PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12N 15/10, C12Q 1/68 C07H 21/04	A1	**(11) International Publication Number:** WO 93/05149 **(43) International Publication Date:** 18 March 1993 (18.03.93)

(21) International Application Number: PCT/US92/07516

(22) International Filing Date: 4 September 1992 (04.09.92)

(30) Priority data:
754,990 5 September 1991 (05.09.91) US

(71) Applicant: FRED HUTCHINSON CANCER RESEARCH CENTER [US/US]; 1124 Columbia Street, Seattle, WA 98104 (US).

(72) Inventor: FOURNIER, Keith ; 8410 S.E. 47th Street, Mercer Island, WA 98040 (US).

(74) Agents: MONROY, Gladys, H. et al.; Morrison & Foerster, 755 Page Mill Road, Palo Alto, CA 94304 (US).

(81) Designated States: AU, CA, JP, KP, KR, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, SE).

Published
With international search report.

(54) Title: SUBTRACTIVE HYBRIDIZATION CLONING FOR PROBES FOR GENE ISOLATION AND MAPPING

(57) Abstract

Simplified methods to isolate eukaryotic genes based on their genomic location are provided. The methods utilize probes that are enriched in a sequence of a target eukaryotic gene. The probes are obtained by subtractive hybridization utilizing at least two hybrid cell lines which are nearly isogenic. The hybrid cell lines contain genetic sequences from a eukaryotic chromosome that is heterologous to the host cell, but differ primarily in the sequence of the target gene within the heterologous eukaryotic chromosome. Using the lines, probes that are enriched for sequences encoded in the region of non-overlap between the provided somatic cell hybrid lines are generated by subtractive hybridization. The subtractive hybridization of the method does not rely only on differential expression. Rather, the 3′-sequence divergence between the transcripts of the non-overlap region of the heterologous chromosome and the transcripts of the hybrid host recipient cell allows the specific removal of host sequences from the probe population. Thus, the method overcomes the problem of homology of genes between the targeted gene and the analogous gene (if any) in the recipient hybrid host cell.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**FI**	Finland	**MN**	Mongolia
AU	Australia	**FR**	France	**MR**	Mauritania
BB	Barbados	**GA**	Gabon	**MW**	Malawi
BE	Belgium	**GB**	United Kingdom	**NL**	Netherlands
BF	Burkina Faso	**GN**	Guinea	**NO**	Norway
BG	Bulgaria	**GR**	Greece	**NZ**	New Zealand
BJ	Benin	**HU**	Hungary	**PL**	Poland
BR	Brazil	**IE**	Ireland	**PT**	Portugal
CA	Canada	**IT**	Italy	**RO**	Romania
CF	Central African Republic	**JP**	Japan	**RU**	Russian Federation
CG	Congo	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CH	Switzerland			**SE**	Sweden
CI	Côte d'Ivoire	**KR**	Republic of Korea	**SK**	Slovak Republic
CM	Cameroon	**LI**	Liechtenstein	**SN**	Senegal
CS	Czechoslovakia	**LK**	Sri Lanka	**SU**	Soviet Union
CZ	Czech Republic	**LU**	Luxembourg	**TD**	Chad
DE	Germany	**MC**	Monaco	**TG**	Togo
DK	Denmark	**MG**	Madagascar	**UA**	Ukraine
ES	Spain	**ML**	Mali	**US**	United States of America

SUBTRACTIVE HYBRIDIZATION CLONING FOR PROBES FOR GENE ISOLATION AND MAPPING

Technical Field

The invention relates to methods for the isolation of eukaryotic genes, and more specifically, to methods which employ hybrid cell lines and subtractive hybridization cloning in the preparation of probes for the isolation and/or mapping of expressed genes.

Background

The advent of significant technical advances in molecular and somatic cell genetics has led to the creation of maps of the human genome. The use of restriction fragment length polymorphisms (RFLP) in conjunction with genetic linkage analysis has allowed the construction of meiotic linkage maps for each of the 23 human chromosomes with an average resolution of 10 to 15 centiMorgans (cM)(H. Donis-Keller et al. (1987), Cell 51:319). Techniques which allow the separation of human chromosomes from one another, either in rodent-human somatic cell hybrids or by physical chromosome sorting, have also led to the assignment of hundreds of human loci to specific human chromosomes. (P.J. McAlpine et al. (1989), Cytogenet. Cell Genet. 51: 13). Moreover, in situ hybridization provides a means of localizing sequences which hybridize to molecular probes to specific positions on human chromosomes. (P. Lichter et al. (1990), Science 247:64.)

Despite these technical advances, present-day maps of human chromosomes are very crude in molecular terms. On average, 1 percent meiotic recombination between two markers on a human chromosome corresponds to

1 megabase pairs (Mb) of DNA. In situ hybridization can localize markers to within two percent of total chromosome length, but in molecular terms, this again represents several million base pairs. Pulsed field gel electrophoresis (PFGE), which can separate DNA fragments of several million base pairs in agarose gels, provides a potentially powerful means for constructing long-range physical maps of human chromosomes when used in conjunction with restriction enzymes that cut infrequently in human DNA. (D.C. Schwartz and C.R. Cantor (1984), Cell 37:67.) However, the paucity of useful rare-cutter enzymes and the nonrandom distribution of rare-cutter sites in human genomic DNA make it difficult to order DNA sequences more than a few hundred kilobase pairs (KB) apart with this technique alone. (D.R. Cox et al. (1990), Science 250:245.)

The recent isolation of genes responsible for several human inherited disorders has relied on their chromosomal location, rather than a phenotypic or biochemical characteristic of these genes. These "positional cloning" methods utilize tightly linked markers as landmarks for isolating contiguous sequences of genomic DNA which span the interval between markers. Candidate exon sequences are identified by hybridization of genomic sequences across species (Riordan et al. (1989), Science 245:1066; Viskochil et al. (1990), Cell 62:187; Call et al. (1990), Cell 60:509), or identification of CpG islands which are often adjacent to transcribed sequences (Bird (1986), Nature 231:209, Oberle et al. (1991), Science 252:1097). Although these methods have been successful in the isolation of a few genes, they have proven to be laborious. Development of simplified methods to isolate genes based on their genomic location are needed to facilitate the search for medically-relevant genes and aid in the construction of a high-resolution expressed tagged sequence map.

Relevant Art

M.R. Wallace et al. (1990), Science 249: 181 reports the identification of a putative Type 1 Neurofibromatosis (NF1) gene, using techniques involving chromosome jumping and yeast artificial chromosomes in the identification process.

D. Viskochil et al. (1990), Cell 62:187 reports the detection and characterization of three NF1 mutations using techniques including pulsed-field gel and Southern blot analyses.

R.M. Cawthton et al. (1990), Cell 62:193 reports the sequencing of overlapping cDNA clones from the translocation breakpoint region (TBR) gene at the NF1 locus, and a characterization of a segment of the NF1 genomic DNA.

B. Kerem et al. (1989), Science 245:1073 reports the identification of the Cystic Fibrosis (CF) gene using techniques including restriction enzyme length polymorphism (RFLP), genetic analysis, and cDNA sequencing.

J. Riordan et al. (1989), Science 245:1066 reports the identification of the CF gene using techniques including the isolation and characterization of overlapping cDNA clones from epithelial cell libraries with a genomic DNA segment containing a portion of a putative CF locus.

E. rose et al. (1990), Cell 60:495 reports the development of a physical map of the 11p13 region containing the Wilms' tumor (WT) locus, and the localization of a candidate WT gene. The reported strategy for the construction of the map involved PFGE analysis of DNA obtain from irradiation-reduced somatic cell hybrids which contained limited segments of human chromosome 11. In the mapping, all DNA fragments of the WAGR region are directly visualized using cloned DNA probes derived from the hybrid cells.

K.M. Call et al. (1990), Cell 60:509 reports the isolation and characterization of a Zinc Finger Polypeptide Gene at the WT locus. The isolation technique included increasing the density of genomic DNA sequences within the WT region using cosmid libraries from a hybrid cell line in which the short arm of chromosome 11 had been segregated from the remainder of the human genome in a Chinese hamster background. Clones within the WAGR region were identified using a mapping panel of somatic cell hybrids containing different fragments of human chromosome 11p.

S.M. Hedrick et al. (1984), Nature 308:149, reports the isolation of cDNA clones encoding T cell specific membrane-associated proteins; the isolation method included an enriched probe for the screening of cDNA libraries. The probe reportedly was prepared by the synthesis of labelled cDNAs of the membrane-bound polysomal RNA of antigen-specific T cells and the removal by RNA hybridization of those sequences also expressed in B cells. R.L. Davis et al. (1987), Cell 51:987, reportedly prepared probes enriched for putative myogenic regulatory sequences. The probes were prepared using a method of subtractive hybridization, in which cDNAs to proliferating myoblast poly $(A)^+$ RNA were hybridized to RNA from a precursor cell line.

Summary of the Invention

The invention provides simplified methods to isolate eukaryotic genes based on their genomic location. The methods utilize probes that are enriched in a sequence of a target eukaryotic gene. The probes are obtained by subtractive hybridization utilizing at least two hybrid cell lines which are nearly isogenic. The hybrid cell lines contain genetic sequences from a eukaryotic chromosome that is heterologous to the host cell, but differ primarily in the sequence of the target gene within the heterologous eukaryotic chromosome.

Using the lines, probes that are enriched for sequences encoded in the region of non-overlap between the provided somatic cell hybrid lines are generated by subtractive hybridization. The subtractive hybridization of the method does not rely only on differential expression. Rather, the 3'-sequence divergence between the transcripts of the nonoverlap region of the heterologous chromosome and the transcripts of the hybrid host recipient cell allows the specific removal of host sequences from the probe population. Thus, the technique overcomes the problem of homology of genes between the targeted gene and the analogous gene (if any) in the recipient hybrid host cell.

The enriched probes are used for the detection, isolation and characterization of sequences from the target region.

Accordingly an embodiment of the invention is a method of preparing a probe for detection of a target eukaryotic gene sequence comprising:

(a) providing a composition containing polynucleotides with sequences complementary to those of transcripts from a (+) hybrid cell, wherein the (+) hybrid cell contains heterologous eukaryotic chromosomal DNA comprised of a target gene sequence, and wherein the polynucleotide is comprised of a sequence complementary to the 3'-end of the transcript;

(b) providing a composition comprised of polynucleotides containing the sequences of transcripts from a (-) hybrid cell, wherein the (-) cell contains a non-overlap region in the heterologous chromosomal DNA, and wherein the non-overlap region encompasses the target eukaryotic gene sequence;

(c) incubating the composition of (a) with the composition of (b) under stringent conditions which allow the formation of heteroduplexes;

(d) isolating polynucleotides of (a) which are not in heteroduplexes with the polynucleotides of (b).

Another embodiment of the invention is a composition comprised of a polynucleotide probe prepared by the above-described method.

Yet another embodiment of the invention is a method of detecting a target gene comprising:

(a) providing a polynucleotide probe prepared by the above-described method;

(b) providing polynucleotides from a sample suspected of containing the target gene;

(c) incubating the polynucleotide probe composition of (a) with the polynucleotides of (b) under stringent conditions which allow the formation of heteroduplexes between the polynucleotides of (a) and (b), if any; and

(d) detecting heteroduplexes formed in (c), if any.

Brief Description of the Figures

Figure 1 is a diagrammatic representation of steps involved in the preparation of microcell hybrids.

Figure 2 is a schematic diagram of human chromosome 17 showing the approximate positions of marker loci. The chromosome fragments retained by five different 7A series hybrids, deduced from marker analysis as described by Leach et al. (1989) are indicated.

Figure 3 is a schematic diagram of fine mapping within the nonoverlap region of 7AE-27 and 7AD-7.

Figure 4 is a 229 bp sequence from the 3'-end of the human Ria transcript. This sequence represents the 3'-end of the 2700 bp λ157 EcoRI fragment, and contains a polyadenylation signal and a short poly(A) tract.

Detailed Description of the Invention

The present invention provides methods which simplify the isolation and/or mapping of eukaryotic genes, particularly human genes which may be medically

relevant or domestic animal genes of agricultural significance. Methods are described herein for the preparation of probes which are enriched in the desired gene sequences, and methods for the use of the enriched probes.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA, genetics and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See e.g., Maniatis, Fitsch & Sambrook, MOLECULAR CLONING; A LABORATORY MANUAL (1982), and Second Edition, (1989); DNA CLONING, VOLUMES I AND II (D.N Glover ed. 1985); OLIGONUCLEOTIDE SYNTHESIS (M.J. Gait ed, 1984); NUCLEIC ACID HYBRIDIZATION (B.D. Hames & S.J. Higgins eds. 1984); TRANSCRIPTION AND TRANSLATION (B.D. Hames & S.J. Higgins eds. 1984); ANIMAL CELL CULTURE (R.I. Freshney ed. 1986); IMMOBILIZED CELLS AND ENZYMES (IRL Press, 1986); B. Perbal, A PRACTICAL GUIDE TO MOLECULAR CLONING (1984); the series, METHODS IN ENZYMOLOGY (Academic Press, Inc.); GENE TRANSFER VECTORS FOR MAMMALIAN CELLS (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory), Methods in Enzymology Vol. 154 and Vol. 155 (Wu and Grossman, and Wu, eds., respectively), Mayer and Walker, eds. (1987), IMMUNOCHEMICAL METHODS IN CELL AND MOLECULAR BIOLOGY (Academic Press, London), Scopes, (1987), PROTEIN PURIFICATION: PRINCIPLES AND PRACTICE, Second Edition (Springer-Verlag, N.Y.), and HANDBOOK OF EXPERIMENTAL IMMUNOLOGY, VOLUMES I-IV (D.M. Weir and C. C. Blackwell eds 1986). All patents, patent applications, and publications mentioned herein, both supra and infra, are hereby incorporated herein by reference.

The methods of the invention include the preparation of probes generated from at least one hybrid cell line, and enriched for a target gene sequence(s) by subtractive hybridization and cloning, utilizing

transcripts from a nearly isogenic cell line for a subtractive hybridization procedure. The hybrid cell lines used for the preparation of the enriched probes contain genetic sequences from a eukaryotic chromosome that is heterologous to the host cell, but differ primarily in a sequence of nonoverlap which contains a target gene(s) within the heterologous eukaryotic chromosome. As used herein, the term "heterologous eukaryotic chromosome" connotes that the chromosome is derived from a different species than the host cell. The target gene(s) is expressible in the hybrid cell line. Host recipient cells are chosen such that sequence divergence in the 3'-untranslated regions of host cell mRNA and the cDNA to the mRNA of the target gene(s) on the heterologous eukaryotic chromosome is sufficient to prevent heteroduplex formation when hybridization is performed under stringent conditions.

Methods of preparing hybrid cell lines which contain whole or fragments of heterologous eukaryotic chromosomes are known in the art. These techniques include, for example, microcell-mediated transfer of chromosomes into somatic cells (R.E.K. Fournier and F.H. Ruddle (1977), Proc. Natl. Acad. Sci. USA 74:319), somatic cell hybridization (Ruddle and Creagan (1975), Ann. Rev. Genet. 9:407), chromosome-mediated gene transfer (McBride and Ozer (1973), Proc. Natl. Acad. Sci. USA 70:1258), DNA-mediated gene transfer (Wigler et al., (1978), Cell 14:725), and irradiation-fusion techniques (Goss and Harris (1975), Nature 255:680, Benham et al. (1989) Genomics 4:509; Cox et al. (1990), Science 250:245).

In a preferred embodiment, the hybrid cells are created by microcell-mediated chromosome transfer (for a review, see Fournier in Techniques in Somatic Cell Genetics (J.W. Shay, ed., Plenum Publishing Corp., 1982), pp. 309-327). The steps involved in this technique are illustrated in the diagram in Figure 1 (taken from

Fournier (1982), id.). Populations of eukaryotic donor cells containing the gene of interest are incubated under conditions which induce an aberrant mitosis. This can be accomplished by incubating growing cultures in the presence of an inhibitor of microtubule polymerization which is known in the art, for example, colchicine or Colcemid. In particular, nuclear division is perturbed such that many small micronuclei are produced when the cells enter interphase. This micronucleation step serves to partition the donor chromosome complement into discrete subnuclear packets which can be physically isolated from the cells. Enucleation is accomplished using standard procedures, e.g., the micronucleate cells are centrifuged in the presence of cytochalasin B. During centrifugation, the micronuclei are drawn from the cell on long stalks, which subsequently break to yield free microcells. Thus, a microcell consists of a single micronucleus surrounded by a thin rim of cytoplasm and an intact plasma membrane. The isolated microcells are fused with intact recipient host cells. Fusion is accomplished by techniques known in the art, including, for example, techniques which use inactivated Sendai virus or polyethylene glycol (PEG). Under appropriate selective conditions, a fraction of the microcell heterokaryons will proliferate to yield microcell hybrid clones. Such clones typically retain 1-5 introduced donor chromosomes in addition to the recipient host cell chromosome complement.

It is desirable to have hybrid cell clones that retain a single donor chromosome (or fragment) which is fixed in the cells by direct selective pressure. One strategy for fixing different donor chromosomes in a series of microcell hybrid clones is to use wild-type donors in microcell fusions with a series of mutant recipients harboring recessive lesions. Another strategy is to use donor cells containing defined translocations between chromosomes carrying a selectable marker and

other autosomes. In some cases, the selectable marker may reside on the chromosome bearing the target gene. A few human chromosomes naturally carry selectable marker genes (e.g., thymidine kinase or dihydrofolate reductase), but the majority do not. Therefore, it may be desirable to introduce a selectable marker into the chromosomes of the donor cells. Methods for introducing selectable markers into eukaryotic cells, and particularly into mammalian chromsomes, are known in the art. For example, bacterial genes such as the gpt gene and the neo gene can confer selectable phenotypes to mammalian cells. The neo gene, which confers resistance to the antibiotic G418A, is a dominant selectable marker, so that recipient cells with recessive mutations are not required.

Marker genes may be introduced into the cell lines by a variety of techniques including transformation and transduction. One method of transfer, for example, is by calcium phosphate precipitation followed by fusion (Nelson et al. (1984), J. Mol. Appl. Genet. 2:563). Another method of introduction of a selectable marker, for example, the neo gene, utilizes defective amphotropic retroviruses. (Lugo et al. (1987), Molecular and Cellular Biology 7:2814.)

The hybrid cells used for the preparation of an enriched probe population by subtractive hybridization are further selected to obtain cell lines containing the chromosome (or fragment) of interest. A selection criterion is usually a chromosomal marker linked to the gene of interest; preferably the marker allows phenotypic selection of the cells. A number of gene loci have been mapped to different chromosomes; some of these are described in, for example Lugo et al. (1987), id.; Saxon et al. (1985), Mol. and Cell. Biol. 5:140; Athwal et al. (1985), Somatic Cell and Mol. Genetics 11:177; and Siden et al. (1989), Somatic Cell and Mol. Genetics 15:245.

At least two different hybrid cell lines which are nearly isogenic are required for the preparation of a probe population enriched for a desired gene sequence. In these isogenic cell lines, the primary difference between the hybrid cell lines is in the sequences of at least one target gene within the heterologous eukaryotic chromosome. These differences are such that they can be detected. I.e., there is a detectable region of nonoverlap in the heterologous eukaryotic chromosomal sequence in the nearly isogenic hybrid cells. Preferably, at least one of the cell lines used is a hybrid cell containing chromosome fragments (deletion hybrids), wherein the deletion is such that it spans at least part or all of the target gene, i.e., such that a transcript of the gene is not obtained from the hybrid cell (a (-) hybrid cell). At least one other cell line used contains an intact target gene such that it allows transcription of the gene in the hybrid host cell (a (+) hybrid cell). Selection of the appropriate hybrid cell lines may be accomplished, for example, by size analysis of restriction enzyme fragments, and/or by marker analysis.

As used herein, the term "plus (+) hybrid cell" connotes a hybrid cell containing heterologous eukaryotic chromosomal DNA which is comprised of sufficient genetic information from a target gene to allow transcription of the target gene in the hybrid host cell. The term "minus (-) hybrid cell" connotes a hybrid cell which is "nearly isogenic" with the "(+) hybrid cell", and which does not allow transcription of the target gene in the hybrid host cell.

As used herein, "nearly isogenic" means that the transcripts produced from the (+) hybrid cell and (-) hybrid cell are essentially homologous, except for those transcribed from the non-overlap region of the heterologous eukaryotic chromosome.

As used herein, the term "nonoverlap region" means that homology of the heterologous chromosome in the (-) hybrid cells is sufficiently lacking or different so that at least one transcript from this region will not form stable heteroduplexes with cDNA to the mRNA transcribed from the chromosomal sequence in the (+) hybrid cells, when hybridized under stringent conditions. Thus, the nonoverlap region is due to mutations in the sequence of the (-) hybrid cell relative to the analogous sequence in the (+) hybrid cell. Nonoverlap may be due to a variety of mutations, and is preferably due to deletions. The nonoverlap region in the (+) cell encodes the target gene, and may encode several genes. The nonoverlap region will be of a size to encode at least one gene; however, it may encode at least 5 genes, in some cases it may encode at least 20 genes, and in some cases it may encode at least 50 genes.

As used herein, "heterologous chromosome" is defined as a chromosome derived from a different species than that of the host-recipient cell; this chromosome may be derived from any eukaryotic species, and is preferably from a vertebrate species. For example, a heterologous chromosome may be human, bovine, ovine, canine, feline, reptilian, or avian and may be transferred into a recipient rodent or murine cell line.

The recipient "host" cell line may be derived from any species except that of the chromosome to be transferred.

As used herein, a "target gene" is a specific region of a polynucleotide containing a gene sequence to be detected, isolated and/or mapped; this term includes wild-type and mutant genes.

The term "polynucleotide" as used herein refers to a polymeric form of nucleotides of any length, either ribonucleotides or deoxyribonucleotides. This term refers only to the primary structure of the molecule. Thus, this term includes double- and single-stranded DNA,

as well as double- and single stranded RNA. It also includes modified, for example, by methylation and/or by capping, and unmodified forms of the polynucleotide.

Hybrid cells containing chromosome fragments may be prepared by a variety of means including, for example, irradiation and fusion (Benham et al. (1989), supra.), and microcell-mediated transfer of chromosomal material (Leach et al. (1989), Genomics 5:167).

In the Examples, infra., 7AE-27 and 7AD-7 are two microcell deletion hybrid lines which were generated by stably transferring fragments of a human fibroblast chromosome 17 into FTO-2B rat hepatoma cells. Resolution of restriction-digested genomic DNA by field inversion gel electrophoresis revealed that 7AE-27 contained 2-4 MB more human DNA than 7AD-7. Thus, 7AE-27 is a (+) hybrid cell, and 7AD-7 is a (-) hybrid cell. Marker analysis revealed that this size difference due to the nonoverlap region mapped to the Col1A1-D1754 interval on the distal portion of the long arm of chromosome 17.

Probes to be used for gene isolation and/or mapping are enriched for sequences expressed from a region of non-overlap between the (+) cell hybrids and the (-) cell hybrids. The enrichment results from removal of essentially all sequences from the host hybrid cell, except for those derived from the nonoverlap region. By "essentially all" is meant at least 70%, preferably at least 80%, and more preferably at least 90% of the host cell sequences which are common to both (+) and (-) hybrid cell lines. Enrichment is accomplished by subtractive hybridization and cloning.

In one embodiment of the invention subtractive hybridization is accomplished as follows. cDNA is synthesized by oligo-dT priming reverse transcription of poly(A)$^+$ RNA from a (+) hybrid cell line. For convenience, the cDNA is labeled. Labeling may be by any means known in the art (e.g., radiolabeling, fluorescence). The conditions used for synthesis are

those that maximize recovery of DNA fragments from the 3'-ends of mature transcripts. Since, surprisingly, the transcripts from the hybrid host cell 3'-untranslated region are sufficiently divergent from those of the 3'-end of the mature transcripts from the heterologous chromosome, these conditions allow the cloning of the heterologous transcript cDNAs irrespective of homologies in coding sequences between the heterologous gene and the host hybrid cell gene.

The cDNA to the (+) cell transcripts is then hybridized with excess poly(A)$^+$ RNA from (-) hybrid cells under stringent conditions, and heteroduplexes formed as a result of the hybridization are removed from solution. Stringent conditions for hybridization depend upon the length and sequence of the polynucleotides to be hybridized; however, general parameters for stringent conditions are described in Maniatis et al. (1989). Removal of the heteroduplexes may be by means known in the art. For example, the transcripts from the (-) hybrid cells may be affixed to solid supports, or may be tagged with a molecule which allows removal from solution. In a preferred mode, transcripts from the (-) hybrid cells are biotinylated in vitro. The biotin RNA-cDNA heteroduplexes are complexed with streptavidin and removed by phenol extraction. Procedures for biotin-streptavidin subtractions are known in the art (see, for example, Sive and St. John (1988), Nucl. Acids. Res. 18:10837). In order to further enrich the population by removal of hybrid host cell sequences, single-stranded cDNA fragments recovered after hybridization with transcripts from (-) cells may be subjected to further rounds of hybridization with the transcripts.

cDNAs to transcripts from the nonoverlap region can be further isolated and characterized by subsequent screening, including screening of cDNA and genomic libraries from the donor species from which the transferred chromosome (or fragment) is derived. The

positive clones derived from the libraries may then be further screened with a second (+) hybrid cell - (-) hybrid cell subtracted probe and, in parallel with a (-) hybrid cell - (-) hybrid cell self-subtracted probe. Recombinant clones which display differential hybridization can then further screened by rehybridization a third time with a (+) hybrid cell - (-) hybrid cell subtracted probe. The individual hybridizing clones are isolated, and the insert DNA is hybridized to a variety of genomic DNAs, including the species from which the eukaryotic chromosomal DNA in the hybrid cell lines was derived, and preferably also with a sample containing an intact chromosome representative of the inserted heterologous eukaryotic chromosomal material. This type of mapping allows identification of distinct cDNAs encoded within the nonoverlap region.

In order to determine whether the subtractive probe cDNAs from the nonoverlap region are indicative of a specific gene, they can be assessed for their concordance with known genotypes and/or phenotypes. Usually, the cDNAs will be sequenced, and the sequence compared with known sequences in known gene banks, for example, Genbank. In addition, the cDNAs can be tested for their ability to confer phenotypes on recipient host cells. In the event that phenotypic assessment in vitro is not possible, the subtractive probes can be used to determine whether genetic traits are linked to mutations within analogous chromosomal sequences. For example, in humans the presence or absence of mutations within a sequence analogous to an isolated cDNA can be correlated with the presence or absence of the disease or carrier state in genetic family studies.

Another embodiment of the invention are compositions comprised of polynucleotide probes prepared by the above described methods, and synthetic counterparts thereof. Synthetic counterparts may be prepared by chemical synthesis using the polynucleotide

sequence information derived from the probe. Methods of preparing polynucleotides of defined sequence by chemical methods are known in the art. Synthetic counterparts may also be produced by recombinant methods. One or more polynucleotides containing a probe sequence may be introduced into a recombinant vector, and the recombinant vector replicated in a host organism. Cells containing the vector may be cloned, utilizing techniques known in the art. Moreover, if desired, the recombinant vector, and/or polynucleotide contained therein may be isolated from non-polynucleotide components.

The synthetic counterparts of the polynucleotide probes of the method need not contain the entire probe sequence. Rather, they will contain sufficient sequence information and length to form stable heteroduplexes with the target gene when hybridized under stringent conditions. The synthetic counterparts may be comprised of a minimum of 8 polynucleotides of the probe sequence, preferably may be comprised of a minimum of 20 polynucleotides of the probe sequence, more preferably may be comprised of at least 50 nucleotides of the probe sequence, and even more preferably may be comprised of at least 70 nucleotides of the probe sequence.

Still another embodiment of the invention are methods of detecting target genes utilizing the enriched probe preparations described above, or the synthetic counterparts thereof. In order to detect the target gene, a sample suspected of containing the target gene is provided. A probe prepared by the above described methods, or the synthetic counterpart thereof, is also provided. The sample and the probe are reacted under stringent conditions which allow the formation of a heteroduplex between a target gene (if any), and the probe to the target gene. Heteroduplexes formed in the reaction, if any, are detected. Detection may be by means known in the art. For example, the probe may be labeled; labels for polynucleotide probes are known in

the art, and include, for example, radiolabels, fluorescent labels, and labels which form complexes with other molecules (e.g., antigens or antibodies, biotin, etc.).

Probes prepared by the method described herein, and synthetic counteparts therof, may also be used in the detection of target genes which are mutants. The wild type gene detected by the probe is sequenced. If the gene causes a phenotypic difference in cells, the target gene if any, is also isolated from the phenotypically different cells. In this case, however, fragments of the probe which are sufficient to allow heteroduplex formation with the gene sequence are used in the detection. The mutated target gene isolated with the probe is then sequenced. Probes to the mutated gene may then be constructed (by chemical or recombinant synthesis), using the sequence difference from the wild-type as part of the probe sequence. In the event that phenotypic selection is not available at the cell level, genetic studies which examine inherited disorders and/or traits can replace phenotypic selection. For example, family studies can be used to determine if a genetic disease state or tendency is present in certain individuals. Probes to the target gene can then be used for isolation of the putative mutant gene. Sequencing of the putative mutant gene will disclose the site and type of mutation (if any).

The probes to mutant genes will then be useful for the detection of the mutated form in individuals. An "individual", as used herein, refers to vertebrates, particularly members of the mammalian species, and includes but is not limited to domestic animals, sports animals, primates, and humans.

Described below are examples of the present invention which are provided only for illustrative purposes, and not to limit the scope of the present invention. In light of the present disclosure, numerous

embodiments within the scope of the claims will be apparent to those of ordinary skill in the art.

Examples

Tissue-specific extinguisher 1 (TSE1) is a trans-acting locus on human chromosome 17 that down-regulates expression of seven liver genes in hepatoma x fibroblast hybrids. Subtractive cDNA hybridization was used to clone transcripts encoded within a 2-4 Mb segment of chromosome 17 that includes TSE1. High resolution mapping within this region indicated that 8 of 9 different human cDNAs so obtained were distinct from TSE1. The remaining cDNA clone mapped concordantly with TSE1 in a panel of fragment-containing hybrids. DNA sequencing indicated that this cDNA encoded regulatory subunit RIα of cAMP-dependent protein kinase, and RIα mRNA levels correlated with TSE1 activity in various hybrid lines. Stable transfection of wild-type of cAMP-binding mutant RIα alleles into hepatoma recipients produced an extinction phenotype indistinguishable from that encoded by human TSE1. We conclude that TSE1 encodes a regulatory subunit of protein kinase A.

Preparation of Somatic Cell Hybrids Containing Chromosome 17 Fragments

Microcell hybrid clones L(17n)D and L(17n)E were prepared by transferring neo-marked human chromosomes into mouse La^-t^- cells and selecting TK^-, $G418^+$ hybrids (Leach et al., 1989). These L(17n) clones retain a single neo-marked human chromosome 17. The 7A series microcell hybrids were generated by fusing L(17n)D or E microcells with PCTA-7A rat hepatoma recipients (Wynshaw-Boris et al. (1984), Biol. Chem. 259:12161) and selecting $G418^r$ microcell hybrids (Leach et al., 1989). DCR-1 and MH41 are mouse x human hybrids that contain constitutional translocations L(1;17)(p34.3;q11.2) and

(17;19)(q23;p13), respectively Menon et al. (1989), Genomics 5:245; van Tuinen et al. (1987), Genomics 1:374.

Maintenance of Cell Lines

FTO-2B (TK^-, Oua^+) and FAO-1 (HPRT', Oua^+) rat hepatoma cells are derivatives of H411EC3 (Killary and Fournier (1984), Cell 36:523). Diploid MEFs were prepared from C57BL/6J embryos at 12-14 days of gestation according to standard techniques (Kozak et al., 1975). Diploid HSFs were isolated as previously described (Riegner et al. (1976), Tissue Culture Assoc. Man. 2:273. Rat-1 cells are an SV-40-transformed line of rat embryo fibroblasts (Botchan et al. (1976), Cell 9:259).

All cell lines were cultured in 1:1 Ham's F12:Dulbecco's modified Eagle's medium with 5% fetal bovine serum (GIBCO) without antibiotics. L(17n)D, L(17n)E, and 7A series microcell hybrids were grown in media containing 500 μg/ml G418, and MH41 and DCR-1 were grown in media containing HAT. Mycoplasma tests (Chen (1977), Exp. Cell Res. 104:255) performed at intervals, were uniformly negative.

Regional Localization of Human TSE1 on Chromosome 17

In order to localize TSE1 to a specific chromosomal region, we screened a collection of chromosome fragment-containing hybrids for expression of TSE1-responsive genes. This information allowed us to identify hybrid cell lines that were potentially useful for molecular cloning.

The 7A series microcell hybrids (preparation discussed supra.) are rat hepatoma cells that retain various fragments of chromosome 17 derived from diploid human fibroblasts; patterns of human chromosome 17 marker retention in these clones have been reported (Leach et al. (1989), Genomics 5:157-176.). We screened 27 different fragment-containing 7A hybrids for expression

of liver-specific genes by RNA blot hybridization, and identified 16 hybrid clones that were extinguished for expression of tyrosine aminotransferase (TAT), phosphoenolpyruvate carboxykinase (PEPCK), argininosuccinate synthetase (AS), and other TSE1-responsive genes (data not shown). Thus, these hybrids retained human fibroblast-derived TSE1. In contrast, 11 other hybrids in this collection expressed TAT, PEPCK, and AS mRNAs at levels comparable with those of parental rat hepatoma cells, indicating that human TSE1 was not retained. By comparing the chromosome fragments present in the various hybrid lines, we were able to position TSE1 on human chromosome 17. For example, hybrid 7AE-5 expressed TAT, PEPCK, and AS mRNAs, indicating that human TSE1 was not retained. This hybrid clone contained a large fragment of human chromosome 17 with an interstitial deletion: 12 of 13 chromosome 17 marker loci were present, but the PKCA locus was deleted in this clone (Figure 2). This suggests that human TSE1 maps on the long arm of human chromosome 17 in the region of PKCA. Hybrid 7AD-14 retained a centric fragment that included alphoid repeats from the human chromosome 17 centromere (D17Z1) and proximal 17q markers D17S33 and THRA1, plus a distal fragment containing PKCA, D17S4, and TK. This hybrid was extinguished for expression of TSE1-responsive genes. These data also suggest that human TSE1 maps near PKCA, possibly in the PKCA-D17S4 segment. Assignment of human TSE1 to this interval, which is likely located in 17q23-24, was parsimonious with patterns of TSE1 retention among the entire collection of 7A hybrid lines.

Hybrid clones 7AE-27 and 7AD-7 were chosen as starting materials for molecular cloning. 7AE-27 contained two segments of a human chromosome 17, a proximal fragment that contained D17Z1, D17S33, and THRA1, and distal 17q fragment extending from GH1 to TK. This region includes human TSE1 and its proximal and

distal flanking markers. Fluorescence in situ hybridization of labeled human DNA to 7AE-27 metaphases identified a single human chromosome fragment that had been translocated to a rat chromosome (data not shown), resulting in stable retention. This chromosome fragment was approximately 15 Mb in length, representing some 15%-20% of human chromosome 17. 7AD-7 cells also contained a D17Z1-D17S33-THRA1 proximal fragment plus a distal fragment from 17q, but human GH1, PKCA, and TSE1 were not retained. By comparing human NotI and MluI restriction fragments resolved by field-inversion gel electrophoresis of hybrid cell genomic DNA, we estimate that 7AE-27 cells contained 2-4 Mb of human chromosome 17 sequence that were not present in hybrid 7AD-7 (data not shown). We therefore sought to isolate human cDNA clones representing 7AE-27-specific transcripts that were encoded within this region.

Isolation of Human cDNA Clones by Subtractive Hybridization In order to use subtractive hybridization to clone expressed DNA sequences from hybrid cell lines, we prepared an enriched cDNA probe under conditions that maximized recovery of DNA fragments from the 3' ends of mature transcripts. This allowed us to clone human cDNAs from specific regions of human chromosome 17 irrespective of rat-human coding sequence homologies.

Single-stranded, radiolabeled cDNA was synthesized by oligo(dT)-primed reverse transcription of poly(A)$^+$ RNA from 7AE-27 cells under conditions that maximized the yield of cDNA fragments less than 500 bp in length. The cDNA fragments were hybridized in solution with a 20-fold molar excess of poly(A)$^+$7AD-7 RNA that had been biotinylated in vitro (Forster et al. (1985), Nucl. Acids Res. 13:745). The biotin RNA-cDNA heteroduplexes were complexed with streptavidin and removed by phenol extraction (Sive and St. John (1988), Nucl. Acids Res. 16:10937). Single-stranded cDNA fragments were recovered

and used for a second round of subtraction, and the resulting probe, highly enriched for 7AE-27-specific sequences, was used to screen a human skin fibroblast cDNA library. A library of HSF cDNA in λgt10 was plated at high density (40,000 plaques per 150 mm plate) and hybridized with a labeled single-stranded cDNA probe enriched for 7AE-27 sequences by subtractive hybridization. Of approximately 120,000 λgt10 plaques tested in primary, high density screening, 221 were scored as positive.

Phage pools containing the 221 primary positives were rescreened with a second (7AE-27)-(7AD-7) subtracted probe and, in parallel, with a (7AD-7)-(7AD-7) self-subtracted probe, as follows. Inserts from the 221 positive pools were amplified in situ using the polymerase chain reaction (PCR), and duplicate filters containing amplified DNAs were hybridized with the two labeled probes. Thirty-one phage pools displayed differential hybridization. These phage pools were plated at low density and rehybridized with a third (7AE-27)-(7AD-7) subtracted probe. Individual hybridization plaques were isolated, and insert DNA was amplified by PCR. The amplified fragments were labeled and hybridized to Southern blots of PstI-digested rat (FAO-1), human (HSF [human foreskin fibroblast]), and mouse (MEF [mouse embryo fibroblast]) genomic DNA as well as DNA from somatic hybrids containing an intact human chromosome 17 (L(17n)E) or chromosome 17 fragments (7AE-27 and 7AD-7). This allowed us to identify human cDNA clones encoded by genes within the 7AE-27/7AD-7 nonoverlap region. For example, the human cDNA insert of clone λ76 did not cross-hybridize with rat or mouse DNA under stringent conditions, but it hybridized specifically with two PstI fragments of human genomic DNA. Both fragments were detected in PstI-digested DNA from monochromosomal hybrid L(17n)E, indicating that they were derived from human chromosome 17. As these

fragments were also present in 7AE-27 but not 7AD-7, the λ76 cDNA was encoded by a human gene within the nonoverlap region. After mapping all of the candidate cDNA clones in this manner, nine distinct human cDNAs encoded within the 7AE-27-specific region were identified. Some properties of these human cDNAs from the non-overlap region are shown in Table 1.

TABLE 1

Properties of Human cDNAs from the Non-overlap Region

Recombinant Designation	cDNA insert (bp)	mRNA (bp)	Product	%nt Homology	Previous Assignment
λD5	500	800	RPL17a	64 (yeast)	none
λ29	700	2000	γ-actin	98 (human)	17p11-qter
λ69	1800	nd	unknown	--	none
λ76	700	2800	P4HB, THBP, PDI	95 (human)	17q25
λ78	500	1000	unknown	--	none
λ111	600	800	RPL27	84 (rat)	none
λ147	500	900	unknown	--	none
λ157	3500	1500, 3500	PKA-R1α	99 (human)	7
λ180	600	1000	mATP-syn-d	84 (bovine	17

Mapping Candidate cDNA Clones Relative to TSE1

To determine whether any of the nine cDNA clones we isolated represented bona fide candidates for the TSE1 transcript, we prepared a physical map of the chromosome 17 nonoverlap region and mapped the cDNA clones relative to human TSE1. Restriction-digested genomic DNAs from various 7A series microcell hybrids were probed with cosmid markers that we cloned from this region (see infra.), with probes for known marker loci, and with the nine distinct cDNA inserts. Retention of human TSE1 in the hybrids was monitored by scoring phenotypic expression of TAT and PEPCK mRNAs. Cytoplasmic RNAs (5 μg) were size fractionated on an agarose-formaldehyde gel, transferred to a Zetabind membrane, and probed sequentially with radiolabeled rat PEPCK and human α-tubulin DNA probes. These experiments demonstrated that a single candidate cDNA clone mapped concordantly with human TSE1.

Results of these analyses are summarized in Figure 3. The nine newly generated cDNA markers and the six new cosmid markers all mapped distal to COL1A1. This finding indicated that most, if not all, of the 7AE-27-specific nonoverlap region was contained on the distal 17q fragment. Furthermore, 8 of 9 cDNA clones and 6 of 6 cosmid markers mapped between COL1A1 and D17S77, suggesting that the nonoverlap region consisted primarily of sequences from the proximal portion of the distal fragment. This chromosome segment, which includes human GH1, PKCA, and TSE1, is stippled in Figure 3. In total, 16 DNA markers were assigned to this interval, which we estimated to be 2-4 Mb in length. If these markers were distributed uniformly, one would occur in every 125-250 kb segment of genomic DNA. Therefore, concordant segregation of two markers in this interval among fragment containing hybrids suggests tight linkage within a small physical segment.

Eight of the nine different cDNAs were mapped to loci distinct from TSE1 using a panel of ten fragment-containing hybrids (Figure 3). In Figure 3, retention of 17q markers in a panel of fragment-containing hybrids was assayed by Southern blot hybridization. To prepare the figure, marker loci were arranged according to the current 17q map (Fain et al. (1991), Report of the second international workshop on human chromosome 17, Cytogenet. Cell Genet., in press.), and new markers were positioned so as to minimize the number of apparent break points in the hybrids. The chromosome fragment of each hybrid line are not shown to scale, as specific cloning of markers within the 7AE-27/7AD-7 nonoverlap region (stippled) results in apparent expansion of this segment. The markers within brackets segregated concordantly in the hybrids, so that their relative order cannot be established from these data. However, although λ29, λ78, and λ75 showed identical patterns of retention, λ29 is placed proximal to TK and λ75 distal because sequence analyses (data not shown) indicated that these cDNAs were encoded by the human ACTG and P4HB loci, respectively. λ78 was arbitrarily placed in the same interval as λ29. A single cDNA clone, λ157, mapped concordantly with human TSE1.

For six of the cDNA clones, both +/- and -/+ TSE1/cDNA discordancies were observed. The genes encoding λ147 and λ180 mapped together within a chromosome segment just distal to TSE1, a segment retained in the TSE1 hybrid 7AE-9. In contrast, the cDNA insert of λ157 was retained concordantly with human TSE1 in all ten hybrid clones; six hybrids retained both TSE1 and λ157 and four clones retained neither marker. To more critically assess the concordancy between retention of TSE1 and λ157, we assayed both markers in a large panel of 7A series fragment-containing hybrids. Genomic DNA from rat hepatoma cells (FAO-1), human fibroblasts (HSF-113), and 27 different 7A series hybrids were

digested to completion with PstI, size-fractionated on agarose gels, and blotted onto nylon filters. The blots were probed with labeled λ157 cDNA insert to detect human-specific 157 restriction fragments. Sixteen of the hybrid clones retained the human 157 gene; all of these hybrids contained human TSE1. The remaining hybrids retained neither human TSE1 nor human 157. The chromosome 17 genotypes of these hybrids have been reported (Leach et al. (1989)). The studies indicated that the human λ157 gene and TSE1 mapped together within a small physical segment of human chromosome 17. To determine whether the λ157 cDNA was encoded by TSE1, we characterized the λ157 insert.

Identification of the TSE1 Gene Product

λ157 contained a cDNA insert of approximately 3550 bp. EcoRI digestion of λ157 DNA released the insert as two fragments, approximately 850 and 2700 bp in length. Each fragment was subcloned in pBluescriptII, and nucleotide sequences from the ends of each subcloned fragment were determined using the dideoxy chain termination method (Sanger et al. (1977), Proc. Natl. Acad. Sci. USA 74:5463). The derived nucleotide sequences were compared with those in the GenBank sequence database. Both ends of the 850 bp fragment and the 3′ end of the 2700 bp fragment displayed >99% sequence homology with human RIα, a regulatory subunit of protein kinase A (PKA), the cAMP-dependent protein kinase. The 5′ end of the 850 bp fragment contained 11 bp of previously unreported sequence (Sandberg et al., 1987), which apparently derives from the 5′ end of the mature RIα transcript. The 3′ end of the 2700 bp fragment was not homologous to any nucleotide sequence in the database (Figure 4). However, only 3039 bp of human RIα sequence have been reported (Sandberg et al., 1990), while the predominant RIα transcript in several human tissues is approximately 3500 bp in length. Therefore,

it appears that the 5′ end of the 850 bp fragment as well as the 3′ end of the 2700 bp λ157 cDNA fragment contain sequences that have not yet been reported. We note also that the EcoRI restriction site in our cloned cDNA sequence does occur at position 845 bp in the reported RIα sequence (Sandberg et al. (1987), Biochem. Biophys. Res. Commun. 149:939).

The determination that λ157 encoded a regulatory subunit of cAMP-dependent protein kinase was consistent with the possibility that λ157 was the product of human TSE1. In particular, previous studies have shown that TSE1-mediated extinction in microcell hybrids can be reversed by treating the cells with cAMP (Thayer and Fournier, 1989). Furthermore, both TSE1-mediated extinction and cyclic nucleotide induction are transcriptional effects that are mediated through the TAT cAMP response element (CRE) (Boshart et al. (1990), Cell 61:905). Therefore, we considered the possibility that the TSE1 phenotype could be produced by altering expression of a regulatory subunit of PKA.

Steady-state RIα transcript levels in parental and hybrid cells were quantitated by RNA blot hybridization. Cytoplasmic RNAs from rat hepatoma cells (FTO-2B), rat (Rat-1) and human fibroblasts (HSF-113), and 7AE-27 and 7AD-7 microcell hybrids were resolved on agarose-formaldehyde gels and probed with a labeled fragment of mouse RIα cDNA. The 228 bp probe used in this experiment was 98% homologous to rat RIα mRNA and 93% homologous to human RIα mRNA. Rat-1 fibroblasts expressed a major RIα transcript of about 1.5 kb and two minor transcripts 2.9 and 3.3 kb in length. The same RNA species were expressed in FTO-2B rat hepatoma cells, but at much reduced levels, <5% those of Rat-1 cells. Human fibroblasts (HSF-113) expressed two RIα transcripts, a major species 3.5 kb in length and a less abundant 1.5 kb transcript. These transcripts are the products of a single human RIα gene, and they differ in the lengths of

their 3' untranslated regions (Sandberg et al. (1990), Biochem. Biophys. Res. Commun. 167:323). 7AD-7 cells expressed low levels of RIα transcripts, similar to FTO-2B cells, but 7AE-27 cells expressed both rat and human RIα mRNAs. RIα transcript accumulation in 7AE-27 cells was intermediate between those of hepatoma cells and fibroblasts. This suggests that the single human RIα locus of 7AE-27 cells was being expressed at fibroblast-typical levels despite being present in rat hepatoma cells. Thus, human RIα expression was correlated with TSE1 genetic activity in these cells.

Stable Transfection of RIα cDNA Produces the TSE1 Extinction Phenotype

To determine whether overexpression of RIα mRNAs in rat hepatoma cells resulted in an extinction phenotype like that encoded by TSE1, we prepared stable hepatoma transfectants expressing cloned RIα cDNAs. An expression vector, in which the Harvey sarcoma virus long terminal repeat was driving expression of full-length mouse RIα cDNA and the SV40 early promoter was driving expression of the neomycin phosphotransferase gene (Clegg et al., 1987), was introduced into FTO-2B cells by electroporation. Parallel transfections with wild-type RIα cDNA (FRIWT) and an RIα allele with mutations in both cAMP-binding sites (FRIAB) were performed. Cells that had stably integrated the transfected DNAs were selected in medium containing G418. Twelve transfectant clones containing the wild-type RIα expression cassette and 14 clones containing the AB mutant allele were isolated and characterized.

The expression of PEPCK, RIα and α-tubulin mRNAs in the transfectants was analyzed by RNA blot hybridization. Cytoplasmic RNAs from untreated and dibutyryl cAMP-induced rat hepatoma cells (FTO-2B) and stable transfectant clones containing wild-type (FRIWT-8, FRIWT-1) or cAMP-binding mutant RIα transgenes (FRIAB-6,

FRIAB-7) were fractionated on an agarose/formaldehyde gel, blotted, and hybridized sequentially with labeled PEPCK, RIα, and α-tubulin probes. Parental rat hepatoma cells (FTO-2B) expressed readily detectable levels of PEPCK mRNA, and PEPCK expression was inducible by dibutyryl cAMP. Rat RIα transcripts were expressed at very low levels in these cells. FRIWT-8 cells, a transfectant clone expressing high levels of a 2.3 kb transcript from the wild-type murine RIα transgene, were extinguished for basal PEPCK mRNA expression; levels were about 10% those of FTO-2B cells. However, treating FRIWT-8 cells with dibutyryl cAMP reversed RIα-mediated extinction, as PEPCK mRNA accumulated to levels comparable with induced FTO-2B cells. This phenotype, a 10- to 100-fold reduction in basal PEPCK mRNA expression and reversal of extinction by cAMP, is indistinguishable from that of TSE1-containing microcell hybrids (Thayer and Fournier (1989), Mol. Cell. Biol. 9:2837). Another $G418^r$ transfectant clone from this experiment, FRIWT-1, failed to express the RIα transgene although it did contain an apparently intact donor plasmid. Expression of PEPCK mRNA in uninduced or cAMP-treated FRIWT-1 cells was comparable with parental FTO-2B cells. Thus, the presence of the transfected plasmid without RIα expression did not affect basal or cAMP-stimulated PEPCK mRNA expression. Transfectant clone FRIAB-6 expressed mutant RIα molecules that were unable to bind cAMP. Basal PEPCK expression was also extinguished in these cells, and dibutyryl cAMP treated partially reversed extinction. A sister clone that failed to express the mutant RIα transgene, FRIAB-7, was not extinguished for PEPCK mRNA expression. As forced expression of RIα produces an extinction phenotype indistinguishable from that encoded by TSE1, we conclude that RIα is the TSE1 gene product.

Probe Preparation

Cytoplasmic RNA (Favaloro et al. (1980), Meth. Enzymol. 65:718) from 7AE-27 or 7AD-7 cells was isolated and poly(A)$^+$ RNA was selected using oligo(dT)-cellulose columns (Aviv and Leder (1972), Proc. Natl. Acad. Sci. USA 69:1408). Poly (A)$^+$ 7AD-7 RNA was biotinylated *in vitro* with photoreactivable biotin (Clontech) (Forster et al., 1985). Two cycles of photobiotinylation were performed. Radiolabeled cDNAs were prepared essentially as described by Davis et al. (1987), Proc. Natl. Acad. Sci. USA 57:587. Two micrograms of poly (A)$^+$ 7AE-27 or 7AD-7 RNA was incubated in 30 μl of 50 mM Tris (pH 8.3), 10 mM $MgCl_2$, 150 mM KCl, 1.0 mM dGTP. 1.0 mM dATP, 1.0 mM TTP, 40 μM [α-^{32}P]dCTP (800 Cl/mmol, New England Nuclear), 100 μg/ml oligo(dT)$_{12-18}$, 10 mM DTT, 10U of RNasin (Promega), 20 U of AMV reverse transcriptase (Life Sciences) for 30 min at 42°C. The specific activities of the cDNAs so obtained were approximately 1.5 x 10^8 cpm/μg.

Solution hybridizations of radiolabeled, single-stranded cDNAs with biotinylated RNAs were performed as described by Sive and St. John (1988). Approximately 500 ng of radiolabeled cDNA was hybridized with a 20-fold molar excess of biotinylated poly(A)$^+$ 7AD-7 RNA driver. The nucleic acids were dissolved in 5 μl of 50 mM HEPES (pH 7.6), 0.2% SDS, 2 mM EDTA, 500 mM NaCl (subtraction buffer) and incubated at 65°C for 48 hr. Under these conditions, C_ot values greater than 1000 were obtained. Each hybridization mixture was diluted to 100 μl with subtraction buffer without SDS, and 10 μg of streptavidin was added. The cDNA-biotinylated RNA-streptavidin ternary complexes were removed by three successive phenol-chloroform extractions. The remaining single-stranded cDNA molecules were recovered and hybridized with a second aliquot of biotinylated 7AD-7 RNA (:0 μg) as described above. The single-stranded cDNA remaining after two rounds of

subtraction generally contained 1.1 x 10^7 to 3.5 x 10^7 cpm and represented 3% to 13% of the starting material.

cDNA Library Screening

An HSF cDNA library in λgt10 (Clontech HL 1052a) was screened with the subtracted cDNA probe, as follows. Approximately 2 x 10^4 to 4 x 10^4 recombinant phage were plated into each of a series of 150 mm petri dishes and replicate nylon filters (Hybond, Amersham) were prepared essentially as described (Benton and Davis, (1977) Science 196:180. The DNA was immobilized on the filters by UV cross-linking. The filters were prehybridized for 2 or more hr at 42°C in 4 ml of 80% formamide, 5 x SSPE (1 x 8SPE - 150 mM NaCl, 10 mM $NaPO_4$ 1 mM EDTA [pH 7.4]), 5 x Denhardt's solution (1 x Denhardt's solution = 0.02% ficoll, 0.02% polyvinyl pyrrolidone, and 0.02% BSA), 1% SDS, 10 μg/ml poly(A), and 10μg/ml poly(C). Approximately 0.15 x 10^7 to 1.0 x 10^7 cpm of enriched probe was denatured by boiling and added directly to the hybridization buffer. Hybridizations were allowed to proceed for 24 hr at 42°C. The filters were washed sequentially in 2 x SSC (1 x SSC - 150 mM NaCl, 15 mM sodium citrate], 0.1% SDS at room temperature for 20 min, 0.2 x SSC, 0.1% SDS at room temperature for 30 min, and 0.1 x SSC, 0.1% SDS at 65°C for 30 min. Autoradiography was for 11 days at -70°C using Kodak XAR film and one intensifying screen.

Agarose plugs (5 mm) containing each positive plaque were picked and stored in 500 μl of 50 mM tris (pH 7.4), 100 mM NaCl, 10 mM $MgSO_4$. Each plug from primary, high density screening contained a number of recombinant phages with different cDNA inserts. These inserts were amplified in situ by PCR using the following primers from each vector arm: 434R (5'-GCTTATGAGTATTTCTTCCAGGGTA-3') and 434L (5'-TGAGCAAGTTCAGCCTGGTTAAGTC-3'). Twenty-five cycles of amplification were performed, with denaturation at 94°C for 1 min, annealing at 60°C for 2 min, and

elongation at 72°C for 4 min. The products from each PCR reaction were precipitated in 400 mM sodium acetate, 60% ethanol and dissolved in 10 mM Tris (pH 7.4), 1 mM EDTA. Duplicate slot blots were prepared using Nytran filters (Schletcher and Schuell). One filter was hybridized with a second, independently prepared subtracted probe and the other with a (7AD-7)-(7AD-7) self-subtracted probe; both filters were washed as described above. Autoradiography was for 4 to 7 days at -70°C.

Secondary positives were plated at low density and plaque purified after hybridization with a third subtracted probe using the conditions described for primary screening. cDNA inserts from individual plaques were obtained by PCR amplification, as described above.

Nucleic Acid Isolation and Blot Hybridization

High molecular weight genomic DNA was isolated (Bell et al., Proc. Natl. Acad. Sci. USA (1981) 78: 5753), digested to completion with PstI, and size fractionated on 0.7% agarose gels. The DNA was transferred to Zetabind (Cuno) by the method of Southern (1975), J. Mol. Biol. 96:503. The filters were prehybridized for 2 or more hr at 42°C in 50% formamide, 5 x SSPE, 5 x Denhardt's solution, 1% SDS, 10 μg/ml poly(A), and 10 μg/ml poly(C). Random hexamer-primed DNAs were radiolabeled to specific activities of approximately 5 x 107 cpm and hybridized with membrane-bound DNA for 20 hr. The filters were washed sequentially in 2 x SSC, 0.1% SDS at room temperature for 30 min, 0.2 x SSC, 0.1% SDS at room temperature for 30 min. and 0.1 x SSC, 0.1% SDS at 65°C for 30 min. Autoradiography was for 2-4 days at -70°C with one intensifying screen using Kodak XAR film of Hyperfilm-MP (Amersham).

Cytoplasmic RNA was size fractionated on 1.2% agarose-formaldehyde gels, transformed to Zetabind, and immobilized by UV crosslinking. A multimerized 228 bp

Hphi-EcoRI fragment of mouse RIα cDNA, corresponding to nucleotides 817-1045 (G. Stanley McKnight, unpublished data), or cloned DNA sequences from the α tubulin (Kα-1, Cowan et al. (1983), Mol. Cell. Biol. (1983) 56:1738) or PEPCK (pCK-10, Yoo-Warren et al. (1983), Proc. Natl. Acad. Sci. USA 80:3658) genes were used as probes. The filters were hybridized and washed as described above, except that the RIα probe was hybridized at 37°C and washed in 0.5 x SSC, 0.1% SDS at 60°C for 30 min.

Isolation of Cosmid DNA Markers

A library of 7AE-27 DNA in cosmid vector pWE15 (Wahl et al. (1987), Proc. Natl. Acad. Sci. USA, 84:2160) was screened with radiolabeled human genomic DNA, and cosmid clones containing human repetitive DNA were isolated. Individual cosmid clones were labeled and used to probe genomic Southern blots of 7AE-27 and 7AD-7 DNA in the presence of unlabeled human DNA competitor, as described (Sealey et al. (1985), Nucl. Acids Res. 78:1905). Cosmids containing 7AE-27-specific sequences were digested with EcoRI and unique sequence DNA fragments were identified by probing Southern blots of the cosmid fragments with radiolabeled human DNA. Repetitive element-free fragments from each cosmid were isolated and subcloned into pUC18 (Yanisch-Perron et al. (1985), Gene 33:103). The subcloned inserts were used to probe Southern blots of 7AE-27 and 7AD-7 genomic DNA to confirm their location within the human chromosome 17 nonoverlap region.

DNA Sequencing

Phage cDNA inserts were digested to completion with EcoRI and subcloned in pBluescriptII (Stratagene) using standard techniques (Maniatis et al. (1982), Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory, New York)). Recombinant plasmids were isolated (Birnboim and Daly (1979), Nucl. Acids Res.

7:1513), and used as templates for dideoxy sequencing as described (US Biochemical Corp., Sequenase Version 2.0). Nucleotide sequence homology searches were performed using the GenBank nucleotide sequence database (release 65.0, September 1990).

DNA Transfection

The RIα expression vectors used in these experiments were prepared by Clegg et al. (1987), J. Biol. Chem. (1987) 262:13111, $pHLREV_{wt}$ neo and $pMLREV_{wt}$ neo each contain a neomycin-phosphotransferase gene linked to the SV40 early promoter and an RIα cDNA fragment inserted between the Harvey sarcoma virus long terminal repeat and a hepatitis B virus polyadenylation sequence. $pHLREV_{wt}$ neo encodes a wild-type RIα protein, while $pMLREV_{wt}$ neo encodes an RIα protein with mutations in both cAMP-binding sites. Twenty micrograms of each plasmid was linearized with PvuI, an enzyme with a single recognition site in the ampicillin-resistance gene. The linearized plasmids were incubated for 5 min with 9×10^6 FTO-2B cells in ice-cold phosphate-buffered saline. The suspension was electroporated using standard procedures (Chu et al. (1987), Nucl. Acids Res. (1987) 15:1311) with 960 μF and 300 V. The electroporated cells were plated in 1:1 Ham's F12: Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum, 100 U/ml penicillin, and 100 μg/ml streptomycin. After 48 hr, 500 μg/ml G418 was added to the medium to select for neo gene expression. G418-resistant clones were isolated and expanded. The cells were either mock induced with serum-free media or treated with 0.5 mM dibutyryl cAMP and 1.0 mM theophylline in serum-free medium for 16 hr and cytoplasmic RNA was isolated and analyzed as described above.

Industrial Significance

The methods of the invention for the preparation of probes are useful for providing probes to genes which can be of agricultural and/or veterinary or medical significance. The probes can be used for the detection and isolation of genes which encode pharmacologically active molecules, for genes which provide resistance to disease or confer disease, or for genes which provide for the overproduction of desired molecules.

Claims

1. A method of preparing a probe for detection of a target eukaryotic gene sequence comprising:

(a) providing a composition containing polynucleotides with sequences complementary to those of transcripts from a (+) hybrid cell, wherein the (+) hybrid cell contains heterologous eukaryotic chromosomal DNA comprised of a target gene sequence, and wherein the polynucleotide is comprised of a sequence complementary to the 3'-end of the transcript;

(b) providing a composition comprised of polynucleotides containing the sequences of transcripts from a (-) hybrid cell, wherein the (-) cell contains a non-overlap region in the heterologous chromosomal DNA, and wherein the non-overlap region encompasses the target eukaryotic gene sequence;

(c) incubating the composition of (a) with the composition of (b) under stringent conditions which allow the formation of heteroduplexes;

(d) isolating polynucleotides of (a) which are not in heteroduplexes with the polynucleotides of (b).

2. The method of claim 1, further comprising:

(a) providing a genomic library from the species from which the eukaryotic chromosomal DNA was prepared;

(b) incubating the isolated polynucleotides of 1(d) with the genomic library of (a) under stringent conditions which allow the formation of heteroduplexes; and

(c) isolating the polynucleotide heteroduplexes formed in (b).

3. The method of claim 2, further comprising:

(a) providing clones containing the polynucleotides from the heteroduplexes of 2(c); and

(b) selecting clones which display differential hybridization with a (+) hybrid cell - (-) hybrid cell subtracted probe and a (-) hybrid cell - (-) hybrid cell self-subtracted probe;

(c) isolating the individual selected clones.

4. The method of claim 3, further comprising:

(a) providing a sample containing a polynucleotide with the sequence of an intact chromosome representative of the inserted heterologous eukaryotic chromosomal DNA of claim 1;

(b) incubating the polynucleotides of the selected clones of 3(b) with the sample of (a) under stringent conditions which allow the formation of heteroduplexes between the polynucleotide of (a) and the polynucleotides of (b), if any; and

(c) isolating heteroduplexes formed in (b).

5. The method of claim 1, wherein heterologous eukaryotic chromosomal DNA comprised of a target gene sequence is from human chromosome 17.

6. The method of claim 5, wherein the target gene sequence is a TSE1 gene.

7. A composition comprised of a polynucleotide probe prepared by the method of claim 1, and synthetic counterparts thereof.

8. A composition comprised of a polynucleotide probe prepared by the method of claim 2, claim 3, claim 4 or claim 5, and synthetic counterparts thereof.

9. A method of detecting a target gene comprising:

(a) providing a polynucleotide probe composition according to claim 7 or claim 8;

(b) providing polynucleotides from a sample suspected of containing the target gene;

(c) incubating the polynucleotide probe composition of (a) with the polynucleotides of (b) under stringent conditions which allow the formation of heteroduplexes between the polynucleotides of (a) and (b), if any; and

(d) detecting heteroduplexes formed in (c), if any.

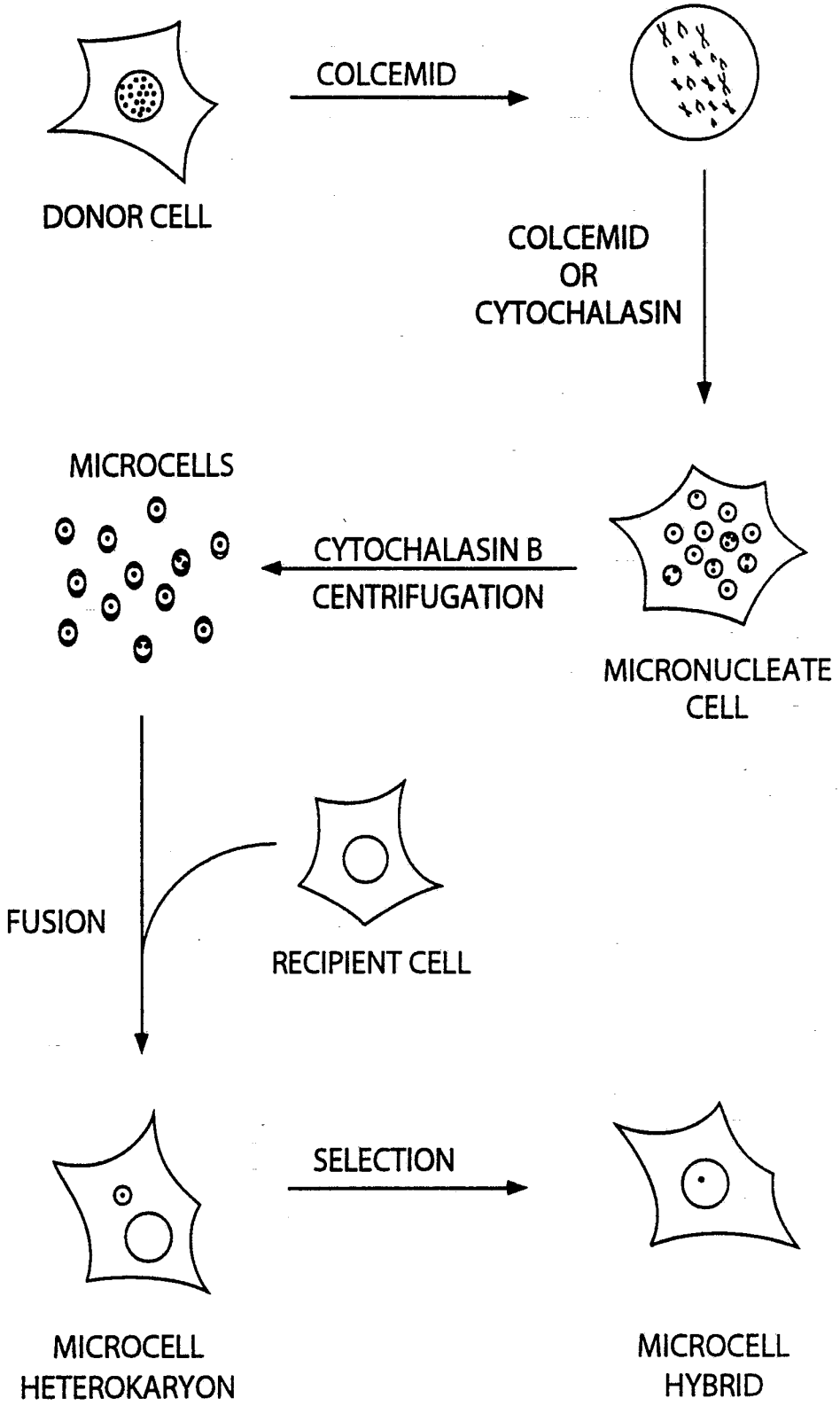
COLCEMID
DONOR CELL
COLCEMID
OR
CYTOCHALASIN
MICROCELLS
CYTOCHALASIN B
CENTRIFUGATION
MICRONUCLEATE
CELL
FUSION
RECIPIENT CELL
SELECTION
MICROCELL
HETEROKARYON
MICROCELL
HYBRID

FIG. 1

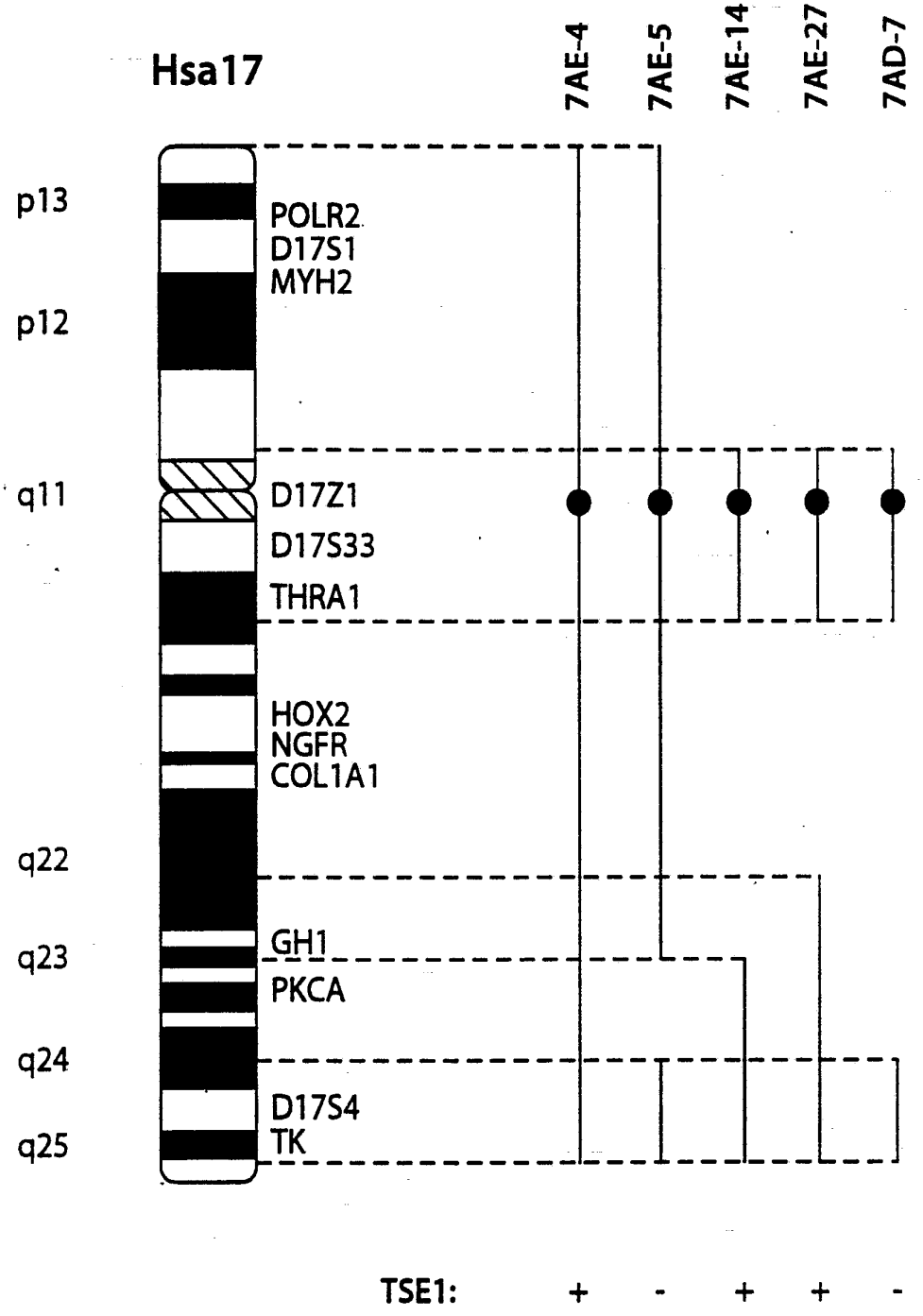
Hsa17
7AE-4
7AE-5
7AE-14
7AE-27
7AD-7
p13
p12
q11
q22
q23
q24
q25
POLR2
D17S1
MYH2
D17Z1
D17S33
THRA1
HOX2
NGFR
COL1A1
GH1
PKCA
D17S4
TK
TSE1: + - + + -

FIG. 2

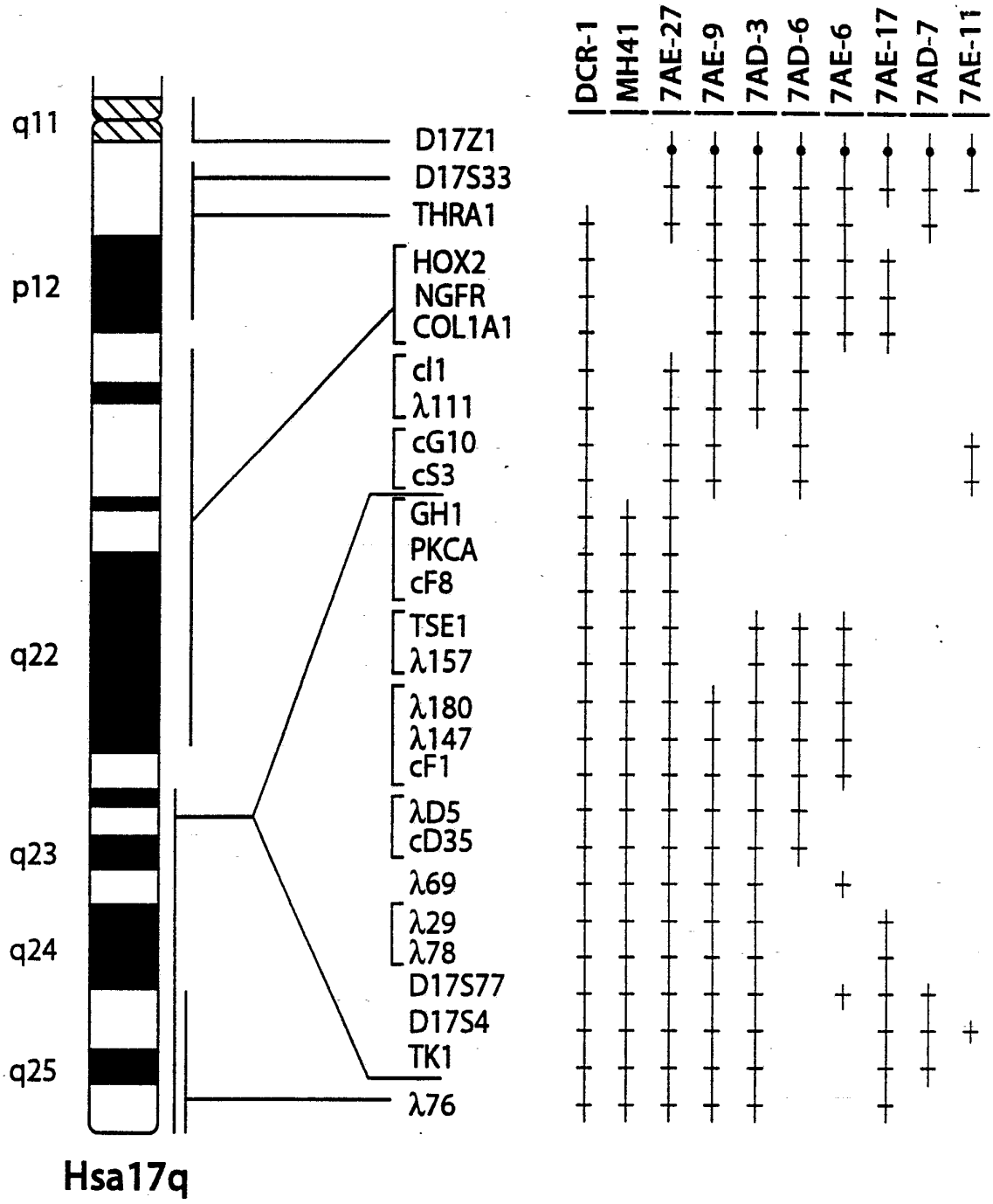
DCR-1
MH41
7AE-27
7AE-9
7AD-3
7AD-6
7AE-6
7AE-17
7AD-7
7AE-11
q11
p12
q22
q23
q24
q25
D17Z1
D17S33
THRA1
HOX2
NGFR
COL1A1
cI1
λ111
cG10
cS3
GH1
PKCA
cF8
TSE1
λ157
λ180
λ147
cF1
λD5
cD35
λ69
λ29
λ78
D17S77
D17S4
TK1
λ76
Hsa17q

FIG. 3

```
TAATTTGTCT TAGTTCATAT TCAAGGCTTT AAAAGTCATT ATTCCTGGGC AAGGTAAGTG    60

AATTTATTGAG ATTTACTGCT CTAGAAAGAT AGATGGCCAA AGGACCGTTT TGTATTGCTT   120

CCTGATTACC AGTCTGATTA TACCATGTGT GCTAATATAC TTTTTTGTTA TAGATTGTCT   180

TAATGGTAGG TCAAGTAATA AAAAGAGATG AAATAAAAAA AAAAAAAAA               229
```

FIGURE 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US92/07516

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) :C12N 15/10; C12Q 1/68; C07H 21/04
US CL :435/6; 536/27
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. :

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

BIOSIS, MEDLINE, APS,
search terms: subtractive hybridization, cDNA, heterologous, ribonucleic acid, transcription

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Science, Volume 246, issued 10 November 1989, P. Liu et al., "Isolation of Human Transcribed Sequences from Human-Rodent Somatic Cell Hybrids", pages 813-815, see entire document.	1-9
Y	J. Sambrook, et al., "MOLECULAR CLONING, A LABORATORY MANUAL", published 1989 by Cold Spring Harbor Laboratory Press (N.Y.), see pages 8.48 and 10.40-10.43.	1-9
Y	Genomics, Volume 5, issued 1989, R. Leach et al., "Physical Mapping of Human Chromosome 17 Using Fragment-Containing Microcell Hybrids", pages 167-176, see entire document.	5
Y	Proc. Natl. Acad. Sci. USA, Volume 85, issued October 1988, J. Lem et al., "Coordinate regulation of two genes encoding gluconeogenic enzymes by the trans-dominant locus Tse-1", pages 7302-7306, see entire document, especially Table 1.	6

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be part of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search
13 November 1992

Date of mailing of the international search report
02DEC 1992

Name and mailing address of the ISA/
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. NOT APPLICABLE

Authorized officer
DAVID SCHREIBER
Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)★