

Construction of a Transcription Map Surrounding the *BRCA1* Locus of Human Chromosome 17

LAWRENCE C. BRODY,^{*,1} KENNETH J. ABEL,^{†,‡,1} LUCIO H. CASTILLA,^{*,1} FERGUS J. COUCH,^{§,1}
 DAWN R. MCKINLEY,^{*} GUIYING YIN,[†] PEGGY P. HO,[¶] SOFIA MERAJVER,[¶]
 SETTARA C. CHANDRASEKHARAPPA,^{*} JUNZHE XU,[¶] JEFFERY L. COLE,[†] JEFFERY P. STRUEWING,^{||}
 JOHN M. VALDES,[†] FRANCIS S. COLLINS,^{*,2} AND BARBARA L. WEBER[§]

^{*}The National Center for Human Genome Research, ^{||}Genetic Epidemiology Branch, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892; [¶]Departments of [¶]Medicine and [†]Human Genetics and [‡]The Human Genome Center, University of Michigan, Ann Arbor, Michigan 48109; and [§]Departments of Internal Medicine and Genetics, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Received October 24, 1994; revised November 28, 1994

We have used a combination of methods (exon amplification, direct selection, direct screening, evolutionary conservation, island rescue-PCR, and direct sequence analysis) to survey approximately 600 kb of genomic DNA surrounding the *BRCA1* gene for transcribed sequences. We have cloned a set of fragments representing at least 26 genes. The DNA sequence of these clones reveals that 5 are previously cloned genes; the precise chromosomal location of 2 was previously unknown, and 3 have been cloned and mapped by others to this interval. Three other genes, including *BRCA1* itself, have recently been mapped independently to this region. Sequences from 11 genes are similar but not identical matches to known genes; 5 of these appear to be the human homologues of genes cloned from other species. Another 7 genes have no similarity with known genes. In addition, 39 putative exons and 14 expressed sequence tags have been identified and mapped to individual cosmids. This transcript map provides a detailed description of gene organization for this region of the genome. © 1995 Academic Press, Inc.

INTRODUCTION

The generation of transcript maps is an important goal of the human genome project (Collins and Galas, 1993). In addition to detailed genetic and physical maps, a map of transcribed sequences is necessary for understanding genome structure and will aid greatly in the isolation of the genes for inherited traits. As part of our effort to isolate the breast cancer susceptibility gene, *BRCA1*, we have constructed a transcript map

covering approximately 1 Mb of chromosome 17q21. The mapping of the *BRCA1* gene to chromosome 17q (Hall *et al.*, 1990; Narod *et al.*, 1991) and further refinement of its location by genetic recombination allowed a minimal candidate interval to be defined (Bowcock *et al.*, 1993; Easton *et al.*, 1993; Chamberlain *et al.*, 1993; Simard *et al.*, 1993; Smith *et al.*, 1994). This interval spans approximately 1 cM of chromosome 17. Since on average 1 cM equals approximately 1 Mb, this amount of DNA should be amenable to positional cloning efforts (Collins, 1992). The first step of such an effort is construction of a physical contig of DNA included in the interval. We and others (Albertsen *et al.*, 1994) have constructed physical maps of this candidate interval, and the accompanying paper (Couch *et al.*, 1995) presents a yeast artificial chromosome (YAC)-, P1-, and cosmid-based contig of this region. Although by its nature a positional cloning approach is targeted toward the isolation of a single gene, a successful outcome in many cases depends upon the isolation of most of the genes contained in the candidate interval.

There are several methods available for isolating transcribed sequences from large genomic regions. Exon amplification/exon trapping (Duyk *et al.*, 1990; Buckler *et al.*, 1991; Krizman and Berget, 1993), direct selection (Parimoo *et al.*, 1991; Lovett *et al.*, 1991), direct screening (Wallace *et al.*, 1990), evolutionary conservation, genomic sequencing (Sulston *et al.*, 1992; Oliver *et al.*, 1992), and CpG island-based (Patel *et al.*, 1991; Valdes *et al.*, 1994) methods have all been used to identify the transcribed portions of genomic DNA. The advantages and disadvantages of these methods have been described (Collins, 1992; Weber *et al.*, 1994). No single method has been shown to be capable of isolating 100% of the transcripts from a given genomic interval of this size. For this reason our collaborative multi-institutional effort employed six methods of transcript identification in parallel. Exon amplification, direct selection, and direct screening were performed us-

¹ Contributed equally to this work.

² To whom correspondence should be addressed at The National Center for Human Genome Research, National Institutes of Health, Building 49, Room 3A18, 9000 Rockville Pike, Bethesda, Maryland 20892. Telephone: (301) 496-0844. Fax: (301) 402-4929.

ing cosmids representing the majority of our contig. Transcript isolation by island rescue PCR and evolutionary conservation was carried out on a subset of these. In the accompanying papers, Osborne-Lawrence *et al.* (1995) and Friedman *et al.* (1995) report on extensive direct selection experiments and YAC-based direct cDNA library screening, respectively. These studies utilized the same collection of genomic clones as were used in this study (Couch *et al.*, 1995). Therefore, some conclusions can be drawn as to the specificity and sensitivity of these methods as the result of this comparison.

While recognizing that these methods can be applied to either YACs or cosmids, we have derived the majority of the data described here from cosmid material due to its relative ease of manipulation. We present here a transcript map of the BRCA1 region. In the interval bounded by the genes *EDH17B2* (17 β -hydroxysteroid dehydrogenase II) and *PPY* (pancreatic polypeptide), we have isolated 26 genes. Of these, 19 have similarities or identities to known genes, while 7 appear to be novel genes. In addition, 53 unique exons and cDNA fragments (expressed sequence tags, ESTs) have been isolated and mapped to specific cosmids in the contig.

MATERIALS AND METHODS

Molecular biology techniques. Except as noted, all molecular manipulations (cloning, phage library screening, PCR) were performed using standard methods (Sambrook *et al.*, 1989).

Cosmids and YACs. The cloning and construction of the *BRCA1*-region cosmid, P1, and YAC contig is described by Couch *et al.* (1995). Well locations of the cosmids presented here have been maintained from the original library provided by L. Deaven (Los Alamos National Laboratory).

cDNA libraries. A human breast (λ gt11 vector, oligo(dT)-primed) cDNA library was constructed at the University of Michigan Genome Center (Swaroop and Xu, 1993). Human fetal brain (Stratagene, LambdaZAPII vector, random and dT primed), normal human ovary (Stratagene, LambdaUNI.ZAPXR, dT primed), and normal human breast (Clontech, λ gt10, random primed) cDNA libraries were purchased. A human endothelial cell cDNA library (λ gt11, random and dT primed) was a gift from D. Ginsburg. For library screening 30,000 to 50,000 plaques were plated per 150-mm dish and transferred to Hybond-N (Amersham) membranes. Three to five hundred thousand plaque forming units were screened with each probe.

Chromosome mapping panels. The chromosomal location of all clones was verified by at least one of two methods. The first was Southern blotting to a somatic cell hybrid mapping panel containing the following human-rodent somatic cell hybrid lines: 7AE4 (containing an intact human chromosome 17; Leach *et al.*, 1989), 12.3B5 (17p-q12, GM10659), MH41 (17q23.2-qter, GM10502), and L(17n)C (chromosome 17q; Leach *et al.*, 1989). An appropriate hybridization pattern allowed assignment to human 17q21-q23. The second method employed a screening Southern blot containing two samples (plus a molecular weight standard). Lane 1 contained total human genomic DNA, and lane 2 contained DNA from a pool of 400 cosmids from the region. The cosmid sample was "normalized" for copy number to the genomic samples (10 μ g/lane) by loading less cosmid pool DNA (10 ng/lane). These samples were restricted with *Pst*I, and multiple "two-lane" blots were produced. The cosmid lane served to represent the genomic interval of interest. Clones were said to map to the region if equivalent hybridization bands were observed in the cosmid and human genomic lanes. Exons and cDNA clones were then mapped to individual cosmids by colony hybridization to cosmid arrays as previously described (Couch *et al.*, 1995).

Exon amplification. Exon amplification was performed essen-

tially as previously described (Buckler *et al.*, 1991) using the modified cloning vector pSPL3 (Church *et al.*, 1994). For a detailed description and discussion of the exon amplification methods, cosmids employed, and characterization of isolated exons, see Abel *et al.* (1994). Briefly, pools of 6-10 cosmids were used for each exon amplification reaction. These pools included many redundant cosmids not shown in the minimal set (Couch *et al.*, 1995, Fig. 1). Plasmid clones containing putative exons derived from these experiments were isolated to microtiter dishes to be further characterized. Filter replicas of these dishes were then hybridized sequentially with human Cot1 DNA, the amplification vector, and PCR-amplified inserts of individual clones. These hybridizations allowed the identification of clones containing human reiterated sequences, cloning artifacts, and redundant exons, respectively. The resulting set of unique clones is shown in Fig. 1. Clones containing trapped exons were mapped as described above and used as probes to screen cDNA libraries.

Direct selection. Direct selection was carried out using the magnetic bead capture protocol as described (Tagle *et al.*, 1993; Couch *et al.*, 1994). Cosmid DNAs spanning the region between cosmids 96H6 and 119D4 (Fig. 1) were placed in two pools (seven cosmids per pool) and digested with *Sau*3AI or *Alu*I. After linker addition and amplification with biotinylated primers, the cosmid DNA was hybridized in solution to an excess of amplified breast cDNA library inserts. Specifically bound material was eluted at high temperature and cloned into plasmid vectors. Inserts from these clones were then analyzed as outlined above.

Direct screening. Cosmids were digested with the restriction enzyme *Not*I, and inserts were isolated from low-melting-temperature agarose (SeaPlaque, FMC) and random-prime labeled individually using standard methods (Feinberg and Vogelstein, 1983). The labeled products of up to five cosmids were then pooled. Pooled probes were mixed with 150 μ g of sheared human placental and 300 μ g of sheared salmon sperm DNAs in TE (500 μ l). After boiling (10 min), the DNAs were preannealed at 65°C for 1 h. Prehybridization and hybridization of cDNA library filters was carried out in 7% SDS, 0.25 M Na₂HPO₄, 1 mM EDTA, and 1% bovine serum albumin. Filters were washed in 2 \times SSC, 1% SDS at 65°C (2 \times , 30 min) prior to exposing to X-ray film. After a 24-h exposure, primary plaque signal intensity varied from just over background to several-fold over background.

Evolutionary conservation and island rescue PCR. Cosmid fragments containing single-copy DNA were isolated by subcloning restriction fragments that failed to hybridize with labeled total human genomic DNA. These subclones were used to probe Southern blots containing human, gorilla, rodent (mouse, hamster, and rat), and *Torpedo californica* (Pacific electric ray) DNA. Probes hybridizing to humans and other species were sequenced.

The generation of island rescue PCR (IR-PCR) probes from the 300-kb, nonchimeric YAC, 26F3, has been previously reported (Valdes *et al.*, 1994).

Sequence analysis. Double-stranded DNA sequencing was performed on an automated sequencer (Applied Biosystems or Pharmacia) using fluorescence methodology as recommended by the manufacturer. The BLASTN (Altschul *et al.*, 1990) and BLASTX (Gish and States, 1993) programs were used to search GenBank (National Center for Biotechnology Information) with sequence from putative transcribed clones. Gene Recognition and Analysis Internet Link (GRAIL, Uberbacher and Mural, 1991) program version 1.2 was accessed through the XGRAIL client (Shah *et al.*, 1994) on a Sun Sparc10 workstation. GRAIL 2 analysis was performed on the 21 kb of genomic sequence from the *EDH17B2* locus. This sequence was deposited in GenBank by Peltoketo *et al.* (1988) (Accession No. M84472). The sequences of clones reported in Table 1 have been deposited with GenBank.

RESULTS AND DISCUSSION

A map of transcripts is shown in Fig. 1. More detailed information on individual transcripts is listed in Table 1. For orientation purposes, the clones are listed in centromere to telomere order. The transcripts are de-

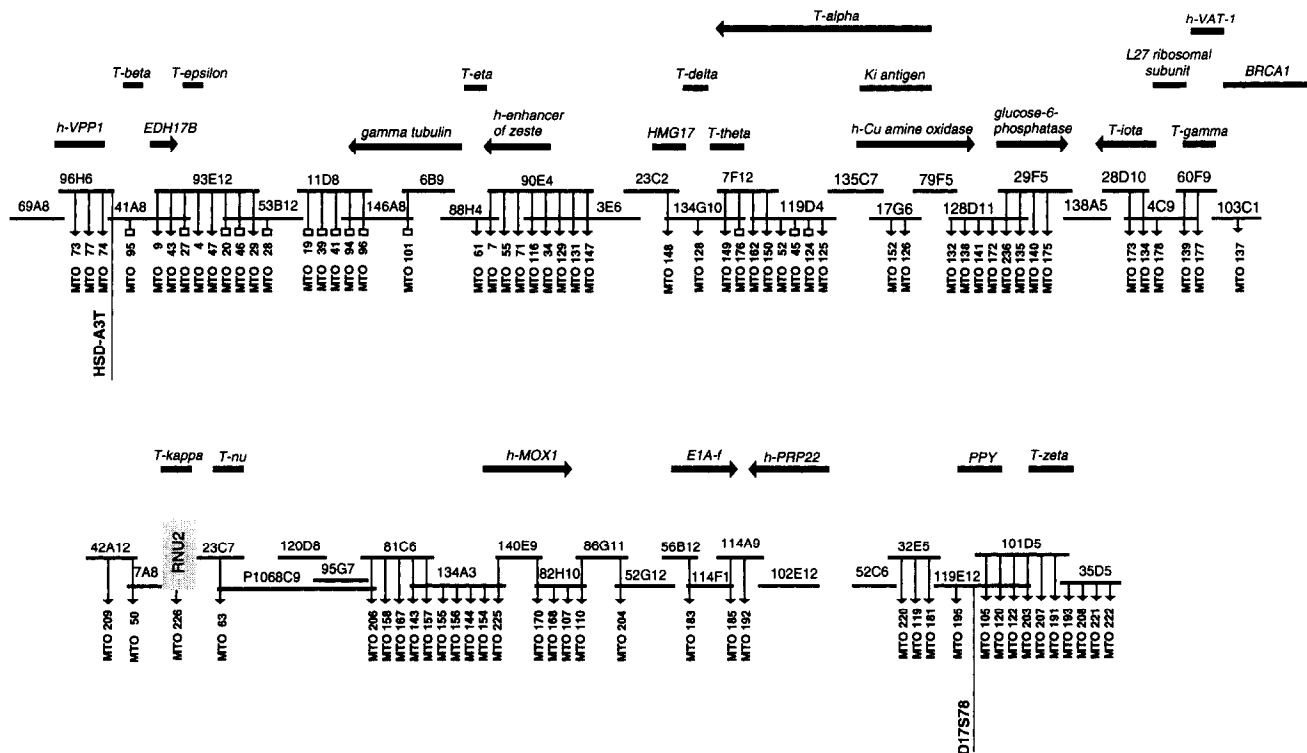


FIG. 1. Schematic diagram of the transcript map in the *BRCA1* region between markers HSD-A3T (Friedman *et al.*, 1993) and the *PPY* gene. The minimal cosmid contig is represented by thin horizontal bars (Couch *et al.*, 1995) from centromeric end (**top left**) to telomeric end (**bottom right**). Cloned transcripts are represented by thick bars. Arrows at the ends of transcripts indicate the direction of transcription, if known. Previously unidentified genes are named with the prefix "T" (transcript) and a Greek letter. Human homologues of genes cloned from other species are named with the prefix "h" (human). Downward vertical thin arrows indicate the map location of trapped exons, and vertical bars with squares direct selected cDNA clones. The RNU2 region is represented by a gray box (see text). Cosmids, genes, and exons are not drawn to scale. Genes are drawn to represent accurately the hybridization of cDNA clones to specific cosmids. The overlapping cosmid and P1 contig contains two gaps, one at the break of the top right edge and one between cosmids 102E12 and 52C6 on the bottom.

scribed below, first in terms of isolation methods and then by sequence information.

Evaluation of Transcription Isolation Methods

Exon amplification. Cosmid DNAs representing 39 of the 45 cosmids in our contig were used as templates for exon amplification. Two exon amplification experiments were performed: the first using unordered cosmid pools and the second with cosmids pooled by their contig location (Abel *et al.*, 1994). Over 2000 primary clones were isolated and arrayed in microtiter dishes. From this starting material a nonredundant collection of exons was derived. This set contained a total of 76 exons that mapped to the region of interest (Fig. 1). Exon amplification was applied more extensively than any other method presented here and proved highly effective. Putative exons were isolated from almost every cosmid tested. Trapped exons were isolated as plasmid clones that can be readily sequenced, thus allowing rapid comparison to nucleotide databases. Because they are directionally cloned, the orientation of transcription can be determined from this sequence. An added advantage of this method is the ease of determining intron-exon borders by comparing the sequence of trapped exons to cDNA sequences. In several cases, multiple unique exon clones were shown to be con-

tained in a single cDNA (E1A-f, glucose-6-phosphatase, human Enhancer of zeste, and others). In this study, the major disadvantages of transcript isolation by exon amplification was the small size of some of the trapped exons. Although the average trapped exon was 200 bp, the distribution was clearly skewed toward the 100- to 150-bp range (median = 130; Abel *et al.*, 1994). In the case of the *BRCA1* gene, we isolated clones corresponding to exons 20 (84 bp) and 22 (74 bp) (Miki *et al.*, 1994). Clones of this size are difficult to use as hybridization probes for screening cDNA libraries using conventional methods.

Direct selection. A total of 14 cosmids were digested with *Sau3AI* and *AluI* and hybridized to PCR-amplified inserts from a breast cDNA library for the direct selection of clones. Of 32 clones analyzed from one experiment, 23 (72%) mapped to the region of interest of chromosome 17, although more than half (14) of these contained *Alu* sequences in addition to single-copy sequence (Couch *et al.*, 1995). Due to the difficulty of using these clones in hybridization experiments, they were not characterized further. This set of clones identified 2 genes that were not isolated by other methods (Table 1).

Direct screening. The inserts of 25 cosmids represented in the minimal contig were used to probe the

TABLE 1
Summary of Isolated BRCA1-Region Clones

Map feature	Clone name	Initial isolation method	Accession No.	Gene name	Database homology	Accession No. of best match	mRNA size
ExST	MTO-73	A	T27155				
Gene	MTO-77	A	T27156	<i>VPP1</i>	<i>Bos taurus</i> VPP1	L31770	
ExST	MTO-74	A	T27157				
EST	MTO-95	B				M84472	
Gene	MTO-92	E		T-beta		M84472	
Gene	MTOs-9+, 43	A			EDH17B2	M84472	
EST	MTO-27	B				M84472	
Gene	MTOs-4, 14, 20, 46, 47	A, B, C	T27158-62	T-epsilon	EST NIB800	T17366	4.4
EST	MTO-29	B	T27163				
EST	MTO-28	B	T27164				
EST	MTO-19	B	T27165				
EST	MTO-39*	B	T27166				
EST	MTO-41*	B	T27167				
EST	MTO-94	B	T27168				
EST	MTO-96	B	T27169				
Gene	MTO-101	B			γ -Tubulin	M61764	
Gene	MTO-2, 61	A, F	T27170	T-eta	Rat neurexin	S27887	5.0
Gene	MTOs-1, 34, 55, 71, 116	A, F		enhancer of zeste	Drosophila enhancer of zeste	U00180	4.6
ExST	MTOs-129, 131	A	T27171, T27223				
ExST	MTO-147	A	T27172				
Gene	MTOs-148, 242	A			HMG17	M26676	
Gene	MTO-128	A	T27173	T-delta			
ExST	MTO-149		T27174				
Gene	MTO-176	B	T27175-6	T-theta	EST HIBAB96	T08097	1.2
ExST	MTO-162	A	T27177				
Gene	MTOs-52, 150	A	T27178-9	T-alpha			2.5
ExST	MTO-45	A	T27180				
EST	MTO-124	B	T27181				
EST	MTO-125	B	T27182				
ExST	MTO-152	A	T27183				
Gene	MTO-126	A			Ki Antigen	A60537	2.8
Gene	MTO-98	C		Cu amine oxidase	<i>Bos taurus</i> Cu amine oxidase	S34657	2.5
ExST	MTO-132	A	T27184				
ExST	MTO-138	A	T27185				
ExST	MTO-141	A	T27186				
ExST	MTO-172	A	T27187				
ExST	MTO-236	A					
Gene	MTOs-108, 135+, 140, 175	A, D			Human glucose-6-phosphatase	U01120	
Gene	MTOs-134, 173	A	T27188-9	T-iota	Human p23	L24998	7.4
ExST	MTO-178	A	T27190				
Gene	MTO-257	C		h-L27 ribosomal	L27-ribosomal	L19527	
Gene	MTO-139	A	T27191	T-gamma			3.8
Gene	MTO-258	C		VAT-1	VAT-1	T05306	
ExST	MTO-177*	A	T27192				
Gene	MTOs-133, 137	A			<i>BRCA1</i>	U14680	
ExST	MTO-209	A	T27193				
EST	MTO-50	A	T27194				
Gene	MTO-226	A	T27195	T-kappa	Flt4-kinase	X69878	
Gene	MTO-63	A	T27196	T-nu			
ExST	MTO-206	A	T27197				
ExST	MTO-158	A	T27198				
EST	MTO-167	A	T27199				
ExST	MTO-143	A	T27200				
ExST	MTO-157	A	T27201				
ExST	MTO-155	A	T27202				
ExST	MTO-156	A	T27203				
ExST	MTO-144	A	T27204				
ExST	MTO-154	A	T27205				
ExST	MTO-225	A	T27206				
Gene	MTOs-170, 228	A, D		<i>MOX1</i>	MOX1	Z15103	

TABLE 1—Continued

Map feature	Clone name	Initial isolation method	Accession No.	Gene name	Database homology	Accession No. of best match	mRNA size
ExST	MTO-168	A	T27207				
ExST	MTO-107	A	T27208				
ExST	MTO-110	A	T27209				
ExST	MTO-204	A	T27210				
Gene	MTOs-183, 185	A			E1A-f	D12765	
Gene	MTO-192	A		<i>PRP22</i>	Yeast PRP-22	Z29560	
ExST	MTO-220	A	T27211				
Gene	MTO-119+	A			PPY	M11726	
ExST	MTO-181	A	T27212				
ExST	MTO-195	A	T27213				
ExST	MTO-105	A	T27214				
Gene	MTOs-120, 122	A	T27215-6	T-zeta	Xenopus FGF3	S25713	
ExST	MTO-203	A	T27217				
ExST	MTO-207	A	T27218				
ExST	MTO-191	A					
ExST	MTO-193	A	T27219				
ExST	MTO-208	A	T27220				
ExST	MTO-221	A	T27221				
ExST	MTO-222	A	T27222				

Note. Rows containing map features are in centromere to telomere order as in Fig. 1. Clones marked with an asterisk contain repetitive and single-copy sequences (see text). A plus sign is used to denote exon clones that have been trapped from the minus strand of known genes (see text). Initial isolation methods are as follows: A, exon amplification; B, direct selection; C, direct screening; D, evolutionary conservation; E, GRAIL; F, island rescue PCR. Accession numbers are not given when the sequence of clones presented here was identical to those in GenBank. Likewise, a name does not appear in the gene name column if the nucleotide sequence of a cloned fragment was identical to the sequence of the database entry. Blank entries in the database homology column indicate that no significant matches were found. The accession number of the best GenBank match is shown. In many cases additional GenBank entries scored nearly as high as the best match. mRNA size is shown if the isolated clones were used to probe Northern blots. Blank entries indicate that a Northern blot was not performed.

oligo(dT)-primed, normal breast cDNA library. These screens resulted in the isolation of four genes. Direct hybridization methods are often complicated by the isolation of transcribed repeats. In this study one cDNA clone, MTO-14, was found to contain a repetitive sequence and adjacent unique sequences. These unique sequences hybridized specifically to the 93E12 and 41A8 cosmids. In general, direct library screening yielded few genes per pooled cosmid probe. The major disadvantage of this method was its low efficiency of transcript isolation. Its simplicity and the fact that it isolates intact cDNA clones is one advantage of the direct screening method over those described above.

Island rescue and evolutionary conservation. Although these two methods were only applied to three cosmids and one YAC, this application did yield four genes. Island rescue was performed on the 26F3 YAC, resulting in the isolation of two genes from the human endothelial cell cDNA library (Table 1).

Clones containing evolutionarily conserved sequences were isolated as a by-product of experiments aimed at cloning of "single-copy" DNA fragments from the region. In the process of mapping these clones to the chromosome 17 somatic cell hybrid panel, signals were noted in lanes containing the DNA of other vertebrates (Fig. 2A). Subsequent sequence analysis of these clones revealed that two of three clones had significant similarity to sequences in GenBank. A subclone from cosmid 29F5 contains a portion of the glucose-6-phos-

phatase gene (Fig. 3), and the subclone spanning cosmids 134A3, 140E9, and 86G11 contains a portion of the *MOX1* gene (data not shown).

Sequence-based exon prediction. GRAIL analysis of the published 21,788-bp region of genomic DNA containing the *EDH17B1* and *EDH17B2* genes predicted the presence of a previously undescribed transcription unit proximal to the *EDH17B1* gene (Fig. 4). Three putative exons were amplified from genomic DNA and used to screen a breast cDNA library, resulting in the isolation of several cDNA clones containing the predicted exons spliced together.

Efficiency of transcript isolation. The goal of this study was to isolate the maximum number of transcripts. Transcript identification was carried out as a dynamic process. Regions of the contig in which transcripts were efficiently isolated by the first method applied were not as intensely characterized by additional methods. Products of each of the methods were cross-hybridized to each other, and in several cases fragments of the same transcript were isolated by independent methods (Table 1). In the case of glucose-6-phosphatase (Fig. 3), different segments of the same gene were isolated by different methods. Clones isolated in subsequent experiments that were completely contained within existing clones (regardless of the isolation method) were not characterized further and do not appear in Table 1. Therefore, these results cannot be used rigorously to compare transcript isolation meth-

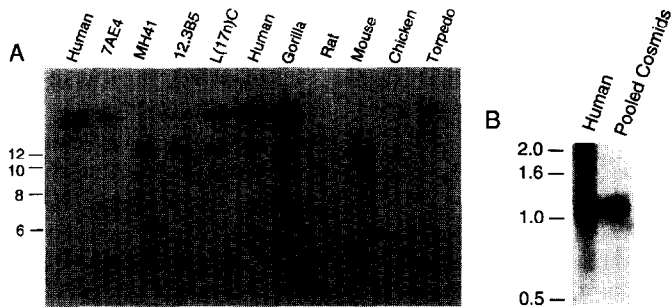


FIG. 2. Mapping panels for the localization of transcripts to the *BRCA1* region. (A) Somatic cell hybrid panel. Southern blot analysis using cell lines containing various regions of human chromosome 17 and genomic DNAs (restricted with *Xba*I) from various other species. Cell lines (lanes 2–5) contained the following regions of human chromosome 17 in the specified rodent background: 7AE4 (intact chromosome 17, rat background), MH41 (17q23.2–qter, mouse background), 12.3B5 (17p–q12, mouse background), and L(17n)C (17q, mouse background). The relative migration of the molecular weight marker (in kb) is shown to the left. The filter was hybridized with a probe from the MTO-108 clone (contains part of the *G6PT* gene). A band (>12 kb) is found in the 7AE4 and L(17n)C cell lines, as well as in human and gorilla DNA. The absence of this band in the MH41 and 12.3B5 cell lines allows assignment of this clone to the 17q12–q22 interval. A mouse-specific band (\approx 11 kb) and a rat-specific band (\approx 7 kb, visible on longer exposures) are present in hybrid cells with the appropriate backgrounds. The presence of bands in other vertebrate DNAs indicates a high degree of evolutionary conservation of this sequence. (B) An example of a “two-lane” mapping panel (restricted with *Pst*I) as described under Materials and Methods. Lane 1 contains total human genomic DNA (10 μ g). Lane 2 contains pooled DNA (10 ng) of 400 cosmids from the *BRCA1* region. The migration of molecular weight standards (kb) is shown to the left. The blot was hybridized with the probe from the exon clone MTO-110 (169-bp insert). Bands of the same size (\approx 1.2 kb) are present in the cosmid pool and human genomic lanes, indicating that this exon maps to the interval included in the cosmid contig. The uniform background hybridization in the human lane was often observed when inserts smaller than 200 bp were used as probes.

ods. In most cases, the partial cDNAs were then used to screen cDNA libraries to extend the initial clones.

We have failed to clone fragments of at least two genes known to be in this interval. The VHR phosphatase gene has been independently isolated and localized to this region by Kamb *et al.* (1994) and Friedman *et al.* (1995). A 4.3-kb cDNA encoding the putative

CA125 antigen, a B-box protein, has also been mapped to this interval by Campbell *et al.* (1994) and Osborne-Lawrence *et al.* (1995). Both of these genes are known to be represented in our cosmid collection, highlighting the fact that current applications of transcript isolation methodologies are less than 100% effective.

Characterization of Isolated Transcripts

Placement of features on the map. Borrowing a term from the cartographer’s lexicon, we refer to the transcript clones as “features.” For classification purposes we divided the products of our transcript search into three categories: genes, expressed sequence tags (EST), and exon sequence tags (ExST) (Table 1). These features are distributed across the entire contig (Fig. 1). The chromosomal location of clones was first verified by hybridization to somatic cell hybrid panels as described under Materials and Methods (Fig. 2). Higher resolution mapping was achieved by hybridizing individual clones to gridded arrays of the complete cosmid contig. At least one feature was mapped to each cosmid used for transcript searching. The cosmids 120D8, 95G7, 102E12, and 52C6 were not used in our transcript isolation effort. The transcript sizes of several genes have been determined by Northern blot analysis (Table 1).

It should be noted that the exact physical distance spanned by these cosmids is not known. Our cosmid contig materials contained three gaps (Couch *et al.*, 1995), one of which is now known to contain the majority of the *BRCA1* gene (Miki *et al.*, 1994; Futreal *et al.*, 1994b). The 40 cosmids used for transcript isolation overlap each other to varying extents (Couch *et al.*, 1995). Assuming conservatively that each cosmid represents only 15 kb of unique genomic DNA (i.e., not found in the adjacent cosmid), we have surveyed approximately 600 kb of genomic material. This estimate does not include the region containing the *RNU2* gene cluster, which is known to contain approximately 20 copies of this gene, spanning 120 kb of genomic DNA (Westin *et al.*, 1984).

Genes. There are two overlapping criteria by which cloned material was placed into the “gene” category. (1)

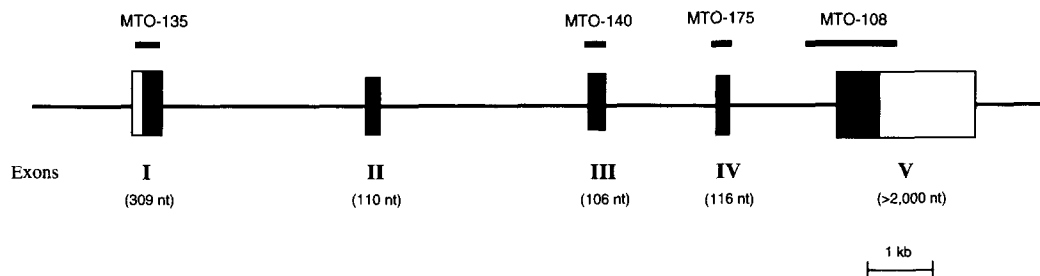


FIG. 3. Schematic diagram of the glucose-6-phosphatase (*G6PT*) gene adapted from Lei *et al.* (1993). Exons are represented as boxes. Shaded regions correspond to the translated portion of the *G6PT* cDNA. Regions of this gene independently cloned in this report are shown above the schematic. Clones MTO-135, -140, and -175 were isolated by exon amplification. Clones MTO-140 and -175 correspond exactly to *G6PT* exons III and IV as reported in Lei *et al.* (1993). The sequence of clone MTO-135 reveals an exact match to nucleotides 134–245 of exon I. Exon I spans from nucleotides 1 to 309 of the *G6PT* cDNA. Because the exon-amplified fragments are directionally cloned, it can be determined that MTO-135 was originally trapped from the DNA strand opposite to that of *G6PT* exon I. The clone MTO-108 was isolated as an evolutionarily conserved fragment (1.2-kb *Sac*I, subclone) of the cosmid 29F5. This genomic clone spans part of *G6PT* intron IV and exon V.

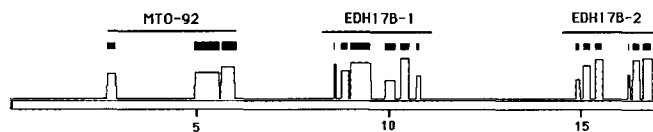


FIG. 4. GRAIL analysis of the genomic sequence containing the *EDH17B1* and *EDH17B2* loci. GRAIL 2 analysis was performed on 21 kb of genomic sequence (Accession No. M84472). The results of the analysis for the first 17 kb of this sequence are shown. The bottom-most horizontal line is scaled in kilobases. Putative coding sequences are represented by GRAIL as flattened peaks along the next horizontal line. Thick bars above each peak delimit the exons as predicted by GRAIL. The location of the *EDHB-1* and *-2* genes are shown by the thin horizontal lines. The presence of three novel exons is predicted by GRAIL. These putative exons are found to span the following positions: 2541–2769, 4783–5391, and 5459–5794 bp of the published sequence. cDNA clones were shown to contain these predicted exons spliced together and not the intervening genomic sequences.

Sequence of the cloned fragment revealed significant homology to a known transcript. The vacuolar proton pump 1 (VPP1), *E1A-f*, glucose-6-phosphatase, and *MOX1* genes are examples of cases in which the isolation of a single fragment provided sufficient information to infer that a clone was part of a *bona fide* gene. (2) The fragment is part of a cDNA clone. Intact cDNA library clones were isolated by the island rescue and direct screening methods. These clones are therefore considered “genes” for mapping purposes. DNA fragments cloned by exon amplification or direct selection were placed in the gene category only when they met criterion one or were successfully used as probes to isolate clones from cDNA libraries (criterion two).

In total, 26 genes were placed on the map. Fragments of 2 genes already known to map to this interval, *EDH17B2* and *PPY*, were isolated. In addition, a previously unrecognized transcription unit was identified in genomic sequence of the *EDHB* region (see above). We isolated fragments of two previously cloned genes: glucose-6-phosphatase (G6PT), the gene responsible for glycogen storage disease type 1a (Lei *et al.*, 1993), and *E1A-f*, an *ets*-oncogene related transcription factor (Higashino *et al.*, 1993). The precise chromosomal location of these genes was previously unknown, although the G6PT gene has been localized to chromosome 17 (Lei *et al.*, 1994). Clones corresponding to three additional genes, the *Ki antigen* (Albertsen *et al.*, 1994), *MOX1* (Futreal *et al.*, 1994a), and *BRCA1* (Miki *et al.*, 1994) were also independently isolated and mapped to this interval by others during the course of this work. Of the remaining 21 genes, 11 detect similar sequences in GenBank by BLAST analysis. Of these, 3 genes, HMG-17 (Landsman *et al.*, 1986, 1989), γ -tubulin (Zheng *et al.*, 1991), and ribosomal protein L27 (Gallagher *et al.*, 1994) are members of gene families. A fourth, the *FLT4* tyrosine kinase (Aprelikova *et al.*, 1992), hybridizes to specific cosmids from this interval and to additional sites in the genome. Sequence obtained to date from this cDNA clone is restricted to the 3' untranslated region of the *FLT4* gene (data not shown). In the absence of genomic sequence from this region, we are

unable to determine whether the cDNAs isolated for these genes were transcribed from these exact loci on chromosome 17 or from the loci of other family members. Five additional genes appear to be human homologues of genes cloned from other species. As mentioned above, the human *MOX1* clones isolated represent the human equivalent of this mouse gene (Futreal *et al.*, 1994a).

The human vacuolar proton pump-1 cDNA shares 97% amino acid identity over 29 residues with the 116-kDa subunit of the bovine vacuolar proton pump (Peng *et al.*, 1994). A rat gene probe was previously used to localize the human VPP-1 homologue to 17q21–qter in somatic cell hybrids (Ozelik *et al.*, 1991). The human *Enhancer of Zeste (E(z))* gene shares 69% amino acid similarity with the *Drosophila E(z)* gene, a negative regulator of the *Antennapedia* and *Bithorax* gene complexes (Jones and Gelbart, 1993). The complete cDNA sequence of the human *E(z)* gene will be reported elsewhere (Abel *et al.*, manuscript in preparation). Partial sequence of the human copper amine oxidase cDNA reveals a 77% amino acid identity over a 193-amino-acid segment with the bovine serum copper amine oxidase (Mu *et al.*, 1994). The determination of the complete sequence of this gene is in progress. We have isolated a clone identical to a portion of EST-HFBCV14 (Adams *et al.*, 1991), which we infer to be part of the human vesicle amine transporter-1 gene (VAT-1) by the virtue of the homology between the EST and *Torpedo californica* (Pacific electric ray) VAT-1 gene.

The GenBank/BLAST alignments of four genes revealed limited but significant similarity (smallest sum probability $< 1 \times 10^{-5}$) to sequence in GenBank. Partial sequence of clone MTO-61 reveals limited homology (46% amino acid identity in a 48-amino-acid overlap) to the rat and bovine neurexins (Ushkaryov and Sudhof, 1993). Exon clone MTO-134 shares 44% amino acid identity (82% similarity over a 29-amino-acid overlap) to the human and chicken 23-kDa component (p23) of the progesterone receptor complex (Johnson *et al.*, 1994; Smith and Toft, 1992). The exon clone, MTO-173, is contained within cDNAs isolated with MTO-134 (data not shown). This exon does not share homology to the *p23* gene, suggesting that the homology between these two genes is limited. Sequence of the exon MTO-192 reveals 70% amino acid similarity with the *Caenorhabditis elegans* and *Saccharomyces cerevisiae* pre-mRNA splicing factors 22 and 16 (Company *et al.*, 1991; Sultson *et al.*, 1992). Additional cDNA-derived sequences demonstrate that this homology is restricted to the carboxyl half of these genes. (Ostermeyer *et al.*, in preparation). The sequence of trapped exon clone MTO-122 is 40% identical and 60% similar (over a 60-amino-acid overlap) to the *Xenopus* int-2 (FGF-3) protein (Tannahill *et al.*, 1992). This region is also common to mouse keratinocyte growth factor Fgf-7 (Dickson and Mason, 1993) and the DNA binding protein PRDII-BF1 (Fan and Maniatis, 1990).

Partial sequence from the seven remaining genes reveals no significant nucleotide or amino acid similarity

ties to the entries in GenBank. It will be necessary to obtain full-length cDNA sequence of these genes before it can be determined whether they represent previously described transcription units.

ExST's. Our second map feature is products of exon amplification experiments that map to specific cosmids but have not detected cDNA clones in breast or fetal brain cDNA libraries after screening at least 5×10^5 recombinant phage. The complete sequence of these trapped exons has been determined and searched against GenBank; similar sequences were not located. These 39 unique clones in several cases appear to be clustered in the contig. Most notably, the 81C6 and 134A3 cosmids (Fig. 1) contain 10 ExST's. At least two other less impressive clusters are found in the contig (e.g., distal end of cosmid 90E4 and cosmid 32E5). Even though these exons are distributed across this area in a pattern resembling a gene, we have been unable to demonstrate either by sequence similarity to known genes or by the isolation of cDNAs that they belong to a contiguous transcription unit. A likely explanation for this is that an insufficient number of tissues or developmental stages have been surveyed for expression. It is also possible that a more sensitive test such as RT-PCR will be required to detect low-level transcripts from these areas. Further studies will be needed to determine whether these ExST's are constituents of true transcription units or artifacts of the exon amplification process.

ESTs. The third category is ESTs. These clones are the products of direct selection experiments. Because they are products amplified from a cDNA library, they represent fragments of transcribed sequences. Sufficient evidence does not exist to classify them as genes as outlined above. As is the case for the ExST's, the sequences of these fragments are not similar to sequences in the public databases. Like many randomly generated ESTs (Adams *et al.*, 1991), these fragments are of unknown function. Unlike the majority of ESTs, the precise chromosomal location of these putative genes is now known.

In conclusion, we have isolated transcripts from approximately 600 kb of the human genome. It is likely that this interval of the genome contains at least 26 unique genes. Although we are unable to calculate the exact gene density of this region, our finding of 1 gene approximately every 20 kb is close to the value expected if the human genome contains 100,000 randomly distributed genes (e.g., 1 gene per 30 kb). This transcript map and further refinements thereof will add to our knowledge of genome structure, gene organization, and expression.

ACKNOWLEDGMENTS

The authors thank Joan Effiom for assistance in the preparation of the manuscript and Sonye Danoff for critical reading and helpful suggestions. This work was supported by NIH Grant CA-57601 (B.W.).

Note added in proof. The cloning of the human homologue of yeast Prp22 has recently been reported by Y. Ono, M. Ohno, and Y. Shimura (1994). *Mol. Cell. Biol.* **14**: 7611–7620. The sequence of the cDNA clone designated by Ono as HRH1 is identical to the Prp22 clones reported in this paper.

REFERENCES

- Abel, K. J., Castilla, L. H., Buckler, A. J., Couch, F. J., Ho, P., Schaefer, I., Chandrasekharappa, S. C., Collins, F. S., and Weber, B. (1994). Isolation of gene sequences from the BRCA1 region of chromosome 17q21 by exon amplification. In "Identification of Transcribed Sequences" (U. Hochgeschwender and K. Gardiner, Eds.), pp. 183–198, Plenum, New York.
- Adams, M. D., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulos, M. H., Xiao, H., Merril, C. R., Wu, A., Olde, B., Moreno, R. F., Kerlavage, A. R., McCombie, W. R., and Venter, J. C. (1991). Complementary DNA sequencing: Expressed sequence tags and the human genome project. *Science* **252**: 1651–1656.
- Albertsen, H. M., Smith, S. A., Mazoyer, S., Fujimoto, E., Stevens, J., Williams, B., Rodriguez, P., Cropp, C. S., Slijepcevic, P., Carlson, M., Robertson, M., Bradley, P., Lawrence, E., Harrington, T., Mei Sheng, Z., Hoopes, R., Sternberg, N., Brothman, A., Callahan, R., Ponder, B. A. J., and White, R. (1994). A physical map and candidate genes in the BRCA1 region on chromosome 17q12–21. *Nature Genet.* **7**: 472–479.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**: 403–410.
- Aprelikova, O., Pajusola, K., Partanen, J., Armstrong, E., Alitalo, R., Bailey, S. K., McMahon, J., Wasmuth, J., Huebner, K., and Alitalo, K. (1992). FLT4, a novel class III receptor tyrosine kinase in chromosome 5q33–qter. *Cancer Res.* **52**: 746–748.
- Bowcock, A. M., Anderson, L. A., Friedman, L. S., Black, D. M., Osborne-Lawrence, S., Rowell, S. E., Hall, J. M., Solomon, E., and King, M.-C. (1993). THRA and D17S183 flank an interval of 4 cM for the breast-ovarian-cancer gene (BRCA1). *Am. J. Hum. Genet.* **52**: 718–722.
- Buckler, A. J., Chang, D. D., Graw, S. L., Brook, J. D., Haber, D. A., Sharp, P. A., and Housman, D. E. (1991). Exon amplification: A strategy to isolate mammalian genes based on RNA splicing. *Proc. Natl. Acad. Sci. USA* **88**: 4005–4009.
- Campbell, I. G., Nicolai, H. M., Foulkes, W. D., Senger, G., Stamp, G. W., Allan, G., Boyer, C., Jones, K., Bast, R. C., Soloman, E., Trowsdale, J., and Black, D. M. (1994). A novel gene encoding a B-box protein within the BRCA1 region at 17q21.1. *Hum. Mol. Genet.* **3**: 589–594.
- Chamberlain, J. S., Boehnke, M., Frank, T. S., Kioussis, S., Xu, J., Guo, S.-W., Hauser, E. R., Norum, R. A., Helmbold, E. A., Markel, D. S., Keshavarzi, S. M., Jackson, C. E., Calzone, K., Garber, J., Collins, F. S., and Weber, B. L. (1993). BRCA1 maps proximal to D17S579 on chromosome 17q21 by genetic analysis. *Am. J. Hum. Genet.* **52**: 792–798.
- Church, D. M., Stotler, C. J., Rutter, J. L., Murrell, J. R., Trofatter, J. A., and Buckler, A. J. (1994). Isolation of genes from complex sources of mammalian genomic DNA using exon amplification. *Nature Genet.* **6**: 98.
- Collins, F. S. (1992). Positional cloning: Let's not call it reverse any more. *Nature Genet.* **1**: 3–6.
- Collins, F. S., and Galas, D. (1993). A new five year plan for the U.S. Human Genome Project. *Science* **262**: 43.
- Company, M., Arenas, J., and Abelson, J. (1991). Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes. *Nature* **349**: 487–493.
- Couch, F. J., Castilla, L. H., Xu, J., Abel, K. A., Welcsh, P., King, S. E., Wong, L., Ho, P. P., Merajver, S., Brody, L. C., Yin, G., Hayes, S. T., Geiser, L. M., Flejter, W. L., Glover, T. W., Friedman, L. S., Lynch, E. D., Meza, J. E., King, M.-C., Law, D. J., Deaven, L., Bowcock, A. M., Collins, F. S., Weber, B. L., and

- Chandrasekharappa, S. C. (1995). A YAC-, P1-, and cosmid-based physical map of the BRCA1 region on chromosome 17q21. *Genomics* **25**: 264–273.
- Couch, F. J., Weber, B. L., Collins, F. S., Tagle, D. A. (1994). Isolation of expressed sequences from the chromosome 17q21 BRCA1 region by magnetic bead capture. In "Identification of Transcribed Sequences" (U. Hochgeschwender and K. Gardiner, Eds.), pp. 51–63, Plenum, New York.
- Dickson, C., and Mason, I. J. (1993). FGF-7 keratinocyte growth factor expression during mouse development suggests roles in myogenesis, forebrain regionalisation and epithelial–mesenchymal interaction. EMBL Accession No. Z22703.
- Duyk, G. M., Kim, S., Myers, R. M., and Cox, D. R. (1990). Exon trapping: A genetic screen to identify candidate transcribed sequences in cloned mammalian genomic DNA. *Proc. Natl. Acad. Sci. USA* **87**: 8995–8999.
- Easton, D. F., Bishop, D. T., Ford, D., and Crockford, G. P. (1993). The breast cancer linkage consortium: Genetic linkage analysis in familial breast and ovarian cancer—Results from 214 families. *Am. J. Hum. Genet.* **52**(4): 678–701.
- Fan, C. M., and Maniatis, T. (1990). A DNA-binding protein containing two widely separated zinc finger motifs that recognize the same DNA sequence. *Genes Dev.* **4**: 29–42.
- Feinberg, A. P., and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction fragments to high specific activity. *Anal. Biochem.* **132**: 6.
- Friedman, L. S., Lynch, E. D., and King, M. C. (1993). Two independent polymorphisms at the 17 β -hydroxysteroid dehydrogenase (EDH17B) gene (17q21). *Hum. Mol. Genet.* **2**: 821.
- Friedman, L. S., Ostermeyer, E. A., Lynch, E. D., Welcsh, P., Szabo, C. I., Meza, J. E., Anderson, L. A., Dowd, P., Lee, M. K., Rowell, S. E., Ellison, J., Boyd, J., and King, M.-C. (1995). 22 genes from chromosome 17q21: Cloning, sequencing, and characterization of mutations in breast cancer families and tumors. *Genomics* **25**: 256–263.
- Futreal, P. A., Cochran, C., Rosenthal, J., Miki, Y., Swenson, J., Hobbs, M., Bennett, L. M., Haugen-Strano, A., Marks, J., Barrett, J. C., Tavtigian, S. V., Shattuck-Eidens, D., Kamb, A., Skolnick, M., and Wiseman, R. W. (1994a). Isolation of a diverged homeobox gene, MOX1, from the BRCA1 region on 17q21 by solution hybrid capture. *Hum. Mol. Genet.* **3**: 1359–1364.
- Futreal, A. P., Liu, Q., Shattuck-Eidens, D., Cochran, C., Harshman, K., Tavtigian, S., Bennett, L. M., Haugen-Strano, A., Swensen, J., Miki, Y., Eddington, K., McClure, M., Frye, C., Weaver-Feldhaus, J., Ding, W., Gholami, Z., Soderkvist, P., Terry, L., Jhanwar, S., Berchuck, A., Iglehart, J. D., Marks, J., Ballinger, D. G., Barrett, J. C., Skolnick, M. H., Kamb, A., and Wiseman, R. (1994b). BRCA1 mutations in primary breast and ovarian carcinomas. *Science* **266**: 120–122.
- Gallagher, R. A., McClean, P. M., and Malik, A. N. (1994). Cloning and nucleotide sequence of a full length cDNA encoding ribosomal protein L27 from human fetal kidney. *Biochim. Biophys. Acta* **1217**: 329–332.
- Gish, W., and States, D. J. (1993). Identification of protein coding regions by database similarity search. *Nature Genet.* **3**: 266–272.
- Hall, J. M., Lee, M. K., Newman, B., Morrow, J. E., Anderson, L. A., Huey, B., and King, M.-C. (1990). Linkage of early onset breast cancer to chromosome 17q21. *Science* **250**: 1684–1689.
- Higashino, F., Yoshida, K., Fujinaga, Y., Kamio, K., and Fujinaga, K. (1993). Isolation of a cDNA encoding the adenovirus E1A enhancer binding protein: A new human member of the ets oncogene family. *Nucleic Acids Res.* **21**: 547.
- Johnson, J. L., Beito, T. G., Krco, C. J., and Toft, D. O. (1994). Characterization of a novel 23-kilodalton protein of inactive progesterone receptor complexes. *Mol. Cell. Biol.* **14**: 1956–1963.
- Jones, R. S., and Gelbart, W. M. (1993). The drosophila polycomb-group gene enhancer of zeste contains a region with sequence similarity to trithorax. *Mol. Cell. Biol.* **13**: 6357–6366.
- Kamb, A., Futreal, P. A., Rosenthal, J., Cochran, C., Harshman, K. D., Liu, Q., Phelps, R. S., Tavtigian, S. V., Tran, T., Hussey, C., Bell, R., Miki, Y., Swensen, J., Hobbs, M. R., Marks, J., Bennett, L. M., Barrett, J. C., Wiseman, R. W., and Shattuck-Eidens, D. (1994). Localization of the VHR phosphatase gene and its analysis as a candidate for BRCA1. *Genomics* **23**: 163–167.
- Krizman, D. B., and Berget, S. M. (1993). 3'-Terminal exon trapping: Identification of genes from vertebrate DNA. *Focus* **15**: 106–108.
- Landsman, D., Soares, N., Gonzalez, F. J., and Bustin, M. (1986). Chromosomal protein HMG-17: Complete human cDNA sequence and evidence for a multigene family. *J. Biol. Chem.* **261**: 7479–7484.
- Landsman, D., McBride, O. W., and Bustin, M. (1989). Human non-histone chromosomal protein HMG-17: Identification, characterization, chromosome localization and RFLPs of a functional gene from the large multigene family. *Nucleic Acids Res.* **17**: 2301–2314.
- Leach, R. J., Thayer, M. J., Schafer, A. J., and Fournier, R. E. K. (1989). Physical mapping of human chromosome 17 using fragment-containing microcell hybrids. *Genomics* **5**: 167–176.
- Lei, K. J., Pan, C. J., Shelly, L. L., Liu, J. L., and Chou, J. Y. (1994). Identification of mutations in the gene for glucose-6-phosphatase, the enzyme deficient in glycogen storage disease type 1A. *J. Clin. Invest.* **93**: 1994–1999.
- Lei, K. J., Shelly, L. L., Pan, C. J., Sidbury, J. B., and Chou, J. Y. (1993). Mutations in the glucose-6-phosphatase gene that cause glycogen storage disease type 1a. *Science* **262**: 580–583.
- Lovett, M., Kere, J., and Hinton, L. M. (1991). Direct selection: A method for the isolation of cDNAs encoded by large genomic regions. *Proc. Natl. Acad. Sci. USA* **88**: 9628–9632.
- 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, H., Yakumo, K., Gholami, Z., Shaffer, D., Stone, S., Bayer, S., Wray, C., Bogden, R., Dayananth, P., Ward, J., Tonin, P., Narod, S., Bristow, P. K., Norris, F. H., Helvering, L., Morrison, P., Rosteck, P., Lai, M., Barrett, J. C., Lewis, C., Neuhausen, S., Cannon-Albright, L., Goldgar, D., Wiseman, R., Kamb, A., and Skolnick, M. H. (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* **266**: 66.
- Mu, D., Medzihradzky, K. F., Adams, G. W., Mayer, P., Hines, W. M., Burlingame, A. L., Smith, A. J., Cai, D., and Klinman, J. P. (1994). Primary structures for a mammalian cellular and serum copper amine oxidase. *J. Biol. Chem.* **269**: 9926–9932.
- Narod, S. A., Feunteun, J., Lynch, H. T., Watson, P., Conway, T., Lynch, J., and Lenoir, G. M. (1991). Familial breast–ovarian cancer locus on chromosome 17q12–23. *Lancet* **338**: 82–83.
- Oliver, S. G., van der Aart, Q. J., Agostoni-Carbone, M. L., Aigle, M., Alberghina, L., Alexandraki, D., Antoine, G., Anwar, R., and Ballesta, J. P. (1992). The complete DNA sequence of yeast chromosome III. *Nature* **357**: 38–46.
- Osborne-Lawrence, S., Welcsh, P. L., Spillman, M., Chandrasekharappa, S. C., Gallardo, T. D., Lovett, M., and Bowcock, A. M. (1995). Direct selection of expressed sequences within a 1-Mb region flanking BRCA1 on human chromosome 17q21. *Genomics* **25**: 248–255.
- Ozcelik, T., Sudhof, T. C., and Francke, U. (1991). Chromosomal assignments of genes for vacuolar (endomembrane) proton pump subunits VPP1/Vpp-1 (116 kDa) and VPP3/Vpp-3 (58 kDa) in human and mouse (Abstract). *Cytogenet. Cell Genet.* **58**: 2008–2009.
- Parimoo, S., Patanjali, S. R., Shukla, H., Chaplin, D. D., and Weissman, S. M. (1991). cDNA selection: Efficient PCR approach for the selection of cDNAs encoded in large chromosomal DNA fragments. *Proc. Natl. Acad. Sci. USA* **88**: 9623–9627.
- Patel, K., Cox, R., Shipley, J., Kiely, F., Frazer, K., Cox, D. R., Leh-rach, H., and Sheer, D. (1991). A novel and rapid method for isolating sequences adjacent to rare cutting sites and their use in physical mapping. *Nucleic Acids Res.* **19**: 4371–4375.
- Peltoketo, H., Isomaa, V., Maeentausta, O., and Vihko, R. (1988).

- Complete amino acid sequence of human placental 17-beta-hydroxysteroid dehydrogenase deduced from cDNA. *FEBS Lett.* **239**: 73–77.
- Peng, S. B., Crider, B. P., Xie, X. S., and Stone, D. K. (1994). Alternative mRNA splicing generates tissue-specific isoforms of 116-kDa polypeptide of vacuolar proton pump. *J. Biol. Chem.* **269**: 17262–17266.
- Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). "Molecular Cloning: A Laboratory Manual," 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Shah, M. B., Guan, X., Einstein, J. R., Matis, S., Xu, Y., Mural, R. J., and Uberbacher, E. C. (1994). User's guide to GRAIL and GENQUEST (sequence analysis, gene assembly and sequence comparison systems) e-mail servers and XGRAIL (Version 1.2) and XGENQUEST (Version 1.1) client-server systems. Available by anonymous ftp to arthur.epm.ornl.gov (128.219.9.76) from directory pub/xgrail or pub/xgenQuest as file Manual.grail-genquest.
- Simard, J., Feuteun, J., Lenoir, G., Tonin, P., Normand, T., Luu-The, V., Vivier, A., Lasko, D., Morgan, K., Rouleau, G. A., Lynch, H., Labrie, F., and Narod, S. A. (1993). Genetic mapping of the breast-ovarian cancer syndrome to a small interval on chromosome 17q12–21: Exclusion of candidate genes EDH17B2 and RARA. *Hum. Mol. Genet.* **2**: 1193–1199.
- Smith, D. F., and Toft, D. O. (1992). Composition, assembly and activation of the avian progesterone receptor. *J. Steroid Biochem. Mol. Biol.* **41**: 201–207.
- Smith, S. A., DiCioccio, R. A., Struewing, J. P., Easton, D. F., Gallion, H. H., Albertsen, H., Mazoyer, S., Johansson, B., Steichen-Gersdorf, E., Stratton, M., Ford, D., Marshall, G., White, R. L., Piver, M. S., and Ponder, B. A. J. (1994). Localisation of the breast-ovarian cancer susceptibility gene (BRCA1) on 17q12–21 to an interval of ≤ 1 cM. *Genes Chrom. Cancer* **10**: 71–76.
- Sultson, J., Du, Z., Thomas, K., Wilson, R., Hillier, L., Staden, R., Halloran, N., Green, P., Thierry-Mieg, J., Qiu, L., Dear, S., Coulson, A., Craxton, M., Durbin, R., Berks, M., Metzstein, M., Hawkins, T., Ainscough, R., and Waterson, R. (1992). The *C. elegans* genome sequencing project: A beginning. *Nature* **356**: 37–41.
- Swaroop, A., and Xu, J. (1993). cDNA libraries from human tissues and cells lines. *Cytogenet. Cell Genet.* **64**: 292.
- Tagle, D. A., Swaroop, M., Lovett, M., and Collins, F. S. (1993). Magnetic bead capture of expressed sequences within large genomic segments. *Nature* **161**: 751–756.
- Tannahill, D., Isaacs, H. V., Close, M. J., Peters, G., and Slack, J. M. (1992). Developmental expression of the xenopus int-2 (FGF-3) gene: Activation by mesodermal and neural induction. *Development* **115**: 695–702.
- Uberbacher, E. C., and Mural, R. J. (1991). Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. *Proc. Natl. Acad. Sci. USA* **88**: 11261–11265.
- Ushkaryov, Y. A., and Sudhof, T. C. (1993). Neurexin III alpha: Extensive alternative splicing generates membrane-bound and soluble forms. *Proc. Natl. Acad. Sci. USA* **90**: 6410–6414.
- Valdes, J. M., Tagle, D. A., and Collins, F. S. (1994). Island rescue PCR: A rapid and efficient method for isolating transcribed sequences from yeast artificial chromosomes and cosmids. *Proc. Natl. Acad. Sci. USA* **91**: 5377–5381.
- Wallace, M. R., Marchuk, D. A., Anderson, L. B., Letcher, R., Odeh, H. M., Saulino, A. M., Fountain, J. W., Brereton, A., Nicholson, J., and Mitchell, A. L. (1990). Type I neurofibromatosis gene: Identification of a large transcript disrupted in three NFL patients. *Science* **249**: 181–186.
- Weber, B. L., Abel, K., Brody, L. C., Flejter, W. L., Chandrasekhara, S. C., Couch, F. J., Merajver, S. D., and Collins, F. S. (1994). Familial breast cancer: Approaching the isolation of a susceptibility gene. *Cancer* **74**: 1013–1020.
- Westin, G., Zabielski, J., Hammarstrom, K., Monstein, H., Bark, C., and Pettersson, U. (1984). Clustered genes for human U2 RNA. *Proc. Natl. Acad. Sci. USA* **81**: 3811–3815.
- Zheng, Y., Jung, M. K., and Oakley, B. R. (1991). Gamma-tubulin is present in *Drosophila melanogaster* and *Homo sapiens* and associated with the centrosome. *Cell* **65**: 817–823.