed on cDNA derived from RNA from tumors or the appropriate tissue of breast cancer patients, or on genomic DNA, with the problems illustrated in Figure 5. Even once it is possible to amplify all exons by PCR, detection of mutations in genomic DNA may be hard if the location within the gene is variable.

Once one detects a variant cDNA, one needs to confirm that it is not a polymorphism. Determination that the cDNA is consistently altered in linked family members and in tumor tissue from affected family members would be reasonable evidence that one had isolated BRCA1. The existence of mutations in the gene in sporadic tumors would also be significant, although it is not yet known what percent of sporadic breast and ovarian tumors can be attributed to BRCA1 alterations. It is unlikely that major rearrangements of BRCA1 occur frequently, since it has not been possible to identify translocations in this region in patients. If BRCA1 is similar to other tumor suppressor genes, it is likely that a series of different mutations may predispose to tumor formation: i.e. mis-sense and nonsense mutations and small deletions. Ultimately, functional assays for BRCA1 (e.g. reversal of the transformed phenotype in breast carcinoma cell lines) will be necessary to confirm that it has been isolated. This may pave the way for better treatment of breast cancer attributable to a primary lesion at BRCA1, early detection of mutations, and possibly reversion of the malignant phenotype.

Population screening for women with germline BRCA1 mutations

Estimates of the proportion of women with germline BRCA1 mutations, who are at greatly increased risk of developing breast and/or ovarian cancer, range from 1/150 to 1/500 women. Even if the incidence were 1/500, it would still be a sufficiently high number to warrant population screens for women with germ-line BRCA1 mutations. A population screen would probably be most straightforward if it could be performed on cDNA that is easily accessible (e.g. from blood cells). If there is a common mutation, population-based screening would be facilitated. Even in the case of cystic fibrosis, where over 300 mutations in the CFTR gene have been identified, popula-

tion-based screening is made somewhat easier by the deltaF508 mutation which accounts for approximately 70% of the mutations in Northern Europeans [55].

Another problem with population-based screening is that if one does not have an RNA source, one is required to sequence genomic DNA which may be labor-intensive and therefore expensive.

It may be that population screens would be more efficiently performed via another route, such as antibody screening. If alterations at BRCA1 are shown to account for a reasonable number of sporadic tumors, additional methods may be developed to detect very early lesions in this gene in breast and ovarian cells which would permit diagnosis of very early tumors or pre-malignant lesions.

Acknowledgements

Work described here was supported in part by NIH grants CA27632 and CA60650, and a Susan G. Komen Breast Cancer Foundation Fellowship awarded to Dr. Piri Welcsh. I would like to thank Drs. Piri Welcsh (Univ. of Texas Southwestern Medical Center at Dallas) and Anthony Brown (Cornell University, New York) for critical comments on the manuscript, and Mrs. Colleen Campbell for help in preparation of the manuscript. I would also like to acknowledge the "Weekend to Wipe Out Cancer" organization of Dallas.

References

1. Anderson L, Friedman L, Osborne-Lawrence S, et al: High-density genetic map of the BRCA1 region of chromosome 17q12-q21. Genomics 17:618-641, 1993
2. Ashton-Rickardt PG, Dunlop MG, Nakamura Y, Morris RG, Purdie CA, Steel CM, Evans HJ, Bird CC, Wyllie AH: High frequency of APC loss in sporadic colorectal carcinoma due to breaks clustered in 5q21-22. Oncogene 4:1169-1174, 1989
3. Bates G, Valdes J, Hummerich H, Baxendale S, Le Paslier D, Monaco A, Tagle D, MacDonald M, Altherr M, Ross M, Brownstein B, Bentley D, Wasmuth J, Gusella J, Cohen D, Collins F, Lehrach H: Characterization of a yeast artificial chromosome contig spanning the Huntington's disease gene candidate region. Nature Genet 1:180-187, 1993
4. Bird A: CpG islands as gene markers in the vertebrate nucleus. Trends Genet 3:342-347, 1987
5. Bishop J, Morton J, Rosebach M, Richardson M: Three abundance classes in HeLa cell messenger RNA. Nature 250:199-204, 1974
6. Bowcock A, Anderson L, Black D, Osborne-Lawrence S, Hall J, King M-C: THRA1 and D17S183 flank an interval of < 4cM for the breast-ovarian cancer gene (BRCA1) on chromosome 17q21. Am J Hum Genet 52:718-722, 1993
7. Buckler A, Chang D, Graw S, et al: Exon amplification: A strategy to isolate mammalian genes based on RNA splicing. Proc Natl Acad Sci USA 88:4005-4009, 1991.
8. Bystrom C, Larsson C, Blomberg C, Sandelin K, Falkmer U, Skogseid B, Oberg K, Werner S, Nordenskjold M: Localization of the MEN1 gene to a small region within chromosome 11q13 by deletion mapping in tumors. Proc Natl Acad Sci USA 87:1968-1972, 1990
9. Chumakov IF, Rigault P, Guillou S, Ougen P, Billaut A, Guasconi G, et al: Continuum of overlapping clones spanning the entire human chromosome 21q. Nature 359:380-387, 1992
10. Claus EB, Risch N, Thompson WD: Genetic analysis of breast cancer in the cancer and steroid hormone study. Am J Hum Genet 48:232-242, 1991
11. Cohen AJ, Li FP, Berg S, Marchetto DJ, Tsai S, Jacobs SC, Brown RS: Hereditary renal-cell carcinoma associated with a chromosomal translocation. N Engl J Med 301:592-595, 1979
12. Cotton RG, Rodrigues NR, Campbell RD: Reactivity of cytosine and thymine in single-base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations. Proc Natl Acad Sci USA 85:4397-4401, 1988
13. Cox DR, Burmeister M, Price ER, Kim S, Myers RM: Radiation hybrid mapping: a somatic cell genetic method for constructing high resolution maps of mammalian chromosomes. Science 250:245-250, 1990
14. Duyk G, Kim S, Myers R, Cox D: Exon trapping: A genetic screen to identify candidate transcribed sequences in cloned mammalian genomic DNA. Proc Natl Acad Sci USA 87:8995-8999, 1990
15. Easton DF, Bishop T, Ford D, Crockford GP: The Breast Cancer Linkage Consortium: Genetic linkage analysis in familial breast and ovarian cancer: Results from 214 families. Am J Hum Genet 52:678-701, 1993
16. Fontaine B, Sanson M, Delattre O, Menon AG, Rouleau

GA, Seizinger BR, Jewell AF, Hanson MP, Aurias A, Martuza RL, et al: Parental origin of chromosome 22 loss in sporadic and NF2 neuromas. Genomics 10:280-283, 1991

17. Fearon ER, Cho KR, Nigro JM, Kern SE, Simons JW, Ruppert JM, Hamilton SR, Preisinger AC, Thomas G, Kinzler KW, Vogelstein B: Identification of a chromosome 18q gene that is altered in colorectal cancers. Science 247:49-56, 1990
18. Futreal PA, Söderqvist P, Marks JR, Iglehart JD, Cochran C, Barrett JC, Wiseman RW: Detection of frequent allelic loss on proximal chromosome 17q in sporadic breast carcinoma using microsatellite length polymorphisms. Cancer Res 52:2624-2627, 1992
19. Galau G, Klein W, Britten R, Davidson E: Significance of rare mRNA sequences in liver. Arch Biochem Biophys 179:584-599, 1977
20. Green ED, Olson MV: Systematic screening of yeast artificial-chromosome libraries by use of the polymerase chain reaction. Proc Natl Acad Sci USA 87:1213-1217, 1990
21. Groden J, Thliveris A, Samowitz W, Carlson M, Gelbert L, Albertsen H, Yoslyn G, Stevens J, Spirio L, Robertson M, Sargeant L, Krapcho K, Wolff E, Burt R, Hughes JP, Warrington J, McPherson J, Wasmuth J, LePaslier D, Abderrahim H, Cohen D, Leppert M, White R: Identification and characterization of the familial adenomatous polyposis coli gene. Cell 66:589-600, 1991
22. Hall JM, Lee MK, Newman B, et al: Linkage of early-onset familial breast cancer to chromosome 17q21. Science 250:1684-1689, 1990
23. Herrera L: Gardner's syndrome in a man with an interstitial deletion of 5q. Am J Med Genet 25:473-476, 1986
24. Huntington's Disease Collaborative Research Group: A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease. Cell 72:971-983, 1993
25. Jelinek WR: Ubiquitous interspersed repeated sequences in mammalian genomes. Proc Natl Acad Sci USA 77:1398-1402, 1980
26. Kelsey JL: A review of the epidemiology of human breast cancer. Epidemiol Rev 1:74-109, 1979
27. Kovacs G, Szucs S, De Riese W, Baumgartel H: Specific chromosome aberration in human renal cell carcinoma. Cancer Res 40:171-178, 1987
28. Ledbetter SA, Nelson DL, Warren ST, Ledbetter DH: Rapid isolation of DNA probes within specific chromosome regions by interspersed repetitive sequence polymerase chain reaction. Genomics 6:475-481, 1990
29. Lovett M, Kere J, Hinton L: Direct selection: A method for the isolation of cDNAs encoded by large genomic regions. Proc Natl Acad Sci USA 88:9628-9633, 1991
30. Malkin D, Li FP, Strong LC, et al: Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. Science 250:1233-1238, 1990
31. Margaritte P, Bonaiti-Pellie C, King M-C, Clerget-Darpoux F: Linkage of familial breast cancer to chromosome 17q21 may not be restricted to early-onset disease: Am J Hum Genet 50:1231-1234, 1992
32. Milner R, Lai C, Lenoir D, Nave K, Bakhit C, Malfroy B: Brain-specific gene expression. Biochem Soc Symp 52:107-117, 1986
33. Myers RM, Sheffield VC, Cox DR: Detection of single base changes in DNA: ribonuclease cleavage and denaturing gradient gel electrophoresis. *In:* Davies K (ed) Genome Analysis: A Practical approach. IRL Press, 1988, pp 95-130
34. Narod S, Feunteun J, Lynch H, et al: Familial breast-ovarian cancer locus on chromosome 17q12-q23. Lancet 338:82-83, 1991
35. Newman B, Austin M, Lee M, King M: Inheritance of human breast cancer: evidence for autosomal dominant transmission in high-risk families. Proc Natl Acad Sci USA 85:3044-3048, 1988
36. Nishisho I, Nakamura Y, Miyoshi Y, Miki Y, Ando H, Horii A, Koyama K, Utsunomiya J, Baba S, Hedge P, Markham A, Krush AJ, Petersen G, Hamilton SR, Nilbert MC, Levy DB, Bryan TM, Preisinger AC, Smith KJ, Su L, Kinzler KW, Vogelstein B: Mutations of chromosome 5q21 genes in FAP patients and colorectal cancer patients. Science 253:665-669, 1991
37. Olson M, Hood L, Cantor C, Botstein D: A common language for physical mapping of the human genome. Science 245:1434-1435, 1989
38. Ohlsson R: Growth factors, proto-oncogenes and human placental development. Cell Diff Dev 28:1-15, 1989
39. Okubo K, Hori N, Matoba R, et al: Large-scale cDNA mapping for analysis of quantitative and qualitative aspects of gene-expression. Nature Genet 2:173-179, 1992
40. Orita M, Iwahana H, Kanazawa H, Hayashi K, Sekiya T: Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. Proc Natl Acad Sci USA 86:2766-2770, 1989
41. Orita M, Suzuki Y, Sekiya T, Hayashi K: Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. Genomics 5:874-879, 1989
42. Parimoo S, Patanjali S, Shukla H, Chaplin D, Weismann S: cDNA selection: efficient PCR approach for the selection of cDNAs encoded in large chromosomal DNA fragments. Proc Natl Acad Sci USA 88:9623-

9627, 1991
43. Pathak S, Strong LC, Ferrell RE, Trindade A: Familial renal cell carcinoma with a 3;11 chromosome translocation limited to tumor cells. Science 217:939-941, 1982
44. Ponder BAH: Gene losses in human tumours. Nature 335:400-401, 1988
45. Rodrigues NR, Rowan A, Smith ME, Kerr IB, Bodmer WF, Gannon JV, Lane DP: p53 mutations in colorectal cancer. Proc Natl Acad Sci USA 87:7555-7559, 1990
46. Rommens JM, Iannuzzi MC, Kerem B, Drunn ML, Melmer G, Dean M, Rozmahel R, Cole JL, Kennedy D, Hidaka N, Zsiga M, Buchwald M, Riordan JR, Tsui L-C, Collins FS: Identification of the cystic fibrosis gene: chromosome walking and jumping. Science 245: 1059-1065, 1989
47. Sato T, Tanigami A, Yamakawa K, Akiyama F, Kasumi F, Sakamoto G, Nakamura Y: Allelotype of breast cancer: cumulative allele losses promote tumor progression in primary breast cancer. Cancer Res 50:7184-7189, 1990
48. Sato T, Akiyama F, Sakamoto G, Kasumi F, Nakamura Y: Accumulation of genetic alterations and progression of primary breast cancer. Cancer Res 51:5794-5799, 1991
49. Saito H, Inazawa J, Saito S, Kasumi F, Koi S, Sagae S, Kudo R, Saito J, Noda K, Nakamura Y: Detailed deletion mapping of chromosome 17q in ovarian and breast cancers: 2-cM region on 17q21.3 often and commonly deleted in tumors. Cancer Res 53:3382-3385, 1993
50. Schwartz DC, Cantor CR: Separation of yeast chromosome-sized DNAs by pulsed-field gradient gel electrophoresis. Cell 37:67-75, 1984
51. Seizinger BR, Smith DI, Filling-Katz MR, Neumann H, Green JS, Choyke PL, Anderson KM, Freiman RN, Whaley J, et al: Genetic flanking markers refine diagnostic criteria and provide insights into the genetics of Von Hippel Lindau disease. Proc Natl Acad Sci USA 88:2864-2868, 1991
52. Simard J, Feunteun J, Lenoir G, Tonin P, Normand T, Luu The V, Vivier A, Lasko D, Morgan K, Rouleau GA, Lynch H, Labrie F, Narod SA: Genetic mapping of the breast-ovarian cancer syndrome to a small interval on chromosome 17q12-21: exclusion of candidate genes EDH17B2 and RARA. Hum Molec Genet 2:1193-1199, 1993
53. Slamon DJ, Godolphin W, Jones LA, et al: Studies of the HER-2/*neu* oncogene. Science 235:177-182, 1989
54. Smith S, Easton D, Evans D, Bonder B: Allele losses in the region 17q12-q21 in familial breast and ovarian cancer involve the wild-type chromosome. Nature Genet 2:128-131, 1992
55. Tsui LC: The spectrum of cystic fibrosis mutations [review]. Trends in Genetics 8:392-398, 1992
56. Trask BJ: Fluorescence *in situ* hybridization: applications in cytogenetics and gene mapping. Trends in Genetics 7:149-154, 1991