

**Study of the disease associated genes on the long arm  
of chromosome 16, at the region frequently loss in  
breast cancer**

**CHATRI SETTASATIAN**  
**B.Sc. (Medical Technology)**  
**M.Sc. (Biochemistry)**

**A thesis submitted in fulfilment of the requirements for  
the Degree of Doctor of Philosophy**

**Department of Paediatrics**  
**The University of Adelaide**  
**Adelaide, Australia**

**July, 2003**

## *Table of Contents*

	<b>Page</b>
Summary	I
Declaration	V
List of Publications	VI
Acknowledgements	VII
Abbreviations	IX
<b>Chapter 1: Literature Review</b>	1
<b>Chapter 2: Materials and Methods</b>	54
<b>Chapter 3: Characterization of Transcription Unit 3 (T3)</b>	86
<b>Chapter 4: Characterization of Transcription Unit 12 (T12) and Mutation Study of Spastic Paraplegia gene, <i>SPG7</i></b>	118
<b>Chapter 5: Exon Amplification Study and Characterization of Transcription Unit 13 (T13)</b>	145
<b>Chapter 6: Characterization of Transcription Unit 18 (T18)</b>	180
<b>Chapter 7: General Discussion</b>	190
<b>References</b>	195
<b>Appendix: Publication</b>	

## Summary

The loss of the long arm of chromosome 16 has commonly been observed in the sporadic breast and prostate cancers. Loss of heterozygosity (LOH) studies have identified the telomeric region at 16q24.3 to be one of the regions frequently lost in early stage breast cancer, suggesting the presence of tumor suppressor gene(s) involved in breast tumorigenesis. Breast cancer is the most common malignant tumor among women and comprises up to 18% of all cancers occurring in women. Due to the high incidence of breast cancer and the correlation of this cancer with the LOH on 16q, identification of the tumor suppressor gene(s) in this region has been targeted which will provide results applicable for diagnostic and therapeutic approaches. To this end, a detailed physical and transcription map of the 16q24.3 LOH has been established and used as a framework for the identification of candidate genes.

The primary aim of this study was to clone the genes on this chromosome region and identify the candidate gene that may be involved in the development and progression of breast cancer. A number of genes that were characterized showed no evidence of involvement in breast cancer but showed homology to other known functional proteins. Some of these genes were predicted to have significant physiological function and one was shown to be involved in one form of genetic disease.

The initial study was the identification of a minimal region of genomic loss at 16q24.3 in breast cancer. Collaborative studies of detailed LOH of a 16q24.3 region were undertaken in breast cancer cell lines and breast cancer samples. It was shown in some of the tumor samples that LOH is restricted to less than 1 Mb encompassing the 16q24.3 physical map. Subsequently, two genes were characterized from the tentative transcripts mapped to this region.

The first gene, designated transcription unit 3 (*T3*), codes for a protein containing two known functional motifs with sequence homology to those of known protein<sup>(3)</sup> predicted to have a function involved in transcription regulation. To study if this gene<sup>5</sup> involved in breast cancer development, mutation analysis was undertaken in a set of tumor samples showing 16q24.3 LOH. No evidences<sup>7</sup> of mutations<sup>8</sup> were detected and therefore this gene is unlikely to be a tumor suppressor, at least for breast cancer. A partial sequence of the mouse homologue of this gene has been reported and its expression appeared to be developmentally regulated. To extend this study, a cDNA clone of the mouse homologue incorporating the complete coding sequence was isolated and used to study its expression in mouse embryos. Whole body *in situ* hybridization was conducted in mouse embryos and revealed the expression of this gene predominantly in the developing nervous systems of the embryo. The results from this study together with that from a previous report suggested that this gene may be important in the embryonic development.

The second gene, by analysis of its protein sequence, was predicted to have function equivalent to the yeast mitochondrial metalloprotease necessary for mitochondrial protein complex assembly. This gene is found to be identical to a published gene, *SPG7*, involved in one type of neurodegenerative diseases, the hereditary spastic paraplegia (HSP). Subsequent studies were undertaken. The genomic structure of *SPG7* was characterized from a set of overlapping genomic clones and was used to study gene mutations in a number of patients with spastic paraplegia, both hereditary and sporadic cases. Mutations were detected in one HSP family and the only affected case of other family. This study showed that *SPG7* mutations involve in a small subgroup of HSP patients, consistent with the known genetic heterogeneity of this disorder.

Since other genes identified at 16q24.3 also showed no evidence of mutations in breast cancer, the region was extended by analysis of a flanking large genomic clone. Exon

trapping experiments were then used to identify possible genes within this genomic segment. Two groups of trapped exons were selected and characterized. By utilizing several molecular biology techniques, as well as public sequence database homology searches, a large cDNA with a complete open reading frame was assembled. This gene, assigned as transcription unit 13 (*T13*), encodes a high molecular weight protein containing nuclear localization signals. *T13* also contains a set of repeat sequence motifs similar to those of the BRCA1-associated ring domain protein (BARD1) as well as those of the inhibitors of nuclear factor *kappa* B (*I- $\kappa$ B*). Since mutations of BRCA1 are associated with the development of hereditary breast cancer, and BARD1 is required for its function, *T13* was speculated to be one of the components involved in the pathway of BRCA1 function. However, a mutation study failed to detect any tumor-restricted alterations in the *T13* sequence in a set of breast cancer samples with 16q24.3 LOH. This may indicate that *T13* is not the target of mutation inactivation in breast cancer. Alternatively, inactivation of *T13* may be by a mechanism other than gene sequence mutation. ✓

In addition to *T13*, another gene on the same genomic segment was also characterized. This gene, designated transcription unit 18 (*T18*), resulted from the assembly of DNA sequence from a number of partial cDNA sequences deposited in public database, from reverse transcription cDNA amplification products, and from direct sequencing of a cDNA clone. *T18* codes for a protein member of AMP-binding protein, likely belonging to the protein family of fatty acyl CoA ligase. Given the possible function of *T18* in fatty acid metabolism, this gene was not considered to be a tumor suppressor and therefore was not subjected to the mutation analysis in breast cancer.

Although this study failed to identify a candidate tumor suppressor gene involved in breast cancer, the possible functions of the genes which were characterized will provide information for further biological studies for their involvement in other genetic disorders.

Among these, SPG7 is an example of a gene that is involved in a genetic disease. The *T3* gene, which is likely to be involved in embryonic development, may also be the target of a developmentally related genetic disorder or be mutated in other types of tumors. T13, due to its possible association with BRCA1 function and/or relation with the I- $\kappa$ B, may be involved in cancer progression but is targeted by other mechanism of gene inactivation. Continuation of the functional studies of both *T3* and *T13* by others are underway and is beyond the scope of this thesis.

### Declaration

This work contains no material which has been accepted for award of any other degree or diploma in any University or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text

I give consent to this copy of my thesis, when deposited in the University library, being available for loan and photocopying

Signed: \_\_\_\_\_ Date: 20 July 2003

## Acknowledgments

The study presents<sup>ed</sup> in this thesis was conducted in the Department of Cytogenetics and Molecular Genetics, The Women's and Children's Hospital, in Adelaide. I am very much grateful to the Department for providing me all facilities and resources that allow<sup>ed</sup> the initiation and completion of all research work.

I would like to express my respectful thanks to my supervisors, Associate Professor David Callen and Professor Grant Sutherland for their support, encouragement, and expert supervision throughout this work. I also wish to thank them for critical review of this thesis. Thankyou also to the Department of Paediatrics, The University of Adelaide for their coordination of all aspects of the PhD program.

I am extremely grateful to The Royal Thai Government for providing a scholarship and support during my study in Australia.

I also thank all members of the Department who were always very helpful in both laboratory works and other hand-on works while I was in the Department. In particular, I like to thank Dr. Scott Whitmore and Dr. Jozef Gecz for providing their appreciative technical knowledge and expert advice throughout the course of this study.

I also would like to thank Dr. Anne-Marie Cleton-Jansen in Leiden for the LOH data and for providing the precious breast tumor DNA samples; Dr. Ram Seshadri, Sandra Goldup, and Brett McCallum from the Flinders medical Centre in Adelaide for additional LOH data as well as breast tumor DNA samples; Dr. Timothy Cox, and Sonia Donati of



the Department of Genetics, the University of Adelaide for both materials and technical assistance in whole-body in situ hybridization study of the mouse embryos.

Finally I would like to thank my family, in particular my parents for their genuine love and support. I also thank my wife Nongnuch who always provides help and support, her love, patience, and truly understanding are extremely appreciativ<sup>ed</sup>.

## Abbreviations

BAC:	bacterial artificial chromosome.
bp:	base pairs.
BLAST:	basic local alignment tool.
cDNA:	complementary deoxyribonucleic acid.
cM:	centimorgan.
CGH:	comparative genomic hybridization.
dbEST:	database of expressed sequence tags.
dNTP:	deoxynucleotide triphosphate.
DNA:	deoxyribonucleic acid.
DCIS:	ductal carcinoma <i>in situ</i> .
EST:	expressed sequence tag.
ET:	trapped exon.
FAA:	Fanconi anemia complementary group A.
FAB:	Fanconi anemia/Breast cancer.
FISH:	fluorescence <i>in situ</i> hybridization.
HEX:	1-fluorescent 4,7,2'4'5'7',-hexachloro-6-carboxyfluorescein
Kb:	kilobase pairs.
LCIS:	lobular carcinoma <i>in situ</i> .
LOH:	loss of heterozygosity.
Mb:	megabase pairs.
µg:	microgram.
µl:	microlitre.
mg:	milligram.

ml:	millilitre
mRNA:	messenger ribonucleic acid.
NCBI:	National Center for Biotechnology Information.
ng:	nanogram.
ORF:	open reading frame.
PAC:	P1 (phage) artificial chromosome.
PCR:	polymerase chain reaction.
PSI-BLAST:	position-specific iterating BLAST.
PSSM:	position specific scoring matrix.
RACE:	rapid amplification of cDNA ends.
RFLP:	restriction fragment length polymorphism.
RNA:	ribonucleic acid.
RT-PCR:	reverse transcription polymerase chain reaction.
SSCP:	single stranded conformation polymorphism.
STRP:	short tandem repeat polymorphism.
STS:	sequence tagged site.
THC:	tentative human consensus sequence.
TIGR:	The Institute of Genomic Research.
UTR:	untranslated region.
VNTR:	Variable number of tandem repeats.
WCH	Women's and Children's Hospital
YAC:	Yeast artificial chromosome.

# **Chapter 1**

## **Literature Review**

---

---

1.6.1 Construction of the Physical and Transcription map	50
1.6.2 Studies of the Genes within the 16q24.3 minimal LOH region	51
<b>1.7 Aims and Scope of The Study</b>	52

## **1.1 Introduction**

Genes are the fundamental units of every cell governing all cellular functions. Genes are encoded by DNA, the basic biomolecule of the genome. The genetic information stored in DNA dictates the structure of every gene product and delineates every part of the organism. In eukaryotic cells, the genome is organized in the form of chromosomes situated within the nucleus. Genes express their functions by encoding messages through mRNA transcriptions which are then translated into functional proteins. This central dogma of genetics was first proposed by Francis Crick in 1957. The final messages of the genes, proteins, then exert their functions in determining several cellular phenotypes both morphology and physiology. In multicellular organisms, complex organization of the cells and the formation of specific tissues and organs require that gene expression are strictly controlled, such that certain proteins are produced in the right cell, in the right amount, and at the right time. Any deviation of this control may lead to a number of adverse conditions that affect normal physiological processes.

Variations of the genes exist in human as well as other organisms and lead to the phenotypic variations among individuals. Some human phenotype such as red-green color blindness may be regarded as normal variants rather than a clinical disorder. However, some gene variants may cause phenotypes that perturb the normal physiology and development. These types of variant, therefore, associate with disorders or disease syndromes and are generally described as “disease associated genes” or “disease genes”. Some genes are indispensable to embryonic function, so that deleterious mutations result in embryonic lethality and are therefore unrecorded in humans. Some variants that cause abolition of gene function may normally have no effect on the phenotype because other non-allelic genes also supply the same function, so called “genetic redundancy”. Other forms of gene defect

disrupt the cellular systems that control normal cell growth and differentiation. Such defect can lead to uncontrolled cell proliferation and subsequent neoplastic transformation.

The identification of genes associated with human diseases is important for the investigation of cause and mechanism of the disease, which can lead to the development of therapeutic intervention as well as of diagnostic application. To date, only a limited percentage of the genes causing more than 5,000 Mendelian disorders has been identified (McKusick, 1998). A number of identification approaches have been used to identify such disease genes. Before 1980, only a few human disease genes had been identified, these involved a number of diseases with a known biochemical basis where purification of the gene product was possible. Advances in recombinant DNA technology and the establishment of the Human Genome Project, in the mid-1980s, allowed new approaches for disease gene identification. In the last few years, the candidate gene and the positional cloning approaches have played a major role in disease gene identification. The candidate gene approach relies on partial knowledge of the disease gene function. This is based on the availability of previously identified genes, whose features such as sequence domain and expression pattern suggest that they may be implicated in the disease. In positional cloning, the isolation of target gene relies exclusively on the map position of the disease locus in the genome. The resources from the Human Genome Project have greatly facilitated this map-based gene discovery. These include the development of new polymorphic genetic markers and their physical locations on each human chromosome, which are necessary for the genetic mapping of disease loci. In addition, a large collection of partially sequenced cloned cDNAs known as express sequence tags (ESTs) and the collection of complete transcribed sequences of the hypothetical genes deposited in public sequence database, have also been valuable resources for gene identification. Another disease-gene identification strategy, a position candidate gene approach, has become an alternative for positional cloning. This

combines knowledge of the map position of the disease locus with the availability of candidate genes mapped to the same chromosome region.

Human chromosome 16 is one of chromosomes having been targeted for disease gene identification. By linkage analysis, a number of disease loci have been mapped to this chromosome, including the loci for familial Mediterranean fever (FMF) at 16p13.3 (Pras *et al.*, 1992), Crohn's disease at 16q12 (Hugot *et al.*, 1996), fanconi anemia complementation group A (FAA) at 16q24.3 (Pronk *et al.*, 1995), and recently mapped, a recessive hereditary spastic paraplegia (HSP) also at 16q24.3 (De Michele *et al.*, 1998). In addition to linkage mapping for such hereditary disease loci, the loss of heterozygosity (LOH) mapping, was also used to define the chromosome region 16q24.3 as one of the region likely to harbor tumor suppressor gene (s) associated with sporadic breast cancer (discussed in section 1.5.4).

The refined mapping of both the FAA interval and the breast cancer tumor suppressor region indicated that they were both located within the same critical genomic region of 16q24.3 (The Fanconi anaemia/Breast cancer consortium, 1996; Cleton-Jansen *et al.*, 1994). With the objectives of cloning the FAA gene and a putative breast cancer tumor suppressor gene, the primary physical map of the critical region was developed comprising an integrated cosmid contig of approximately 650 kb. Based on this cosmid contig, a combination of exon amplification and cDNA selection methods was successfully used to isolate the candidate gene for FAA (The Fanconi anaemia/Breast cancer consortium, 1996). Further to this, a physical map was extended to nearly 1 Mb utilizing cosmid, BAC, and PAC clones. This physical map was then used as a framework for the development of transcription map for subsequent gene identification (Whitmore *et al.*, 1998a), especially for the isolation of candidate breast cancer tumor suppressor gene (s). As part of the chromosome 16-breast cancer project, the primary aim of the present study was to clone and



characterize the genes within this 16q24.3 LOH region for the candidate breast cancer tumor suppressor, utilizing the available physical and transcription map (Whitmore *et al.*, 1998a).

The following sections provide principles for disease gene mapping and gene isolation. The mechanism of carcinogenesis with colorectal cancer as a model is discussed, and follows<sup>ed</sup> by discussion on the present knowledge of breast cancer genetics and carcinogenesis. Studies on chromosome 16q LOH in sporadic breast cancer are also discussed. Finally, the objective and scope of the present study are given in the last section. x

## 1.2 Mapping of Disease Genes

Identification of genes associated with human diseases primarily requires the knowledge of the approximate location of the genes on chromosomes for subsequent cloning and identification of the candidate genes.

### **1.2.1 Genetic Mapping of Mendelian Trait**

Traits or diseases caused by defects in a single major gene or biochemical pathway are called Mendelian or single gene traits. Mapping of a disease-associated locus is based on a genetic linkage and requires a detailed linkage map of the genome. Genetic linkage is determined by measuring the frequency that two genetic loci are separated by meiotic recombination and is described as "recombination frequency" or "recombination fraction" (using a symbol " $\theta$ ") which is between 0-0.5. The correlation is the closer the two loci are, the lower the  $\theta$  between them. The genetic distance between two loci is then measured based on the  $\theta$  with the measuring unit is "centiMorgans" (cM). Genetic distance is generally correlated with the physical distance, which is defined by the number of base pair (bp) of x

centimorgan is not defined

DNA sequence between two genetic loci. One centiMorgan is correlated with the distance of approximately one megabase ( $1\text{Mb} = 10^6$  base pairs). *see comment.*

Linkage analysis test for co-segregation of a marker and disease phenotype within a pedigree to determine whether a genetic marker and a disease predisposing locus are physically linked, that is, in close physical proximity to each other. By typing marker loci at known locations in the genome, each marker can be tested for linkage to a disease or trait and approximate the location of the disease or trait to the chromosomal region harboring the linked markers. When large, multi-generation pedigrees are available, linkage analysis is a powerful technique for the localization of disease genes.

### 1.2.2 Mapping of Complex Disorder

In contrast to Mendelian traits, complex or multifactorial diseases result from the interaction of multiple genes and environmental factors. These complex disorders include cardiovascular disease, rheumatoid arthritis, diabetes, some form of psychological disorder such as schizophrenia, and cancer. The complex, non-Mendelian pattern of inheritance due to the involvement of multiple genes and the influence of environmental factors in such disorders makes it impossible or rarely *possible* to find the large multi-generation pedigrees suitable for mapping of the associated genes by direct linkage analysis. However, in some form of complex disorders near-Mendelian families can be selected for linkage analysis. For instance, in breast cancer, few families can be found with many affected individuals in a pattern consistent with autosomal dominant inheritance whereby linkage analysis can be used for mapping of disease gene. Alternative procedures for mapping complex traits namely "sib pair analysis" and "association analysis" are described elsewhere (Thomson and Esposito, 1999),

## **1.3 Identification of Human Disease Genes**

Procedures have been developed for the identification of genes associated with human diseases. Without the data of genomic sequence encompassing the candidate region positional cloning approach is the one that has successfully been used to identify several disease-associated genes. It involves construction of a physical map of the candidate region for subsequent identification of the disease-associated gene.

### **1.3.1 Construction of the Physical Map encompassing the Candidate region**

Once the disease locus is mapped to a region of the chromosome, construction of a physical map encompassing such chromosome region is required to allow subsequent identification of candidate genes. Essentially, a physical map consists of an overlapping set of large genomic clones that are arranged in a contiguous order, with respect to the chromosome orientation. Cosmid clones (average insert size of 40 kb) have initially been used to build the genomic contig of a physical map. Alternatively, the larger genomic clones used to build a contig can be yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), or P1 phage recombinant (PACs). YACs have an advantage over other cloning systems in that their large insert size (350-1000 kb on average) facilitates the construction of maps with long-range continuity (Burke *et al.*, 1987; Burke, 1991). However, most YAC libraries have high rates of chimerism and deletions that can limit their utility (Green *et al.*, 1991; Selleri *et al.*, 1992). BAC and PAC clones, though provide smaller insert size (average 120-200 kb and 75 kb respectively) offer high clonal stability and reduced cloning biases (Shizuya *et al.*, 1992; Ioannou *et al.*, 1994). However, a more refined genome contig, utilizing cosmid clones may also be required for subsequent construction of transcript map for further gene isolation.

### 1.3.2 Finding Genes in Cloned DNA

The established contig of cloned genomic DNA provides a basis for the isolation of genes within the critical interval. A variety of strategies <sup>have</sup> been developed and used for gene detection.

#### 1.3.2.1 Gene detection by CpG Island Mapping

This method is based on the knowledge that about half of all vertebrate genes are characterized by the presence of C-G rich regions consisting of nonmethylated CpG dinucleotides, originally called "HpaII Tiny Fragment (HTF) island". The C-G content of these regions is about 60-70% compared with the average 40% for the human genome. These CpG islands are most often found at the 5' end of the genes, associated with the promoter region (Bird, 1987; Antequera and Bird, 1993). CpG islands have been found to be associated with all house-keeping genes and about 40% of tissue restricted genes (Larsen et al., 1992; Antequera and Bird, 1993).

The presence of CpG island can be detected by the use of restriction endonucleases with CpG dinucleotides in their recognition sequence, such as *NotI* (recognition sequence 5' GCGGCCGC 3'), *BssHIII* (5' GCGCGC 3'), *MluI* (5' ACGCGT 3'), and *NarI* (5' GGCGCC 3'). Since most of these restriction enzymes are inhibited by enzymatic methylation of cytosine residues, they are more likely to cleave DNA at the location of CpG islands than at intra- or intergenic regions that are usually methylated. Therefore, long-range restriction maps can be derived that identify the location of genes over large distances (Bird, 1986; 1987). Such an approach has allowed identification of CpG islands and has assisted in the identification of many genes (Rommens *et al.*, 1989; Tribioli *et al.*, 1994). The results, however, sometimes depend on the known tissue-specific variation in methylation patterns at CpG islands. In cloned sources of DNA, where the original pattern of cytosine

methylation is erased, only the clustering of restriction sites for CpG enzymes is often taken as an indication for CpG islands.

### 1.3.2.2 Gene Detection Based on Recognition of Exons

This gene detection method is based on the utilization of splice sites in a functional assay (Auch and Reth, 1990; Duyk *et al.*, 1990; Buckler *et al.*, 1991). In principle, genomic fragments are cloned into the intron of a reporter construct that provides splice acceptor and donor sites. Upon transfection into a suitable cell line, the RNA is expressed and the mRNA splicing machinery generates chimeric mRNAs containing exon fragments from the cloned piece of DNA. Because the sequences flanking the novel exon are known, polymerase chain reaction (PCR) can be used to amplify sufficient amounts of cDNA for cloning and subsequent sequence analysis. This procedure, generally called “*exon trapping*” or “*exon amplification*”, was a popular gene detection method because of its ability to recognize transcribed sequences regardless of their tissue expression pattern and also provided a high success rate in a number of positional cloning projects in the early 1990s.

Because exon trapping is based on splice site recognition, artifacts can be generated. Three classes of exon fragments are always detected: (i) true exons, which are recovered via the utilization of bona fide splice sites, (ii) exon fragments that are trapped using one bona fide splice site and a second cryptic splice site, (iii) fragments that are trapped because two cryptic splice sites are used. The first two types of exon fragments are generally acceptable, because they provide at least partial sequence information about exons. Only the third class is problematic, because it is the result of artifactual splicing reactions.

The classical versions of exon trap/amplification vectors were designed to detect only internal exons, exons that contain both splice acceptor and splice donor sites. However,

alternative modifications of the system have been described, whereby the 3'-terminal exon having only a splice acceptor site can also be detected (Datson *et al.*, 1994).

### **1.3.2.3 Gene Detection Based on Sequence Identity between Genomic DNA and cDNA**

This method, termed "cDNA selection", is based on the formation of heteroduplexes between genomic DNA and cDNA (Lovett *et al.*, 1991; Parimoo *et al.*, 1991). The first development of cDNA selection was based on the use of cDNA inserts, amplified from a cDNA library of interest using vector specific primers, to screen the genomic clones. These cDNA products are radioactively labeled and used to probe YAC or cosmid inserts that have been immobilized on nylon membranes. The cDNA inserts that specifically hybridize to the cloned genomic DNA are then eluted from the membranes and re-amplified with vector specific primers. The resulting cDNA sub-libraries, enriched for expressed sequences from the genomic region, are then cloned and analyzed.

The second-generation version of this method has been developed where both sources of DNA (genomic and cDNA) are converted into a form that is easily amplifiable by PCR. This is achieved by ligation of adapters to their respective ends (Morgan *et al.*, 1992; Korn *et al.*, 1992). By utilizing two adapters that differ in sequence, cDNA and genomic DNA fragments can be amplified independently from each other by their specific adapter primers. In the procedure, briefly, the amplified cDNA fragments are mixed with the amplified fragments of genomic DNA generated from particular genomic region. The mixture is denatured and allowed to reanneal, from which both homoduplexes and heteroduplexes are formed. To allow for the separation of heteroduplexes (cDNA:genomic DNA) from homoduplexes of cDNA fragments, the amplified fragments of genomic DNA are labeled, for instance by biotinylation of their specific primer prior used for PCR. This allows the use of affinity chromatography or affinity beads (eg. streptavidin coated magnetic beads) for

separation. The selected cDNA fragments can subsequently be isolated from heteroduplexes by PCR using cDNA specific-adaptor primers. Since the heteroduplex formation occurs only between the fragment of genomic DNA and cDNA with sequence identity, the selected cDNA, therefore, represents the coding sequence of the gene in particular genomic region. Because the selected cDNA fragments are usually larger in size than the trapped exons, transition to full length transcripts is usually easier than that from trapped exons. This method has been successful in the isolation of a number of genes (Rommens *et al.*, 1993; Peterson *et al.*, 1994; Baens *et al.*, 1995) including those responsible for human diseases (Gecz *et al.*, 1993; Onyango *et al.*, 1998)

#### 1.3.2.4 Positional Candidate Analysis

This approach is based on the fact that <sup>the</sup> number of sequenced genes or part<sup>s</sup> of genes, <sup>is</sup> often in the form of expressed sequenced tags (EST) and their chromosomal localization are increasing rapidly and are publicly available (Ballabio, 1993; Collins, 1995a). This procedure combines the mapping of a disease locus to a defined region with the availability of candidate genes mapped to the same chromosome region. Many mapped genes are of known function, or function can be predicted, though some are novel and are of unknown <sup>^</sup> function. In many cases reasonable hypotheses as to the nature of a gene associated with a given clinical phenotype can be made. Relevant candidate genes can then be analyzed for the presence of disease-associated mutations. This approach has made a successive impact <sup>^</sup> on the isolation of many disease-associated genes (Xu *et al.*, 1998; Pandey and Lewitter, 1999; Hofmann *et al.*, 2001; Bleck *et al.*, 2001).

#### 1.3.2.5 Direct Sequencing of the Critical region

Determination of the nucleotide sequence across a critical interval may provide the best possible way to identify candidate genes. With this approach, the coding region can be identified by computerized gene detection method. Indeed, large-scale sequencing has already been successfully applied in the positional cloning project, for instance the identification of a second breast cancer susceptibility gene, BRCA2 (Wooster *et al.*, 1995).

In addition, a large number of partial cDNA sequences (ESTs) as well as complete sequences of cloned cDNAs or even full range sequences of novel transcripts are available and can be retrieved from public database. These can facilitate the identification of genes utilizing homology search tools for the sequence homology between genomic DNA and cDNA from database.

### 1.3.3 Generation of Transcript Maps

The transcribed sequence information isolated by the above methods either as cDNA clones or trapped exons can then be mapped to their corresponding genomic clones in the critical region. This creates an overall map of the transcribed sequences arranged in order with respect to the genomic clone-contig. The transcribed sequences that arrange in cluster can be grouped and treated as a “transcription unit”, representing a tentative gene. The resulting transcript maps can facilitate cloning of the genes related to the disease locus.

## 1.4 Human Cancer: A Complex Genetic Disease

*Cancer* is a generic term for genetic diseases involving uninhibited cellular proliferation caused by the mutations of multiple genes that result in the disruption of the normal harmonious checks and balances controlling this process within normal cells. These mutations usually occur in genes that play a vital role in the regulation of cell growth and differentiation. In normal tissues, homeostasis is maintained by ensuring that as each stem



cell divides, <sup>on average</sup> only one of the two daughters remains in the stem cell compartment, while the other is committed to a pathway of differentiation (Cairns, 1975). The control of cell multiplication will therefore be the consequence of signals affecting these processes. These signals may be either positive or negative, and the acquisition of tumorigenicity results from genetic changes that affect these control points.

#### 1.4.1 Genes involved in Cancer

The information obtained from the cytogenetic analysis of cancer cells with recurrent chromosome abnormalities together with the study of hereditary cancer syndromes has provided a way for the identification of genes involved in cancer.

The first evidence of a genetic event in cancer was the discovery of the Philadelphia chromosome (Ph<sup>1</sup>) in chronic myelocytic leukemia (CML) (Nowell and Hungerford, 1960). This chromosome was first believed to be an abnormal chromosome 21. After a decade, it was Rowley (in 1973) who demonstrated that Ph<sup>1</sup> originated from a specific translocation between chromosomes 9 and 22. Ten year later molecular cloning of the translocation break-points showed that the translocation interrupted a break-point cluster region (*BCR*) gene on chromosome 22 and the Abelson oncogene (*c-ABL*) on chromosome 9 and created a new hybrid *BCR-ABL* oncogene (De Klein *et al.*, 1982; Heisterkamp *et al.*, 1983, 1985). Expression of this chimeric oncogene produces a tyrosine kinase related to the ABL product but with abnormal transforming properties. Other chromosomal translocations <sup>are</sup> were found, for example those seen in the majority of cases of Burkitt's leukemia involved the chromosome translocations t(8;14), t(2;8), and t(8;22) (Croce and Nowell, 1985; Klein, 1989). These translocations lead to the activation of the *c-Myc* oncogene at the chromosome 8 break-point as a consequence of the control of immunoglobulin gene heavy, *kappa* or *lambda* chain regulatory elements located at the break-points of chromosome 14, 2 and 22

respectively (Rabbits, 1994). Non-random or recurrent chromosomal aberrations have also been reported in a number of solid tumors, many of which involved gene fusion and activation of oncogenes at the break-points (Sánchez-García, 1997; Mitelman *et al.*, 1997).

Activation of "*oncogenes*" provides a positive signal for tumor progression while another group of genes, "*tumor suppressors*", act as negative regulators of cellular proliferation. Inactivation of these tumor suppressor genes is considered to initiate uncontrolled cellular growth and tumor formation.

The concept of a tumor suppressor gene is derived from two lines of evidence. The first, from cell hybridization experiments, showed that the neoplastic phenotype can often be suppressed by fusion of tumor cells with normal cells, resulting in the outgrowth of non-tumorigenic hybrids (reviewed by Harris, 1988). These observations indicated that normal cells were donating genetic information capable of suppressing the neoplastic transformation of the hybridized tumor cells. This suppressing ability is an important functional characteristic of tumor suppressor genes.

The second line of evidence was from epidemiological studies by Knudson in inherited retinoblastoma, a childhood tumor of retinal tissue (Knudson, 1971). Knudson postulated that development of retinoblastoma requires two successive mutations of the homologous alleles in the cell genome. From this "two-hit" hypothesis two forms of mutation were proposed in familial and sporadic retinoblastomas. In familial retinoblastoma, which occurs bilaterally in most cases, the first mutation is constitutional occurring in the germline and is therefore inherited, while the second mutation is a subsequent somatic event in the retinoblast. In contrast, sporadic retinoblastoma results from two independent somatic mutations in the retinoblast and is thus unlikely to be bilateral. The "two-hit" event required for the development of retinoblastoma is likely to inactivate a gene which the normal function is important in regulating the normal cellular growth. Cytogenetic analysis has

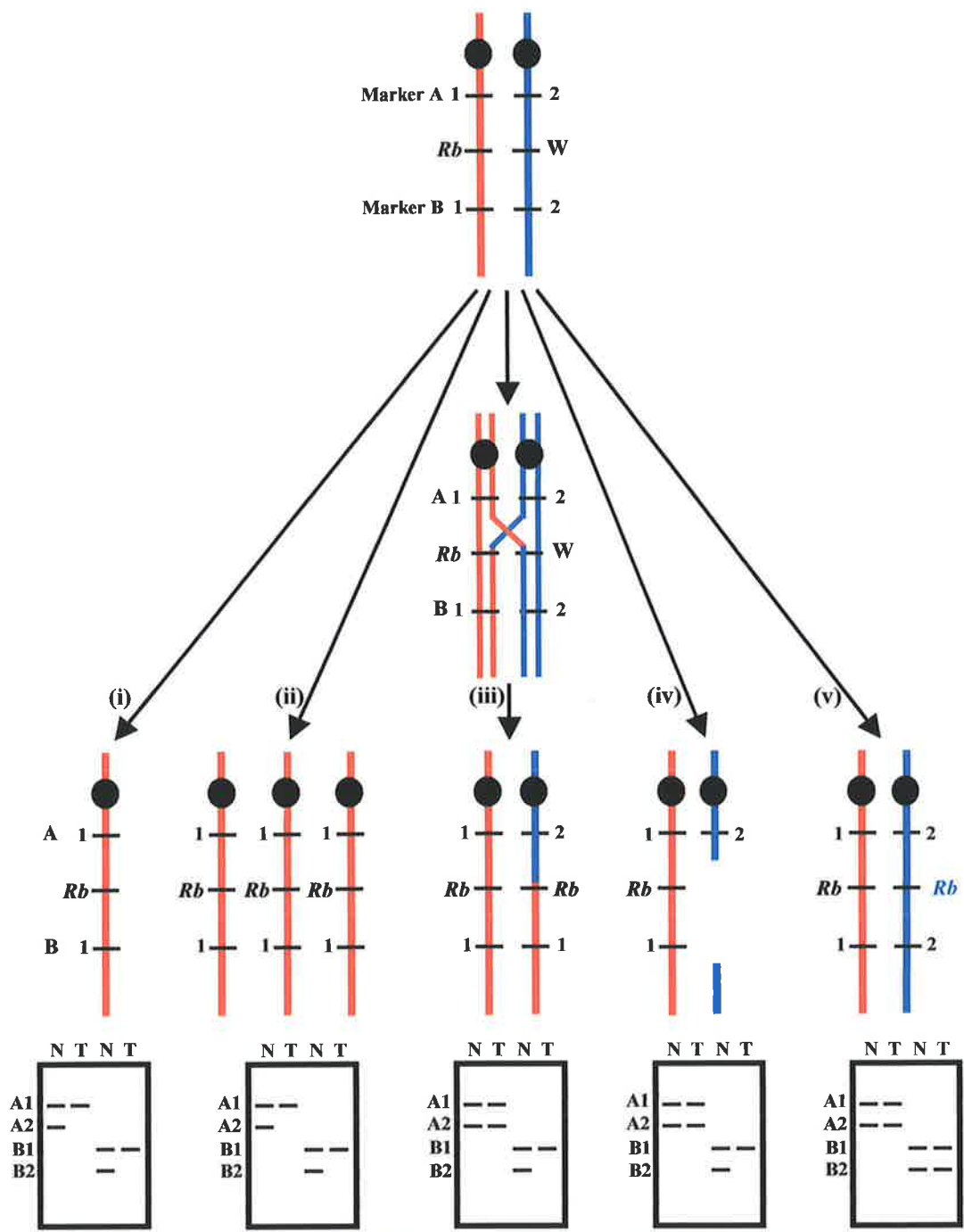
shown that a few hereditary retinoblastoma cases carried a constitutional chromosome band 13q14.1 deletion in all somatic cells (Francke and Kung, 1976; Knudson *et al.*, 1976). Further study in sporadic cases has also occasionally detected 13q14 deletions in tumor cells (Balaban *et al.*, 1982). Molecular genetic studies (Cavenee *et al.*, 1983) as well as the subsequent isolation of the gene (Friend *et al.*, 1986) have shown that Knudson's two elusive genetic targets were the two copies of the *retinoblastoma tumor suppressor gene (RB1)* located on the long arm of chromosome 13 at 13q14. The two mutational events, either constitutional or sporadic involved the inactivation of both functional copies of this gene. The *RB1* gene has subsequently been characterized and found to be mutated or lost in several cancers, including osteosarcoma (Friend *et al.*, 1986), small cell carcinoma of the lung (Harbour *et al.*, 1988; Horowitz *et al.*, 1990), bladder cancer (Horowitz *et al.*, 1990) and breast carcinomas (Lee *et al.*, 1988; T'Ang *et al.*, 1988; Horowitz *et al.*, 1990). Subsequent studies have established that the tumor suppressor function of RB protein (pRB) is through the inhibition of the cell cycle during the G<sub>1</sub>/S phase progression (reviewed by Weinberg, 1995).

Studies in other hereditary cancer syndromes have resulted in the isolation of other tumor suppressor genes with recessive germline mutations predisposing individual to cancer formation. These also are consistent with Knudson's "two-hit" model for complete inactivation of a tumor suppressor gene. These include the *APC* gene in familial adenomatous polyposis (FAP), a syndrome predisposing to colorectal cancer (Bodmer *et al.*, 1987; Groden *et al.*, 1991; Nishisho *et al.*, 1991); *TP53* gene in Li-Fraumeni syndrome, a syndrome of multiple cancer predisposition (Li and Fraumeni, 1975; 1982; Malkin *et al.*, 1990; Srivastava *et al.*, 1990); and the *VHL* gene in von Hippel-Lindau syndrome, a syndrome with multiple hemangioma and predisposition to renal cell carcinoma, pheochromocytoma, and pancreatic tumor (Latif *et al.*, 1993).

### 1.4.2 Tumor Suppressor Gene and Loss of Heterozygosity

The involvement of tumor suppressor genes, such as the *RB1* gene in tumorigenesis of retinoblastoma, appears to be by mutations resulting in loss of function. In the inherited form of retinoblastoma, the somatic mutation that affects the second allele of the *RB1* gene is likely to be <sup>a</sup>the chromosomal event, which uncovers the constitutional recessive mutation. This event was demonstrated by Cavenee and colleagues (1983). They utilized a series of RFLP (restriction fragment length DNA polymorphism) markers at chromosome 13q14 to type DNA from the surgically removed tumor material and from blood samples taken from the same patient. They found that in several patients the constitutional lymphocyte DNA was heterozygous for some 13q14 markers while the tumor cells were apparently homozygous. Such apparent homozygosity or "loss of heterozygosity (LOH)" was suggested to be the somatic equivalent of Knudson's "second hit". This LOH resulted in the loss of one functional copy of a tumor suppressor gene but also included nearby DNA markers on the chromosome. As a result of combined cytogenetic analysis and studies of additional 13q14 markers, a number of chromosomal mechanisms leading to the loss of the wild-type functional allele, were proposed (*Figure 1.1*). These include the loss of an entire chromosome by mitotic non-disjunction, mitotic recombination, unbalanced translocation or chromosome deletion. Subsequently, such events were often associated with reduplication of the remaining chromosome. The frequency of LOH in tumor cells has been estimated to be at least two order of magnitude higher than that of point mutation, and thus it is the favored mechanism to eliminate the wild-type allele of a tumor suppressor gene (Weinberg, 1991).

Since the loss of a specific chromosome region usually affects both the putative tumor suppressor gene and the neighbouring genes or genetic markers, typing of genetic markers can therefore be used to identify the chromosome region likely to harbor the putative tumor



**Figure 1.1** : Diagram representing the mechanism of loss of wild-type allele (w) in retinoblastoma (RB) studied by Cavenee *et al.*, (1983). **(i)** Loss of a whole chromosome by mitotic non-disjunction. **(ii)** Loss and followed by reduplication of the Rb chromosome. **(iii)** Mitotic recombination proximal to the Rb locus, followed by segregation of both Rb-bearing chromosome into one daughter cell. **(iv)** Deletion of the wild-type allele. **(v)** Pathogenic point mutation of the wild type allele. The corresponding results of genotyping normal (N) and tumor (T) DNA for the two markers, A and B, are represented by the figures underneath that indicate the patterns of loss of heterozygosity (LOH) in (i), (ii), (iii), and (iv), but not (v).

suppressor gene through LOH analysis. The LOHs at many chromosome regions have been found in several solid tumors indicating the possible location of multiple tumor suppressor genes involved in tumorigenesis.

### 1.4.3 Multistage Carcinogenesis: The Paradigm of Colorectal Cancer

Human carcinogenesis is a multistage process starting from a clonal expansion of the cell that gains a selective growth advantage by genetic events. Further genetic alterations accumulated during progression cause the phenotypic alteration of the cells to change into more malignant phenotypes. Clinically, cancer development and progression can be divided into sequential events: from normal cells to preneoplastic lesions, to primary tumors and finally to more aggressive metastases. Molecular genetic studies have revealed that such development and progression of human cancer requires multiple genetic alterations. This multistage process is well illustrated by studies of colorectal cancers.

Studies have defined two forms of hereditary predisposition to colorectal cancer: familial adenomatous polyposis (FAP) and hereditary non-polyposis colorectal cancer (HNPCC). These two diseases have provided the information leading to an understanding of colorectal carcinogenesis. It is assumed that other human tumors will follow a similar, though not the same, tumorigenic pathway.

In FAP a stepwise model of colorectal carcinogenesis was developed by Fearon and Vogelstein (1990). Mutations of the adenomatous polyposis coli tumor suppressor gene (*APC*) initiate the formation of preneoplastic lesion, the "adenomatous polyps", in both inherited (FAP) and sporadic cases. <sup>always?</sup> The next stage in progression to malignancy involves oncogenic *K-ras* mutations which <sup>arise</sup> arose during the adenomatous stage. Mutations of *TP53* and deletions of chromosome 18q coincided with the transition to malignancy. Subsequent

work has identified additional genetic events as well as the molecular pathways perturbed by each of these mutations.

The *APC* mutation appears to be the rate-limiting event for colorectal tumorigenesis. Individuals with germline mutations of *APC* are at high risk, but do not necessarily develop colorectal cancer. However, tumor formation will be initiated by a somatic mutation of the *wild type APC* allele inherited from an unaffected parent (Ichii *et al.*, 1992; Levy *et al.*, 1994; Luongo *et al.*, 1994). Furthermore, somatic mutations of both alleles of the *APC* gene were also found in a large number of sporadic colorectal tumors (Miyoshi *et al.*, 1992; Powell *et al.*, 1992). Although the *APC* gene product is also ubiquitously expressed in other tissues than the colonic epithelium, individuals with constitutional *APC* mutations rarely developed tumors in other organs. Thus it was speculated that *APC* gene behaves as the "gatekeeper" of colonic epithelial cell proliferation and its inactivation is required for net cellular growth (Kinzler and Vogelstein, 1996). Normally, "gatekeeper genes" are required for maintaining a constant cell number in renewing cell populations, ensuring that cells respond appropriately to the situations requiring net cell proliferation, such as in tissue damage. Mutation of the gatekeeper thus leads to the permanent imbalance of cell growth over cell death. The gatekeeper function of *APC* is important in colorectal tumor formation. Mutations of other genes known to be involved in the tumorigenic process, such as *TP53* and *K-ras*, fail to promote neoplastic transformation of colonic epithelial cell if the gatekeeper gene, *APC*, is still intact (Kinzler and Vogelstein, 1996). Other genes may also perform the gatekeeper role in other tissues, these are the *NF1* gene in Schwann cells, the *RBI* gene in retinal epithelial cells, and the *VHL* gene in kidney cells (reviewed in Knudson, 1993)

The *APC* tumor suppressor gene encodes a large multidomain protein (APC) which appears to interact and modulate the cytoplasmic level of  $\beta$ -catenin, a downstream effector

of the Wnt signaling pathway (Rubinfeld *et al.*, 1993; Gumbiner, 1995). APC in association with serine threonine glycogen synthase kinase (GSK)-3 $\beta$  regulates the low levels of free  $\beta$ -catenin. GSK-3 $\beta$  phosphorylates both APC and  $\beta$ -catenin facilitating APC/ $\beta$ -catenin complex formation and subsequent targeting  $\beta$ -catenin degradation (Rubinfeld *et al.*, 1996; Aberle *et al.*, 1997; Behrens *et al.*, 1998). In APC mutant colonic cells, degradation of cytoplasmic  $\beta$ -catenin is disrupted and the levels of  $\beta$ -catenin rise dramatically, suggesting its role in the transformation of colonic epithelium into benign polyps (Peifer, 1997).

$\beta$ -Catenin was originally identified on the basis of its association with cadherin adhesion molecules.  $\beta$ -catenin and its homologue, Armadillo protein in *Drosophila*, were recognized as essential components of the Wnt/Wingless signaling pathway (Gumbiner, 1995). Wnt/Wingless signals are known to direct many key developmental decisions including the regulation of anterior-posterior and dorsal-ventral pattern in both flies and vertebrates (Miller and Moon, 1996). The Wnt/Wingless signal, by antagonizing GSK-3 $\beta$  activity, disrupts the APC/ $\beta$ -catenin association, which prevents  $\beta$ -catenin from degradation and therefore stabilizes  $\beta$ -catenin (Papkoff *et al.*, 1996). High level of  $\beta$ -catenin then transduces the signal by translocating into nucleus and interacting with T-cell factor/Lymphoid enhancer factor (Tcf/Lef) family of HMG box transcription factor. TCF-4 is the predominant member of this family of transcription factors in colonic epithelial cell (Korinek *et al.*, 1997). Upon binding with  $\beta$ -catenin, TCF-4 is activated and it up-regulates many responding genes including the oncogenes *MYC* and *CCND1* which play role in cell growth by promoting the G1 $\rightarrow$ S phase of the cell cycle (He *et al.*, 1998; Mann *et al.*, 1999; Tetsu and McCormick, 1999). These findings suggest that  $\beta$ -catenin is a "driver" that up-regulates TCF-responsive gene critical for proliferation and transformation of colonic epithelial cells. Mutations of  $\beta$ -catenin gene (*CTNNB1*) has been reported in a subset of



colorectal tumors with intact APC (Morin *et al.*, 1997; Sparks *et al.*, 1998; Samowitz *et al.*, 1999). These mutations alter the N-terminal domain of  $\beta$ -catenin, a critical region significant for the down-regulation of  $\beta$ -catenin stability, resulting in maintaining its high level and transducing function through TCF-4 binding (Morin *et al.*, 1997; Rubinfeld *et al.*, 1997). These mutations therefore turn *CTNNB1* into an oncogene.

The studies of HNPCC have provided the evidence for another group of genes, mutations of which increase genomic instability. The colorectal tumor in HNPCC patient has been shown to exhibit widespread alterations of simple repeated DNA sequence such as polyA tracts and "microsatellites", including CA repeats (Aaltonen *et al.*, 1993). These alterations generally pronounced as "microsatellite instability" (MSI) have also been described in a subset of sporadic colorectal cancer (Peinado *et al.*, 1992; Ionov *et al.*, 1993; Thibodeau *et al.*, 1993). The genes responsible for HNPCC have now been identified as a group of mismatch repair (MMR) genes, these include *hMSH2*, *hMSH6*, *hMLH1*, *hPMS1*, and *hPMS2* (Peltomaki and de la Chapelle, 1997). However, the majority of HNPCC patients carry mutations of *hMSH2* and *hMLH1*. MMR genes encode proteins involved in DNA mismatch repair, a complex enzymatic proof-reading system that corrects base pair mismatches that arise during DNA replication.

Inactivation of both alleles of <sup>an</sup>MMR gene is required for colorectal tumors to develop in HNPCC patients. In general, patients with HNPCC have one normal allele of the relevant MMR gene: this wild-type allele is sufficient to maintain normal level of MMR protein (Parsons *et al.*, 1993). It is only when the wild-type allele is inactivated during tumorigenesis, through a gross chromosome event, loss of heterozygosity or a subtle intragenic mutation, that MMR is abolished and mutations accumulate (Liu *et al.*, 1995). The progeny of this cell, with "mutator phenotype", then accumulate mutations <sup>in genes generally including</sup> in oncogenes and tumor suppressor genes, thus resulting in clonal expansion and tumorigenesis.

The relationship between MSI mutator phenotype in HNPCC and mutations of *APC*, *K-ras*, or *TP53* remains unsettled. However, it has been demonstrated that mutation rates in tumor cells with MMR deficiency are two to three orders of magnitude higher than in normal cells or tumor cells without the mutator phenotype (Bhattacharyya *et al.* 1994; Shibata *et al.*, 1994; Eshleman *et al.*, 1995). This may accelerate the mutation of genes involved in colorectal tumorigenesis. Mutation of the *APC* gene has been reported in only approximately 21% of tumors with MSI from HNPCC patients (Konishi *et al.*, 1996). However, up to 43% of HNPCC, MSI-positive tumors, with wild-type *APC* has been shown to harbor *CTNNB1* mutations (Miyaki *et al.*, 1999a). The *CTNNB1* mutations detected in these HNPCC tumors resulted in the alteration of regulatory domain at the N-terminus of  $\beta$ -catenin. This likely allows  $\beta$ -catenin to escape from APC-induced degradation and subsequently exerts its activity through TCF-4 binding. Therefore, these indicate the common involvement of the Wnt (APC/ $\beta$ -catenin) signaling pathway in colorectal tumorigenesis either by inactivation of APC in ~~FAB~~<sup>p</sup> or by stabilizing of  $\beta$ -catenin in HNPCC. x

Other signaling pathways are also involved in colorectal carcinogenesis in both ~~FAB~~<sup>p</sup> and HNPCC as well as in sporadic colon cancer. These include the transforming growth factor beta (TGF- $\beta$ ) pathway and the <sup>TP53</sup> p53 pathway. TGF- $\beta$  is a potent inhibitor of normal epithelial cell growth including colonic epithelium (Kurokawa *et al.*, 1987; reviewed in Markowitz and Roberts, 1996). TGF- $\beta$  initiates cellular responses by binding to the type II receptor for TGF- $\beta$  (TGF- $\beta$ RII). This is followed by the recruitment and activation of the type I receptor through phosphorylation. Signaling to the nucleus is mediated through the signaling effectors, SMAD proteins. The type I receptor, with kinase activity, first phosphorylates SMAD2 and SMAD3, which subsequently form a complex with SMAD4. This hetero-oligomeric complex translocates to the nucleus and modulates transcription of specific genes through cis-regulatory SMAD-binding sequences. This results in activation of

such target as the genes for inhibitors of cyclin-dependent kinases (Cdk): p21<sup>WAF1/CIP1</sup>, p15<sup>INK4b</sup>, and p27<sup>KIP1</sup> that inhibit complex formation of Cdk4,6 and Cdk2 with their related cyclins. Cdk4,6-cyclin D and Cdk2-cyclin E complexes are required for the promotion of cell cycle from G1 to S phase by inhibiting the activity of tumor suppressor protein pRB. Inhibition of Cdk-cyclin complex formation therefore arrests the cell cycle progression.

Mutations of the *TGFβRII* gene have been reported in most colorectal tumors with MSI, both HNPCC and MSI positive sporadic tumors, comprising approximately 13% of all colorectal cancer (Lu *et al.*, 1995; Markowitz *et al.*, 1995; Parsons *et al.*, 1995). These mutations occur predominantly in the repeated sequences viz a short GT-repeat or a stretch of 10 adenines, (A)<sub>10</sub>, in coding regions of *TGFβRII*. Functional studies showed that these mutations render cells resistant to the growth-inhibitory effects of TGF-β suggesting the contribution of TGF-β pathway alteration in colorectal tumorigenesis (Markowitz *et al.*, 1995). In addition, suppression of tumorigenicity was also observed in receptor-negative colon cancer cells after restoration of the TGF-β receptor by stable transfection with *TGFβRII* (Wang *et al.*, 1995; Ye *et al.*, 1999). There <sup>has</sup> been reported an increased incidence of *TGFβRII* mutation in advanced stages of colorectal adenomas such as high-grade dysplasia and highly progressed adenomas that contained regions of invasive adenocarcinoma (Grady *et al.*, 1998). This suggests that *TGFβRII* mutation is a late event in colorectal adenoma and correlates with the progression from adenoma to carcinoma.

Although the *TGFβRII* mutation is very common among human colon cancers with MSI (MSI positive), mutation of this gene appears to be low in colon cancers without MSI (MSI negative). Grady *et al.* (1999) have reported the presence of *TGFβRII* mutation in approximately 15% of MSI negative colon cancers tested. The same group, however,

demonstrated that an additional 55% of the MSI negative colon cancers exhibited TGF- $\beta$  signaling abnormalities downstream to TGF- $\beta$ RII (Grady *et al.*, 1999).

A number of studies have revealed defects in the components downstream from TGF- $\beta$ RII in colon cancer. Mutations of *SMAD2* were observed in about 4% (Eppert *et al.*, 1996; Riggins *et al.*, 1996, 1997) while *SMAD4* mutations have been reported in approximately 30% of colorectal cancers (MacGrogan *et al.*, 1997; Takagi *et al.*, 1996; Thiagalingam *et al.*, 1996). The LOH on chromosome 18q have frequently been detected during progression of colorectal carcinomas in both FAP and sporadic cases. Among the genes situated within this LOH region, *SMAD2* and *SMAD4* gene have been investigated. Inactivation mutations, however, were observed mostly in *SMAD4* in invasive carcinomas as well as in distant metastases (Miyaki *et al.*, 1999b). This finding is consistent with the involvement of the TGF- $\beta$  signaling defect in late stage adenoma and the progression to invasive carcinoma (Grady *et al.*, 1998). The importance of the SMAD proteins in colorectal carcinogenesis has also been observed in experimental mice. Mice with compound heterozygous mutation of the *Apc* gene ( $APC^{\Delta 716}$  mice) develop multiple adenomas in small intestine and colon, but rarely progress to invasive carcinoma (Takaku *et al.*, 1998). However, homozygous deletion of *Smad4* resulted in embryonic lethality. Crossbreeding of  $APC^{\Delta 716}$  mice with mice bearing a heterozygous mutation in *Smad4* resulted in mice with intestinal polyps that developed into malignant tumor much more quickly than those observed in  $APC^{\Delta 716}$  mice (Takaku *et al.*, 1998, 1999).

Inactivation of the p53 tumor suppressor pathway is also involved in the progression of colorectal carcinoma. While mutations of *TP53* have been found in the majority of colorectal cancers, the cancers with MSI mutator phenotype are usually wild type for *TP53* (Ionov *et al.*, 1993; Kim *et al.*, 1994; Konishi *et al.*, 1996). In addition to its known cellular function with a central role in cell growth arrest (Donehower and Bradley, 1993; Levine,

1993), p53 is one of the most potent regulators of apoptosis in response to DNA damage (Yonish-Rouach, 1996; Kastan *et al.*, 1991). The p53 protein mediates its apoptotic function by transactivating *BAX* (Miyoshita and Reed, 1995), a member of the *BCL-2* gene family (Cory, 1995; White, 1996) that promote apoptosis (Oltvai *et al.*, 1993). In a number of studies, somatic frame-shift mutations in the *BAX* gene have been found in colon cancers of MSI mutator phenotype, both HNPCC and sporadic cases (Rampino *et al.*, 1997; Yamamoto *et al.*, 1997, 1998).

Overall, the paradigm of colorectal cancer provides a comprehensive view of human carcinogenesis that is multistage and involves multiple cellular pathways. Initiation may occur either by mutations of a gate keeper gene that lead to uncontrolled cell proliferation and or of a gene or genes that <sup>are</sup> involve in genetic integrity. The latter <sup>may</sup> then cause mutator phenotype that accelerates mutation of other genes including those <sup>the</sup> involve in gate keeping pathway.

## 1.5 Breast Cancer

Breast cancer is the most common female malignant tumor with an incidence of one in ten among women in the Western world. Mortality from breast cancer is usually due to unresponsiveness to therapy and the development of distant metastases. With a highly variable clinical course, the disease usually progress rapidly with short survival in some patients, while others have a long disease-free interval, followed by the appearance of distant metastases several years after the initial <sup>therapy</sup> operation (Rutqvist and Wallgren, 1985).

Breast cancers are derived from the epithelial cells that line the milk gland and terminal duct lobular unit. Microscopically, the duct system of the breast consists of branching ducts, which extend from the nipple area into the fibro-adipose tissue and terminates blindly in breast lobules. In each lobule, the duct branches into a cluster of terminal ductules, each

forming a glandular structure or “acinus”. The cells lining of duct and lobular system are cuboidal epithelium and are supported by myoepithelium and basement membrane. Cancer cells that remain within the basement membrane of acinus and terminal duct are classified as “*in situ*” or “non-invasive”. Invasive carcinoma is described when the cancer cells infiltrate through basement membrane of the ducts and lobules into the surrounding normal tissues. There are two distinctive histological types of breast cancer: ductal carcinoma and lobular carcinoma. *Ductal carcinoma* is defined when cancer arose from epithelial cells lining the duct system outside the lobule. It is described as ductal carcinoma in situ (DCIS) if the neoplastic cells confined within the basement membrane. *Lobular carcinoma* is a neoplastic proliferation of epithelial cells lining the terminal ducts and ductules (acini) within the lobular system. Clinically, the most common histological type is the invasive form of ductal carcinoma or “infiltrating ductal carcinoma” which constitutes up to 90% of all breast cancer types.

### **1.5.1 Genetic Predisposition to Breast Cancer**

Among the risk factors for breast cancer, a positive family history appears to be the most important. Clinical epidemiologic studies suggested that breast cancer could be transmitted as an autosomal dominant trait in certain families (Go *et al.*, 1983). Segregation analysis in a series of families indicated that approximately 5-10% of all breast cancer and ovarian cancer cases are consistent with autosomal dominant inheritance and the age of development of breast cancer is significantly reduced compared with sporadic cases (Newman *et al.*, 1988; Schildkraut *et al.*, 1989).

#### **1.5.1.1 Breast Cancer associated Gene: BRCA1 and BRCA2**

Studies of families with early-onset breast cancer, defined as before the age of 46, allowed the localization of the first breast cancer susceptibility gene locus, designated as *BRCA1*, to the chromosome region 17q12-21 (Hall *et al.*, 1990). However, only 40% of the families analyzed showed linkage to the polymorphic marker, D17S74, that is located adjacent to *BRCA1*. This indicates genetic heterogeneity. Other linkage studies in five families, where both breast and ovarian cancer segregated, also demonstrated that predisposition to breast and ovarian cancer was linked to D17S74 marker in three families (Narod *et al.*, 1991).

Additional analysis of 214 families subsequently confirmed the linkage of cancer predisposition to the *BRCA1* locus in nearly all families with both breast and ovarian cancer. However, only 45% of families with breast cancer alone showed linkage to this locus. This provided further evidence of genetic heterogeneity, indicating the existence of one or more additional genes (Easton *et al.*, 1993). Linkage studies in families with at least one case of male breast cancer in addition to early onset female breast cancer also failed to show linkage with *BRCA1* at 17q12-21, again confirming the existence of another gene or genes (Stratton *et al.*, 1994).

A genome-wide genetic linkage studies in families unlinked to *BRCA1* and including several with male breast cancer, resulted in the mapping of the *BRCA2* locus to chromosome 13q12-13 (Wooster *et al.*, 1994). Families with breast cancer linked to *BRCA2* are distinguished by a high incidence of male breast cancer. Risk of breast cancer among males predicted from linkage analysis to carry *BRCA2* mutations is 6% by age 70; inherited mutations in *BRCA2* may be involved in 15% of all male breast cancer (Szabo and King, 1995).

Risks of female breast cancer are similar for *BRCA1* and *BRCA2*. However, *BRCA1* is responsible for a higher proportion of cases of inherited breast and ovarian cancer than

*BRCA2*. Evaluation of 200 families with at least four cases of breast cancer showed that approximately 50% of these families have convincing linkage to *BRCA1*, 30% linkage to *BRCA2* and 20% showed no linkage to either *BRCA1* or *BRCA2* (Szabo and King, 1995). Such a high proportion of unlinked families may reflect the existence of another possible BRCA locus.

In 1994, positional cloning was successfully used to isolate the *BRCA1* gene (Miki *et al.*, 1994) and this was followed by the identification of *BRCA2* a year later (Wooster *et al.*, 1995).

The *BRCA1* gene is composed of 24 exons spanning over 80 kb of genomic DNA and encodes a 7.8 kb transcript which produces a large protein of 1863 amino acids (Miki *et al.*, 1994). Although the BRCA1 protein as a whole shows no sequence similarity to any known protein, it reveals at least three conserved functional domains based on the sequence similarity with those of known proteins. A highly conserved zinc-binding RING finger domain is located between residues 20-68. The RING class zinc fingers are known to be involved in protein-protein interactions (Saurin *et al.*, 1996). By using the BRCA1 RING finger domain as "bait" in a yeast-two hybrid screening Wu *et al.* (1996) identified another RING-domain protein, BARD1 (BRCA1-associated RING domain), which binds to BRCA1. The RING domains of both BRCA1 and BARD1 are necessary, but not sufficient, to mediate this interaction (Wu *et al.*, 1996). There are two tandem copies of a motif, designated as the BRCT domain, located at residues 1699-1736 and 1818-1855 (Koonin *et al.*, 1996). BRCT domains are also found in a duplication at a carboxyl terminus of BARD1 (Wu *et al.*, 1996). The function of the BRCT domain is unknown, however it has been found in several proteins involved in cell cycle regulation or DNA repair such as RAD9, XRCC1, RAD4, RAP1 and in 53BPI, a p53 binding protein (Callebaut and Mornon, 1997). The nuclear magnetic resonance structure of a BRCT domain from the human DNA repair



protein, XRCC1, suggests BRCT's involvement in the interaction of heterodimeric partners (Zhang *et al.*, 1998b). The region between amino acids 1214-1223 in BRCA1 shows a sequence similarity to the granin consensus (Jensen *et al.*, 1996). However, the functional significance of this domain in BRCA1 is not known.

Similar to *BRCA1*, the *BRCA2* gene is a large gene spanning approximately 80 kb of genomic DNA and is composed of 27 exons, encoding a very large protein of 3418 amino acids (Wooster *et al.*, 1995; Tavtigian *et al.*, 1996). The BRCA2 protein, however shows no similarity to BRCA1, and contains no RING finger or BRCT repeated sequence motifs. BRCA2 contains eight copies of a 30-80 amino acid repeat, termed BRC repeat, located between residues 1000-2030. The BRC repeats in BRCA2 show a high degree of sequence conservation among human, mouse, and chicken. This repeat motif has been shown to interact with RAD51, a human homologue of the *E.coli* RecA protein, suggesting a role for the BRCA2 protein in recombination and repair of double-stranded DNA break (Wong *et al.*, 1997; Zhang *et al.*, 1998a)

### 1.5.1.2 *BRCA1* and *BRCA2* functions and Breast Cancer

Evidences<sup>have</sup> have accumulated during the past few years and indicated that BRCA1 and BRCA2 are involved in common biological pathways. Mutations of either of these genes disrupt these pathways and predispose individuals to the development of early-onset breast/ovarian cancer. Both BRCA1 and BRCA2 are localized to the nuclear compartment (Scully *et al.*, 1996; Bertwistle *et al.*, 1997) and appear to play<sup>a</sup> role in maintaining genomic integrity through their involvement in cell cycle checkpoint, regulation of homologous recombination and DNA double-stranded break repair.

Studies have shown that during the cell cycle, BRCA1 and BRCA2 are preferentially expressed at late G1-early S phase transition, with a peak of expression at S phase and

remain elevated during G2-M transition (Gudas *et al.*, 1996; Wang *et al.*, 1997). This suggests a function during or following DNA replication. BRCA1 is hyperphosphorylated during the G1-S transition by the Cdk-cyclin complex (Ruffner *et al.*, 1999) and has been shown to induce cell cycle arrest by binding to hypophosphorylated pRB (Aprelikova *et al.*, 1999). The pRB is an essential component in the regulation of G1-S transition, and pRB is active when it is hypophosphorylated (Ewen *et al.*, 1994; Weinberg, 1995). In addition, BRCA1 may also regulate the G2-M checkpoint as well as control the assembly of mitotic spindles and the appropriate segregation of chromosome to daughter cells. Evidence from mouse embryonic fibroblasts carrying a targeted deletion of *Brca1* exon 11 showed that the G1-S checkpoint was maintained but there was failure to arrest cell at the G2-M checkpoint (Xu *et al.*, 1999). In a significant number of mutant cells, centrosomes were extensively amplified, resulting in abnormal chromosome segregation and aneuploidy. Since BRCA1 localizes to, and interacts with protein compartment of the centrosome during mitosis (Hsu and White, 1998), mutant BRCA1 may also induce genetic instability by disrupting the regulation of centrosome duplication.

*BRCA2* mutations may also involve in the disruption of a mitotic checkpoint. Tumors from *Brca2* knockout mice are defective in the spindle assembly checkpoint and acquire mutations in p53, Bub1 and Mad3L which are the components of the mitotic checkpoints that assess kinetochore activity to determine the corrective alignment of chromosomes on the spindles (Cahill *et al.*, 1998; Lee *et al.*, 1999). In addition, mouse fibroblasts that are homozygous for *Brca2* truncation undergo proliferation arrest and show chromosomal aberrations. However, with the presence of the mutant form of p53 and Bub1, this growth defect and chromosomal aberration can be overcome and neoplastic transformation initiated (Lee *et al.*, 1999). These suggest the role for inactivation of the mitotic checkpoint during tumor development in BRCA2-deficient cells.

A role for BRCA1 and BRCA2 in homologous recombination and DNA break repair is suggested by the biochemical interaction of BRCA1 and BRCA2 with proteins known to be involved in these process, especially the RAD51 protein (Zhang *et al.*, 1998a). RAD51 protein is required for recombination events during mitosis and meiosis as well as recombinational repair of double-stranded DNA breaks (Shinohara *et al.*, 1992; Baumann *et al.*, 1997).

BRCA1 was shown to interact with RAD51 in both mitotic and meiotic cells (Scully *et al.*, 1997a) in a complex that also contains BARD1 (Jin *et al.*, 1997). Furthermore, both BRCA1 and BRCA2 interacted and co-localized to this complex, forming punctate nuclear foci during the S phase of the cell cycle (Chen *et al.*, 1998b). In response to DNA damage, the complex of BRCA1-RAD51-BRCA2 re-localizes into macrostructures containing replicating DNA and proliferating cell nuclear antigens (PCNA), where BRCA1 undergoes phosphorylation (Scully *et al.*, 1997b; Chen *et al.*, 1998b). The DNA damage initiated-BRCA1 phosphorylation is dependent on *ATM* (ataxia-telangiectasia-mutated) gene product and /or ATR (a group of ATM-related kinase) (Cortez *et al.*, 1999; Gatei *et al.*, 2000; Lee *et al.*, 2000). Phosphorylation allows the dissociation of BRCA1 from its protein partner, CtIP, a protein of unknown function that associated with the transcriptional repressor CtBP (Li *et al.*, 1999). Dissociation of BRCA1 from CtIP might enable BRCA1 to mediate its function in association with RAD51/BRCA2 and other components in double-stranded DNA break repair.

Phosphorylated BRCA1 might also activate transcription of other DNA-damage response genes. BRCA1 was shown to induce expression of both *GADD45* and *p21*, the DNA damage-responsive genes which are the specific targets for p53 transactivation (Harkin *et al.*, 1999; reviewed in Irminger-Finger *et al.*, 1999). *GADD45* in association with the c-Jun N-terminal kinase / stress-activated protein kinase (JNK/SAPK) signal pathway

can lead cells to apoptosis (Takekawa and Saito, 1998). The p21<sup>WAF1/CIP1</sup> protein is one of the potent cyclin-dependent kinase (Cdk) inhibitors. Inhibition of the Cdk-cyclin complex formation, Cdk4,6-cyclinD and Cdk2-cyclinE, allows the pRB protein to escape from phosphorylation by Cdk and therefore exerts its cell cycle suppression function.

In addition to contributing to recombinational repair of double strand breaks, BRCA1 has also been implicated in other different DNA repair pathways including transcription-coupled base excision repair of oxidative DNA damage (Gowen *et al.*, 1998; Abbott *et al.*, 1999). The transcription-coupled process repairs damaged DNA more rapidly in transcriptionally active loci compared with the whole genome.

Overall, the tumor suppressing function of both BRCA1 and BRCA2 proteins are likely to involve their role in DNA double-strand break repair as well as in cell cycle checkpoint control in dividing cells. Gene targeting experiments have provided insights into the relationship between the defect of *BRCA* genes and breast cancer. *BRCA1* or *BRCA2* knockout mouse embryos die during the time of gastrulation. These embryos exhibit a proliferative defect and induction of the p53-dependent cell cycle inhibitor, *p21* gene. Since BRCA1 regulates DNA break repair, its inactivation may lead to spontaneous abnormalities in DNA structure. The induction of *p21* in *BRCA1* knockout embryos therefore reflects the activation of a DNA damage-dependent checkpoint. The resulting cell cycle delay may have adverse effects on a gastrulating embryo and lead to “death by checkpoint”. Consistent with this hypothesis, nullizygosity of either *TP53* or *p21* delays the death of *BRCA1* or *BRCA2* knockout embryos (Deng and Scott, 2000; Welch *et al.*, 2000). It was speculated that BRCA1 defect might have similar effects in adult tissues, by precipitating spontaneous DNA structural abnormalities and undergoing “checkpoint mediated growth arrest”. If, however, inactivation of BRCA1 occurred in a pre-malignant cell that had already defected in key checkpoint proteins, such as p53 or p21, the resulting aberrant DNA structure from loss of

BRCA1 function might be tolerated without cell cycle arrest (Welsh *et al.*, 2000). This might promote neoplastic transformation.

BRCA protein functions involve ~~in~~ double-strand break repair and cell cycle checkpoint. Their functions are, therefore, important for the maintenance of genetic stability during cell division. Inactivations~~/of~~ BRCA1 and BRCA2 proteins appear to be restricted for tumorigenesis of breast and ovarian epithelial cells, comparable to the APC inactivation that is specific for colorectal tumorigenesis. However, unlike APC protein “gate keeping function”, the BRCA proteins behave as “caretaker” for genome integrity and are likely to be conserved in most tissues. Their defects likely promote global genomic instability (mutator phenotype) that accelerates tumor progression. This seems to contradict the observed tissue specificity of tumor development in BRCA mutations. Individuals carrying either BRCA1 or BRCA2 germline mutations have never or rarely been reported to develop cancer of other tissues than breast and ovary.

Tissue specificity of *BRCA*-linked cancer risk could be relevant to the hormone-dependent nature of both breast and ovarian epithelium. Proliferation of both tissues is, indeed, under the influence of estrogenic hormones. There is evidence that some estrogen metabolites can <sup>lead to</sup> adduct DNA, and so could act as carcinogens and induce tissue specific DNA damage (Fishman *et al.*, 1995; Liehr, 2000). This effect might be exacerbated by *BRCA* mutations as well as by dysfunction of related DNA repair pathways. In addition, BRCA1 has been reported to inhibit estrogen-dependent transactivation by the estrogen receptor  $\alpha$  (ER $\alpha$ ) through its direct interaction with ER $\alpha$  (Fan *et al.*, 1999). In this regard, BRCA1 <sup>has</sup> play~~s~~ role in the modulation of estrogen signaling pathways and, hence, the expression of hormone-responsive genes.

### 1.5.1.3 Mutation spectrum of *BRCA1* and *BRCA2*

The mutation spectrum of both *BRCA1* and *BRCA2* have been registered in the Breast Cancer Information Core (BIC) data base ([http://www.nhgri.nih.gov/Intramural\\_research/Lab\\_transfer/Bic](http://www.nhgri.nih.gov/Intramural_research/Lab_transfer/Bic)). These include more than 200 different germ-line mutations in *BRCA1* and over 100 such mutations in *BRCA2*, all of which are clearly associated with cancer susceptibility. There is no evidence of mutations clustering in any particular region. Mutations are scattered throughout the coding sequence of both *BRCA1* and *BRCA2*. Of the known *BRCA1* and *BRCA2* mutations, ~ 95% are predicted to result in premature termination of translation.

A survey of large numbers of *BRCA1* linked families and high-risk women have revealed a few mutations that are seen recurrently in some population (Shattuck-Eidens *et al.*, 1995; Couch and Weber, 1996). These include a mutation within the RING-finger domain (185delAG) that is found in approximately 12% of *BRCA1* mutation carriers, a mutation within exon 20 (5382insC), comprises ~10%, and a mutation within exon 11 (4184del4). The 185delAG mutation was seen in ~20% of Ashkenazi breast cancer cases diagnosed under age of 40 (Fitzgerald *et al.*, 1996; Offit *et al.*, 1996).

Recurrent *BRCA2* mutations <sup>have</sup> also been reported. The 6503delTT mutation has been detected in several British and French families. In Ashkenazi Jews, 6174delT is the main *BRCA2* mutation and has been detected in 8% of breast cancer cases, diagnosed before age 40 (Neuhausen *et al.*, 1996). Finally, 999del5 is the main mutation in Icelandic *BRCA2* families (Thorlacius *et al.*, 1996).

#### **1.5.1.4 Other Syndromes showing predisposition to Breast Cancer**

Other hereditary syndromes with predisposition to breast cancer are considerably rarer than *BRCA1* and *BRCA2* mutations. These include Li-Fraumeni Syndrome, Ataxia-telangiectasia, and Cowden disease.

*Li-Fraumeni Syndrome (LFS)* is a rare autosomal dominantly inherited syndrome of early onset breast cancer and other diverse cancers, notably soft tissue sarcoma, osteosarcoma, brain tumor, leukemia, and adrenocortical carcinoma (Li and Fraumeni, 1975; Lynch *et al.*, 1978; Li and Fraumeni, 1982; Li *et al.*, 1988). The average age of breast cancer in LFS is 35 years with a high rate of bilaterality (Garber *et al.*, 1991). The genetic basis for LFS is caused by mutations in the tumor suppressor gene *TP53*, located at chromosome 17p13.1 (Malkin *et al.*, 1990; Srivastava *et al.*, 1990).

Patients with *Ataxia-telangiectasia (AT)*, a rare autosomal recessive disorder, experience progressive neurological degeneration (e.g. cerebral ataxia), telangiectases of skin and eyes, immunodeficiency, developmental abnormalities, hypersensitivity to ionizing radiation, and an increased susceptibility to various cancers including breast cancer and haematologic malignancies (Gatti *et al.*, 1991). A gene mutated in AT (*ATM*), located at chromosome arm 11q22-23, codes for a protein member of protein kinase family homologous to the catalytic subunit of phosphoinositide 3-kinase (PI3-kinase) (Savitsky *et al.*, 1995). Predisposition of AT patient to a wide-range of cancers may be accounted for by the *ATM* protein function, which involves <sup>i</sup> in <sup>d</sup> genomic stability. The *ATM* protein, in response to DNA damage, phosphorylates and activates many key proteins in the regulation of the cell cycle checkpoint and DNA damage repair pathways. These include p53, Chk1/Chk2/hCds1, c-Abl, and Mre11/Rad50/NBS1 (Baskaran *et al.*, 1997; Shafman *et al.*, 1997; Banin *et al.*, 1998; Canman *et al.*, 1998; Matsuoka *et al.*, 1998; review in Dasika *et al.*, 1999).

A number of studies have suggested that women heterozygous for mutations at the *ATM* locus have an increased relative risk for breast cancer (Swift *et al.*, 1991; Easton, 1994; Inskip *et al.*, 1999; Janin *et al.*, 1999). However, this has not been confirmed by other studies (Wooster *et al.*, 1993; Fitzgerald *et al.*, 1997; Chen *et al.*, 1998a). Nevertheless, a

recent study surveying a selected group of patients with breast cancer has demonstrated that there is a high percentage of *ATM* germline mutations among patients with sporadic breast cancer (Broeks *et al.*, 2000). Furthermore, two reports have provided evidence that ATM is directly involved in the regulation of the product of the breast cancer gene *BRCA1* (Cortez *et al.*, 1999; Gatei *et al.*, 2000). As discussed in section 1.5.1.2, *BRCA1* is involved in the cell cycle control and DNA repair processes through its association with Rad51. This functional interaction of the ATM and *BRCA1* proteins argues in favor of the involvement of particular aspects of ATM function in predisposition to breast cancer.

In *Cowden syndrome* (CS), a rare autosomal dominant syndrome, affected members tend to develop multiple hamartomas and other benign tumors including oral papillomatosis, gastrointestinal polyposis, and female genital tract tumors, as well as an increased susceptibility to breast and thyroid malignancies (Hanssen and Fryns, 1995). Linkage analysis studies have localized the gene responsible for CS to chromosome 10q22-23 (Nelen *et al.*, 1996). This chromosome region has also been represented as an important target area in several human tumors including glioblastoma, prostate and breast cancer, endometrial tumor, and haematologic malignancies (Pfeifer *et al.*, 1995; Rasheed *et al.*, 1995; Albarosa *et al.*, 1996; Trybus *et al.*, 1996; Kobayashi *et al.*, 1997; Bose *et al.*, 1998). The gene, designated *PTEN* or *MMAC1*, has been identified as a tumor suppressor located within the CS critical region, 10q23.3, and mutated in several advanced cancers (Li *et al.*, 1997; Steck *et al.*, 1997). This gene has also been designated as *TEP1* (transforming growth factor- $\beta$  regulated and epithelial-cell-enriched phosphatase) for its putative signaling effectors (Li and Sun, 1997). Finally, germline mutation of *PTEN* has been shown to be the cause of Cowden syndrome (Liaw *et al.*, 1997).

The roles of PTEN protein in tumor suppression is attributable to the induction of cell cycle arrest and apoptosis (reviewed in Dahia, 2000). PTEN is a dual specific phosphatase



and functions by antagonizing the phosphoinositide 3-kinase (PI3-kinase) activity, thus preventing the formation of PIP-3 (phosphatidylinositol triphosphate). This results in the down regulation of the PIP-3-dependent activation of Akt, a proto-oncogene serine/threonine kinase. Activated Akt is the upstream component of anti-apoptotic pathways as well as the pathway maintaining cyclinD1 mediated cell proliferation (Coffer *et al.*, 1998). Another cell proliferation arrest pathway mediated by PTEN is the induction of the *p27(KIP1)/CDKN1B* gene expression. The  $p27^{KIP1}$  protein is a cyclin-dependent kinase (CDK) inhibitor. By forming a complex with cyclin E,  $p27^{KIP1}$  can inhibit the CDK2 function, thus preventing the phosphorylation of pRB resulting in the inhibition of G1→S progression in the cell cycle.

### 1.5.2 Genetics of Sporadic Breast Cancer

Sporadic breast cancer comprises over 90% of all breast cancer cases. However, there is little detailed knowledge regarding the biological processes and the genes involved for the genesis of this type of breast cancer. In general, cancer susceptibility genes are tumor suppressors and require mutation of both alleles present within a normal cell for neoplastic transformation to occur. The mutations are usually predicted to result in the inactivation of gene functions. In hereditary cancer, there is a constitutional inactivating mutation of the tumor suppressor gene. Subsequently, in somatic tissue the wild-type allele is usually lost by a variety of mechanisms (see section 1.4.2), which is usually detectable as loss of heterozygosity (LOH) of polymorphic markers in the vicinity of the susceptibility locus. This results in complete inactivation of the tumor suppressor gene in somatic tissue which can lead to the development of tumor. This two-mutation hypothesis of tumor suppressor gene, which was articulated and elaborated by Knudson, also predicts that genes that confer a risk of cancer as a result of germline mutations are likely to be somatically mutated in

sporadic cancer of the same type (Knudson, 1993). This has proved to be the case for the paradigm of the *RBI* gene and for several other genes such as *VHL* and *NF2*. However, these typical features of tumor suppressor gene can, only in part, apply for the most prevalent inherited breast/ovarian cancer predisposing genes, *BRCA1* and *BRCA2*.

In most cancers arising in *BRCA1* or *BRCA2* mutation carriers, inactivation of germline mutation is rendered by loss of the wild-type allele (Collins *et al.*, 1995b; Cornelis *et al.*, 1995). Since, approximately 50-75% of sporadic breast and ovarian cancers exhibit LOH on chromosome 17q12-q21 and about 30-40% show LOH on chromosome 13q12-q13, it was widely anticipated that somatic *BRCA1* and *BRCA2* mutations would be found in sporadic breast and ovarian cancers. However, no clear disease-causing somatic mutations have been described in either gene in sporadic breast cancer, and somatic mutations in ovarian cancer are very rare (Futreal *et al.*, 1994; Lancaster *et al.*, 1996). It is plausible that LOH on chromosome 13q and 17q in sporadic cancers does not target the *BRCA1* and *BRCA2* gene, but other as yet unidentified tumor suppressor genes in the same vicinity. It is also possible that *BRCA1* and *BRCA2* are inactivated by other mechanisms than the mutations of coding sequences. Reduction of *BRCA1* expression have been reported in a portion of sporadic breast cancer patients and is associated with a malignant phenotype (Holt *et al.*, 1996; Holt, 1997; Seery *et al.*, 1999; Wilson *et al.*, 1999; Yoshikawa *et al.*, 1999). This suggests that reduction or absence of *BRCA1* may contribute to the pathogenesis of a significant percentage of sporadic breast cancers. Aberrant methylation of CpG island associated with transcriptional promoter of the gene has been purposed to inactivate the gene and underline such reduction. A number of studies have demonstrated the aberrant methylation of the *BRCA1* promoter region in a subset of sporadic breast cancers that showed reduction or absence of *BRCA1* expression and mostly in advanced stage (Dobrovic and Simpfendorfer, 1997; Magdinier *et al.*, 1998; Mancini *et al.*, 1998; Rice *et al.*, 1998; Catteau *et al.*, 1999;

Esteller *et al.*, 2000). However, not all sporadic breast cancer with reduced *BRCA1* expression exhibited promoter CpG methylation, suggesting other inactivated mechanism<sup>ing</sup> could exist. Methylation of the putative *BRCA2* promoter has not been detected in a large series of sporadic breast cancers (Collins *et al.*, 1997). *BRCA1* and *BRCA2* are therefore likely to play a role in carcinogenesis only in a subset of sporadic breast cancer, and that indicates the existent of other responsible genes.

Amplification and/or overexpression of proto-oncogenes are also the mechanisms that promote tumor progression. Of many well-described oncogenes to date, only a few may have<sup>a</sup> role in breast cancer progression that may contribute to the prognostic implication. The most studied oncogene in breast cancer is *ERBB2*, which also known as *HER2* or *neu*. *ERBB2* located at chromosome 17q12, encodes a tyrosine kinase growth factor receptor with high homology to epidermal growth factor receptor (Coussens *et al.*, 1985; Jardines *et al.*, 1993). It was shown in transgenic mice that activation or overexpression of *ERBB2* resulted in the genesis of mammary tumors (Bouchard *et al.*, 1989; Muller *et al.*, 1988). Amplification and overexpression of *ERBB2* have been reported in ~ 20% of sporadic invasive breast cancers (Slamon *et al.*, 1987; Ali *et al.*, 1988; Tandon *et al.*, 1989; Borg *et al.*, 1991) and in 40-60% of comedocarcinoma (van de Vijver *et al.*, 1988; Allred *et al.*, 1992). Comedocarcinoma is a specific histological type of ductal carcinoma *in situ* (DCIS) of the breast that has a higher proliferation rate than other DCIS type, resulting in greater clinical severity and microinvasion (Lagios *et al.*, 1989). The higher frequency of *ERBB2* overexpression in comedo type of DCIS than that in invasive cancer may suggest that this type of DCIS is a separate biological entity with its own tumor progression, involving early and preponderant deregulation of *ERBB2*. Alternatively, overexpression of *ERBB2* may decline in the later stage of tumorigenesis, assuming that comedocarcinomas are the

precursor of invasive cancers, or development of invasive carcinoma is *de novo* without passing through a comedo-type DCIS.

Amplification of the proto-oncogene *MYC*, also known as *c-Myc*, was first reported in a cultured cell line from breast cancer. It was also shown that overexpression of *c-Myc* in transgenic mice resulted in mammary tumors (Muller *et al.*, 1988). Approximately 20-25% of primary breast cancers have been shown to have *c-Myc* amplification (Escot *et al.*, 1986; Varley *et al.*, 1987; Berns *et al.*, 1992; Borg *et al.*, 1992), while ~ 80% of tumors have shown overexpression without amplification (El-Ashry and Lippman, 1994). This may indicate that *c-Myc* overactivation in mammary tumor cells is by a mechanism other than gene amplification. Amplification of *c-Myc* has been reported to be associated with high-grade breast cancers (Varley *et al.*, 1987). However, studies in lymph node metastases from patients whose primary tumor showed *c-Myc* amplification, failed to detect such amplification in metastatic cells, suggesting that amplification may be an early event in tumor progression and is not a prerequisite for metastatic phenotype (Shiu *et al.*, 1993). The proto-oncogene *c-Myc*, localized at chromosome 8q24, encodes a protein characteristic of a transcription factor that regulate the expression of a number of genes involved in cell cycle progression and apoptosis (Ryan and Bernie, 1996). For example, MYC transactivates the expression of the *Cdc25a*-phosphatase gene the product of which prevents the inhibitory phosphorylation of Cdk2 and Cdk4, components of cell cycle progression (Galaktionov *et al.*, 1996). MYC protein also down-regulates expression of p27<sup>KIP1</sup>, a Cdk2 inhibitor (Alevizopoulos *et al.*, 1997).

Amplification of the chromosome region 11q13 has been observed in about 15-20% of breast cancers (Lammie and Peters, 1991; Schuuring *et al.*, 1992), and has been shown to be involved in the local recurrence of primary cancers (Champème *et al.*, 1995). Among many genes identified in this region only the *cyclin D1* gene (*CCND1*) is likely the target of such

amplification in breast cancer and is overexpressed in ~ 45% of breast carcinomas, most of which are both estrogen and progesterone receptor positive (Gillett *et al.*, 1994; Bartkova *et al.*, 1994). Cyclin (cyc) is a protein family involved in driving the various stages of the cell cycle by forming an active complex with cyclin dependent kinase (Cdk). CyclinD1 in association with Cdk4/6, as the cycD1-Cdk4/6 complex, facilitates the transition from G1 to S phase by phosphorylation inactivation of the RB protein (Ewen, 1994; Weinberg, 1995). Transgenic mice carrying the *CCND1* gene driven for overexpression have been shown to alter mammary cell proliferation and develop mammary adenocarcinomas (Wang *et al.*, 1994). Further studies have shown that mice homozygously null for *CCND1* failed to undergo the proliferative changes of the mammary epithelium in association with pregnancy (Sicinski *et al.*, 1995). This may indicate the role for *CCND1* in steroid-induced proliferation of the mammary epithelium.

### 1.5.3 Loss of Heterozygosity and Sporadic Breast Cancer

Cytogenetic analysis of breast cancer has identified a number of chromosomal changes. The most frequent is numerical alterations including trisomies of chromosomes 7, 18 and 20, and monosomies of 8, 11, 13, 16, 17, 22, and X (reviewed in Devilee and Cornelisse, 1994). However, the most common chromosomal aberrations observed in near-diploid tumors without metastases are trisomy 7, loss of 17 and 19, and over-representation of 1q, 3q, and 6p (Thompson *et al.*, 1993). Structural changes including terminal deletion and unbalanced non-reciprocal translocations are most frequently involved chromosomes 1, 6 and 16q. Furthermore, the breakpoints of structural aberrations cluster to several segments, including 1p22-q11, 3p11, 6p11-13, 7p11-q11, 8p11-q11, 16q, and 19q13 (Thompson *et al.*, 1993). Chromosomal aberrations such as chromosome loss and partial deletion provide evidence for somatic genetic alterations that may affect the genes involved in breast carcinogenesis.

Furthermore, with the development of polymorphic genetic markers with established map locations on each chromosome, a more refined genetic analysis has been used to investigate the recurrent loss of specific chromosome regions in breast cancer.

Allelotyping of breast cancer using several chromosomal markers, both RFLP (restriction fragment length polymorphism) and STRP (short tandem repeat polymorphism), has identified several regions of allelic imbalance (or loss of heterozygosity: LOH) suggesting the possible locations of tumor suppressor genes. Data compiled from more than 30 studies has revealed a consensus of allelic loss affecting over 11 chromosome arms at a frequency of more than 25% (Devilee and Cornelisse, 1994). The chromosome arms 16q and 17p appeared to be the most frequent loss, affecting more than 50% of tumors, whereas the others such as 1p, 1q, 3p, 6q, 8p, 11p, 13q, 17q, 18q, and 22q were affected at a frequency of 25-40%.

Although these chromosomal losses occur at high frequency in breast cancer, the existing difficulties are to determine the relevance of such losses to breast carcinogenesis. Since in most cases the tumors analyzed were of the invasive type or of advanced stages, it is questionable whether these losses are the causative factors of tumorigenesis or the consequences of general genomic instability associated with advanced cancer. It is also possible that certain losses may be associated with the selective progression or clonal evolution of tumors to a more advanced tumor rather than related to early events related to the genesis of tumor. To overcome this problem, LOH analyses have been conducted in preinvasive breast carcinomas to define regions of allelic loss that may be involved in the early stage of breast cancer. A study of a total of 61 preinvasive breast cancers (DCIS) using 51 polymorphic markers covering 39 chromosome arms has identified significant LOH for marker loci at 8p (18.7%), 13q (18%), 16q (28.6%), 17p (37.5%), and 17q (15.9%) (Radford *et al.*, 1995). Another study performed a comparative LOH analysis in 23 DCIS samples and

29 invasive ductal carcinomas utilizing a total of 20 polymorphic markers for each group of samples (Aldaz *et al.*, 1995). In this study the LOH profiles observed in invasive cancers were similar to those of data compiled from 30 other studies (Devilee and Cornelisse, 1994). However, allelic losses at chromosome regions 1p, 3p, 3q, 6p, 11p, 16p, 18p, 18q, and 22q were not observed or observed at a very low percentage in DCIS samples, as compared with those in invasive carcinomas (Aldaz *et al.*, 1995) suggesting their involvement in relatively late stage of breast cancer progression. Furthermore, the losses at 7p, 7q, 16q, 17p, and 17q appeared to be involved in an early event in breast cancer since they were also observed in 25-30% of DCIS cases (Brenner and Aldaz, 1995). The high incidence of allelic loss at the chromosome 16q region, and the involvement in both early and late stage of breast cancer suggests the existent of tumor suppressor gene or genes in this region. Mutation of this tumor suppressor is likely to significantly contribute to tumorigenesis of a large proportion of breast cancers.

#### **1.5.4 Chromosome 16 and Loss of Heterozygosity in Breast Cancer**

Human chromosome 16 has been predicted by LOH studies to harbor putative tumor suppressor genes. Frequent LOH on the long arm of chromosome 16, 16q, has been reported not only in breast cancer but also in other cancers. These include hepatocellular carcinoma (Tsuda *et al.*, 1990), prostate carcinoma (Carter *et al.*, 1990; Bergerheim *et al.*, 1991; Kunimi *et al.*, 1991), ovarian cancer (Sato *et al.*, 1991b), central nervous system primitive neuroectodermal tumor (Thomas and Raffel, 1991), and Wilm's tumor (Maw *et al.*, 1992). This indicates that significant regions within 16q may harbor tumor suppressor genes that play a common role in tumorigenesis in these tumors.

In breast and prostate cancer, a number of detailed LOH studies, using polymorphic genetic markers saturated at 16q, have been reported and indicated that frequent regions of

LOH consistent for both cancers map to at least three regions at 16q22.1, 16q23.2-q24.1, and 16q24.3-qter (summarized in *Figure 1.2*)

The initial study was performed in 219 sporadic primary breast tumors, of which 193 were invasive ductal carcinomas, utilizing seven RFLP markers spanning the long arm of chromosome 16 (Sato *et al.*, 1991a). A significant fraction of tumors tested, ~51% (78/153) has been identified as having LOH for at least one 16q marker. In this study, a chromosome segment between the HP gene locus and a distal marker, D16S157, was detected as a common region of deletion. Of the tumors tested, it was noted that a group of tumors showing progressive malignant phenotype (eg. large tumor size and lymph node metastasis) also appeared to involve the LOH of markers on other chromosomes. This might indicate that accumulation of genetic alterations contribute to tumor pathogenesis in primary breast cancer.

In an analysis utilizing 18 RFLP markers from 16q, detailed LOH mapping was conducted in 78 breast cancers and identified 38 tumors with 16q LOH (Tsuda *et al.*, 1994). Of 38 cancers showing 16q LOH, 23 of which were suggested by analysis to have partial allelic imbalance on 16q while 15 tumors appeared to have LOH of the entire long arm. The most frequent LOH was observed in the 16q24.2-qter region between the markers D16S43 or D16S155 and the telomere. Further analysis for the association of 16q24.2-qter LOH with clinicopathological parameters of the tumors was undertaken in a total of 234 tumors including those previously tested. The overall incidence of such LOH was then determined to be 52% in the total series of tumors examined. This high incidence of 16q24.2-qter LOH appeared in tumors of all histological grades and clinical stages, ranging from low-grade intraductal tumors to high grade and invasive ductal carcinomas and those with lymph node metastases. This suggested the presence of a tumor suppressor gene involved in the early stage of breast cancer development. Eight tumors exhibited LOH of the markers between



16cen-q22.1 but not at 16q24.2-qter. This group consisted mainly of tumors of an advanced stage with an aggressive phenotype. This suggested the existence of additional inactivation of a tumor suppressor gene in this region and this gene may be involved in the promotion of biologically aggressive phenotypes of breast cancer.

The RFLP markers, though initially utilized as a useful tool for LOH analysis, are quite limited in number, and provide a low degree of genetic map resolution and informativeness. The identification of microsatellite repeats has allowed the use of polymorphic microsatellite markers, also known as STRP markers which are highly polymorphic and distributed across the human genome. These new markers have been used in the construction of detailed genetic linkage maps of the human genome (NIH/CEPH collaborative mapping group, 1992; Weissenbach *et al.*, 1992; Buetow *et al.*, 1994; Gyapay *et al.*, 1994; Matise *et al.*, 1994) and are ideal for detailed LOH studies.

The combination of RFLP markers and new STRP markers spanning chromosome 16q were used in LOH mapping of 79 sporadic breast cancer cases (Cleton-Jansen *et al.*, 1994). In this study, with a total of twenty 16q markers selected, 63% of tumors (50/79) showed LOH for at least one marker and about 20% of tumors appeared to have lost the entire long arm. However, in a limited number of tumors examined two common regions of allelic imbalance were observed in the tumors. The first region at 16q24.3, was defined by 7 tumors exhibiting the loss of one or more of 3 markers restricted to the telomeric region, with one tumor indicated the loss between APRT and D16S303. The second region was detected in another 7 tumors displaying the interstitial LOH of 16q markers and was defined by one of these tumors to be restricted at 16q22.1 between markers D16S398 and D16S301. These two LOH regions correspond with those reported by Tsuda *et al.* (1994), suggesting the existence of at least two tumor suppressor genes involved in the genesis of breast cancer.

In addition, the LOH region presented at 16q22.2-q24.2 (Sato *et al.*, 1991a) may indicate yet another region for a tumor suppressor gene.

Similar results were also obtained from another study in 150 sporadic primary breast tumors, where four 16q markers identified 101 tumors (67%) showing allelic imbalance for at least one of these markers (Skirnisdottir *et al.*, 1995). Further analysis with additional markers on these 101 tumors revealed that 53 had partial or interstitial loss. Analysis of the LOH pattern in the tumors exhibiting restricted loss has identified two common regions, both appear to overlap with those in the previous reports (Tsuda *et al.*, 1994; Cleton-Jansen *et al.*, 1995). The first region lies between D16S413 and telomere, at 16q24.3, and the second region spanning from markers D16S260 to D16S265, which covering the 16q22.1 region.

The high incidence of 16q LOH and the common region of restricted loss were also confirmed by another independent studies of sporadic breast carcinomas. Allelic imbalance mapping was undertaken in a set of 46 primary tumors with nine 16q markers and has identified the 16q LOH of one or more markers in 65% (30/46) of these tumors (Dorion-Bonnet *et al.*, 1995). About one third of these 16q LOH tumors involved a region that appeared to overlap with those previously defined LOH. Furthermore, the same group (Driouch *et al.*, 1997) conducted a deletion map of 16q in 24 breast cancer metastases obtained by fine needle biopsy. The results also confirmed three regions that overlapped with the previously identified region of LOHs.

In a study that included 210 sporadic breast cancers and utilized 14 microsatellite markers localized to 16q, similar common LOH regions for the long arm of chromosome 16 were defined. In 67% of these tumors there was LOH for at least one 16q marker (Iida *et al.*, 1997). Two minimal regions were also identified in a sub-set of tumors showing interstitial LOH. The first region located between markers D16S512 and D16S515 at 16q23.2-q24.1

was defined by 7 tumors, and this region overlapped with that previously described (Sato *et al.*, 1991a; Driouch *et al.*, 1997). The second region mapped to 16q24.3 between markers D16S413 and D16S303 was identified in 3 tumors, and this region also agrees with all previous reports.

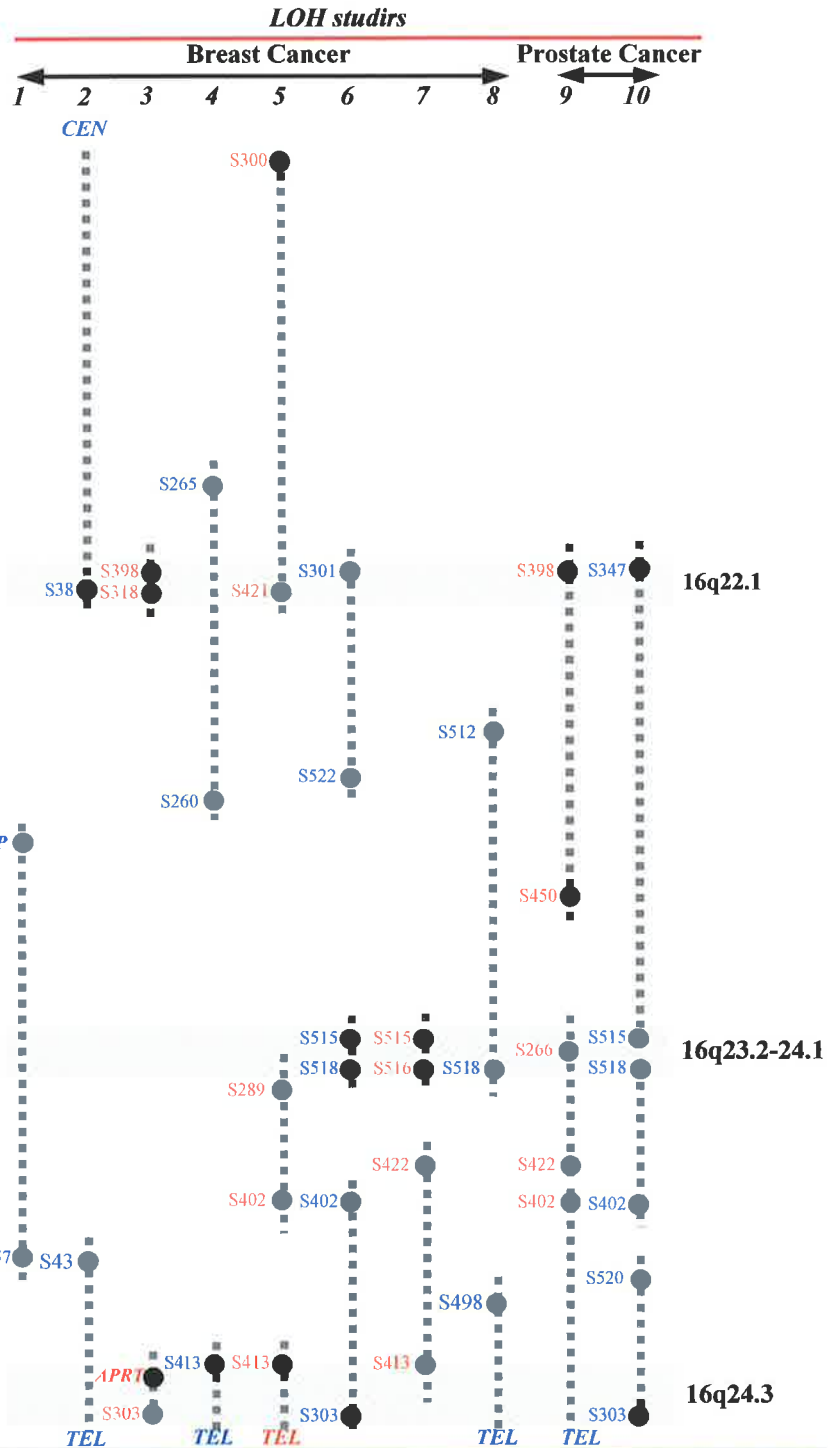
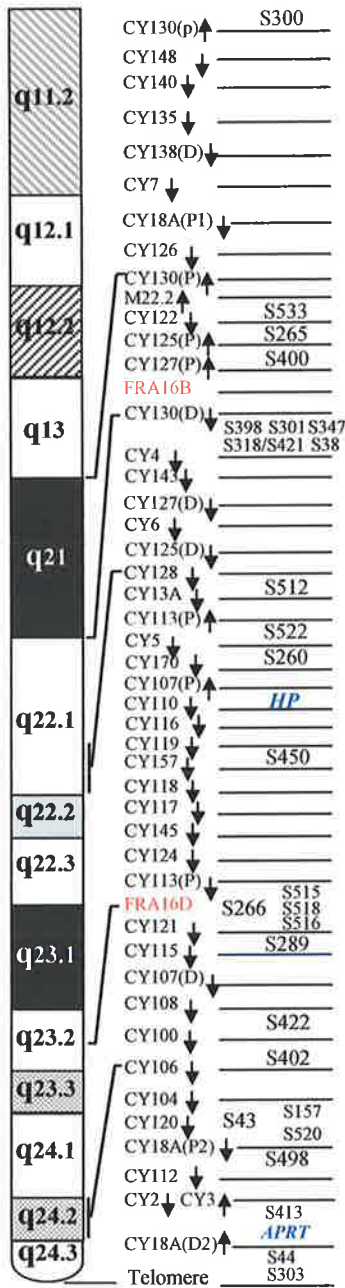
Deletion mapping of chromosome 16q was also undertaken in the early stage breast cancer (DCIS). Studies of 35 tumor samples microdissected from DCIS lesions and the use of 20 microsatellite markers at 16q have identified 31 tumors (89%) with 16q LOH (Chen *et al.*, 1996). The use of microdissected tumor material has provided samples with minimal contamination of normal stromal tissue and may result in the high percentage of allelic imbalance in this study. Although LOHs seen in most of these tumors appeared to be complex losses of 16q markers, the results also showed three overlapping regions of allelic loss commonly reported in other studies. In the previous study of DCIS by the same group (Aldaz *et al.*, 1995), a significant LOH was also observed at the marker D16S413 at 16q24.3 when compared with other loci.

Fine mapping of physical deletion at chromosome 16q in prostate cancers was initially conducted utilizing fluorescence *in situ* hybridization (FISH) (Cher *et al.*, 1995). Region-specific chromosome 16 cosmid contig probes were used to hybridize to the interphase prostatic carcinoma nuclei and identified 15 of 30 tumors (50%) with physical deletion at 16q24. The same group later carried out LOH studies in prostate adenocarcinoma with 10 polymorphic markers on 16q (Godfrey *et al.*, 1997). The data indicated at least two commonly deleted regions at 16q22 and 16q24.2-qter with the frequency of 50% and 56% of informative cases respectively. In a study of 32 paired normal/prostate tumor samples utilizing 11 microsatellite markers mapped to 16q, 16 of 31 informative cases (51%) showed allelic imbalance of at least one of the markers (Osman *et al.*, 1997). The results also

showed that most of the LOH cluster at 6 loci mapped to 16q22.1-23.1. Another report analyzed 48 prostate cancers of both primary and metastases using 17 markers at 16q and showed that 42% of tumors exhibited LOH for at least one marker (Suzuki *et al.*, 1996). Further detailed analysis of LOH map has identified three common regions at 16q22.1-q22.3, 16q23.2-q24.1, and 16q24.3-qter corresponding with those seen in breast cancer. These three commonly LOH regions were also detected in a study of 59 prostate cancer cases of various stages which showed over all 16q LOH in 35 out of 59 tumors (Latil *et al.*, 1997). In the study by Elo *et al.* (1997), examining 50 prostate cancer samples, the region at 16q24.1 was identified to be the most frequent LOH and was associated with clinically aggressive behavior of tumor, metastatic disease, and high-grade tumors.

The recurrent and common regions of LOH established at chromosome 16q as frequently seen in both breast and prostate cancers might indicate the involvement of common tumor suppressor genes. The correlation of allelic imbalance at 16q with some clinico-pathological parameters was also observed in both breast and prostate cancers. In one study of breast cancer, allelic imbalance on 16q has been statistically shown to correlate with a positive estrogen receptor content (Cleton-Jansen *et al.*, 1994). Another group also provided evidence of correlation with a high progesterone receptor content and low S-phase fraction of tumor cells (Skirnisdottir *et al.*, 1995). It was also observed that patients showing a low S-phase fraction have a 19% higher survival rate than those with high S-phase fractions. However these correlations were not reported in other studies. In prostate cancer, the region at 16q24.1 has been identified as the most frequent LOH region associated with a clinically aggressive behavior of tumor, metastatic disease, and high grade tumors (Elo *et al.*, 1997). In contrast, almost all LOH studies in breast cancer have the consistent finding that allelic imbalance on 16q shows no correlation with clinico-

**Chr. 16q Somatic Cell Mapped Hybrids markers**



**Figure 1.2** : Summary of the loss of heterozygosity (LOH) studies involving polymorphic markers that mapped to the long arm of chromosome 16 in sporadic *breast* and *prostate* cancers. The critical regions of LOH based on the analysis of the tumor samples in each study are indicated by *vertical dash lines*. The boundaries for LOH defined by the markers are indicated at both ends of each vertical dash line. Number on top of each group of vertical dash lines indicates each study: *1*. Sato *et al.*, 1991; *2*. Tsuda *et al.*, 1994; *3*. Cleton-Jansen *et al.*, 1994; *4*. Skirnisdottir *et al.*, 1995; *5*. Dorion-Bonnet *et al.*, 1995; *6*. Driouch *et al.*, 1997; *7*. Chen *et al.*, 1996; *8*. Iida *et al.*, 1997; *9*. Suzuki *et al.*, 1996; *10*. Latil *et al.*, 1997. The consensus critical LOH regions for this chromosome arm based on the overlaps of the individual studies are *shaded* and three overlapping regions are defined at 16q22.1, 16q23.2-24.1, and 16q24.3

pathological parameters such as differences in tumor size, histology grade, clinical stage, lymph node status, tumor ploidy, age at diagnosis, or family history of breast cancer.

### 1.5.5 Studies of Candidate Genes localized at 16q Loss of Heterozygosity regions

A number of candidate genes mapped to the 16q LOH regions have been studied for their possible involvement in breast cancers. The epithelial cadherin gene (*CDH1*) which mapped within 16q22.1 region codes for a protein, E-cadherin, important for the maintenance of cell-cell adhesion in epithelial tissues (Mansouri *et al.*, 1988; Berx *et al.*, 1995b). The loss of function of E-cadherin and /or related proteins contributes to the increased proliferation, tumor invasion and metastasis in many solid tumors including breast cancer (Ilyas and Tomlinson, 1997). Mutations of *CDH1* have been reported in breast cancer cell lines (Pierceall *et al.*, 1995; Hiraguri *et al.*, 1998) and over 50% of infiltrating lobular breast tumors (Berox *et al.*, 1995a, 1996). In the latter, mutations have been detected in combination with LOH of the wild-type locus. However, no mutations in *CDH1* have been detected in ductal carcinomas (Kashiwaba *et al.*, 1995; Berx *et al.*, 1995a, 1996). These are the most common type of breast cancer and belong to the major group for which LOH of 16q22.1 has been reported. It is therefore suggested that other candidate genes exist in this region that are involved in the progression of ductal carcinoma of the breast.

The *CTCF* gene located in the 16q22.1 LOH region codes for an evolutionary conserved DNA-binding nuclear protein that has a role in the regulation of vertebrate cellular *MYC* gene expression (Lobanekov *et al.*, 1990; Klenova *et al.*, 1993; Filippova *et al.*, 1996). Over-expression of *MYC* is one of the major oncogenic factors involve in various types of malignancy, including breast cancer progression. Mutations of the genes, including *CTCF*, that regulate *MYC* expression may contribute to the pathogenesis of breast cancer. *CTCF* may therefore be a candidate tumor suppressor for breast carcinogenesis. Genomic

rearrangement of *CTCF* have been reported in two breast cancer cell lines and one of four analyzed paired DNA samples from primary breast cancer patients (Filippova *et al.*, 1998). However, no subsequent additional mutation analysis or other evidence has been published to confirm that *CTCF* is involved in breast cancer.

Two candidate genes have been mapped to the LOH region at 16q23.2-q24.1. The first, a breast cancer anti-estrogen resistance 1 (*BCAR1*) gene has initially been identified through a retroviral insertion mutagenesis of the human estrogen-dependent breast cancer cell line ZR-75-1 (Dorssers *et al.*, 1993). It has been shown that integration of replication defective retrovirus at a common site, the *BCAR1* locus, could transform the cells from estrogen-dependent to an anti-estrogen resistance phenotype. These cells lost estrogen receptor expression and became estrogen-independent (Dorssers *et al.*, 1993). The *BCAR1* was initially mapped to the chromosome band 16q23 by *in situ* hybridization (Dorssers *et al.*, 1993). The gene, *BCAR1*, has subsequently been cloned and mapped to the region between the somatic hybrid breakpoints CY145 and CY117 (Brinkman *et al.*, 2000). This gene, as shown by its protein sequence, is a homologue of the mouse and rat *p130Cas* gene (Brinkman *et al.*, 2000). Overexpression of *BCAR1* has been shown to confer anti-estrogen resistance on ZR-75-1 breast cancer cells and the *in vitro* insertion of retrovirus appears to activate *BCAR1* promoter resulting in its overexpression (Brinkman *et al.*, 2000). In general, about two thirds of all breast cancers produce functionally active estrogen receptor (ER) and are therefore estrogen-dependent (Horwitz *et al.*, 1985; Foekens *et al.*, 1990; Clarke *et al.*, 1997). These tumors respond well to treatment with anti-estrogens which inhibit the estrogen response pathway and thus arrest tumor growth (Jordan *et al.*, 1990; Musgrove *et al.*, 1993). However, nearly half of ER positive tumors develop resistance to this anti-estrogen therapy and eventually the majority of tumors that initially respond to anti-

estrogens develop resistant metastases (Foekens *et al.*, 1994). The involvement of *BCAR1* in breast cancer progression by promoting the anti-estrogen resistant phenotype indicates the likelihood that this gene behaves as a tumor modifier rather than a tumor suppressor.

The second candidate gene has recently been isolated from the genomic region between the markers D16S518 and D16S516 within 16q23.2-q24.1 by mean of physical map construction prior to the identification of the gene (Bednarek *et al.*, 2000). The gene, designated *WFOX*, codes for a protein containing two WW domains coupled to a region with high homology to the short-chain dehydrogenase/reductase family of enzymes. Overexpression of *WFOX* has been demonstrated in breast cancer cell lines when compared with normal breast cells. The high expression of *WFOX* has also been observed in other hormonally regulated tissues such as testis, ovary, and prostate. By its expression pattern, the presence of short-chain dehydrogenase/reductase domain and the specific amino acid features suggest the role for *WFOX* in steroid metabolism. However, mutation studies failed to detect any alteration of this gene in breast cancer showing hemizygous for the 16q genomic region, suggesting that *WFOX* is probably not a tumor suppressor gene (Bednarek *et al.*, 2000).

The LOH region at 16q24.3 has initially been mapped between the marker D16S413 or the *APRT* gene and the marker D16S303 close to the telomere (Cleton-Jansen *et al.*, 1994; Skirnisdottir *et al.*, 1995; Dorion-Bonnet *et al.*, 1995). This region covers the 16q segment between mouse/human somatic cell hybrid breakpoint CY18A(D2) and the telomere (Callen *et al.*, 1995). Within this region, a number of known genes have been localized and some of these genes may be a candidate tumor suppressor, which requires further analysis. A collaborative study to refine the LOH at 16q24.3 utilizing additional polymorphic markers distributed in this region has initially provided evidence suggesting the possible minimal



LOH region between markers D16S3026 and D16S303 (Cleton-Jansen, personal communication). Concurrently, a detailed physical and transcription map has been established that encompasses this minimal region (Whitmore *et al.*, 1998a). The construction of this 16q24.3 physical and transcription map and the studies of some known genes mapped to this region are discussed in the next section.

## **1.6 Physical and Transcription map of 16q24.3 and Studies of Candidate Genes**

The construction of the detail physical and transcription map at 16q24.3 region was initiated to provide a basis for the isolation of candidate tumor suppressor gene or genes (Whitmore *et al.*, 1998a). This was the minimum 16q24.3 region defined by LOH studies.

### **1.6.1 Construction of the Physical and Transcription map**

Initially, a total of nine separate markers that mapped between the somatic hybrid breakpoint CY18A(D2) and the 16q telomere (Callen *et al.*, 1995) were used as probes to screen a chromosome 16 cosmid library. The identified cosmid clones were organized into six groups of cosmid contigs as nucleations for subsequent cosmid walking. Cosmid walking and restriction site mapping was successfully used to extend and link these clusters. This generated a physical map of approximately 950 kb extending from the marker D16S303 to the region proximal to D16S3026. This physical map contained a total of 103 overlapping clones with a minimum tiling path of 35 cosmid clones (Whitmore *et al.*, 1998a).

To identify the transcripts within this physical map, exon trapping (Buckler *et al.*, 1991; Burn *et al.*, 1995) was used on selected overlapping cosmids. A total of 71 trapped exons were identified, 17 of which were identical to characterized genes known to map to this LOH region (Whitmore *et al.*, 1998a). Among these are the *FAA*, *PISSLRE*, and *PRSM1* genes. The DNA sequences of 23 clones were homologous to ESTs in the sequence database

at NCBI and were grouped into 15 different transcripts. The remaining 31 trapped exons showed no database sequence homologies, although 23 had the open reading frames suggesting that they were parts of genes.

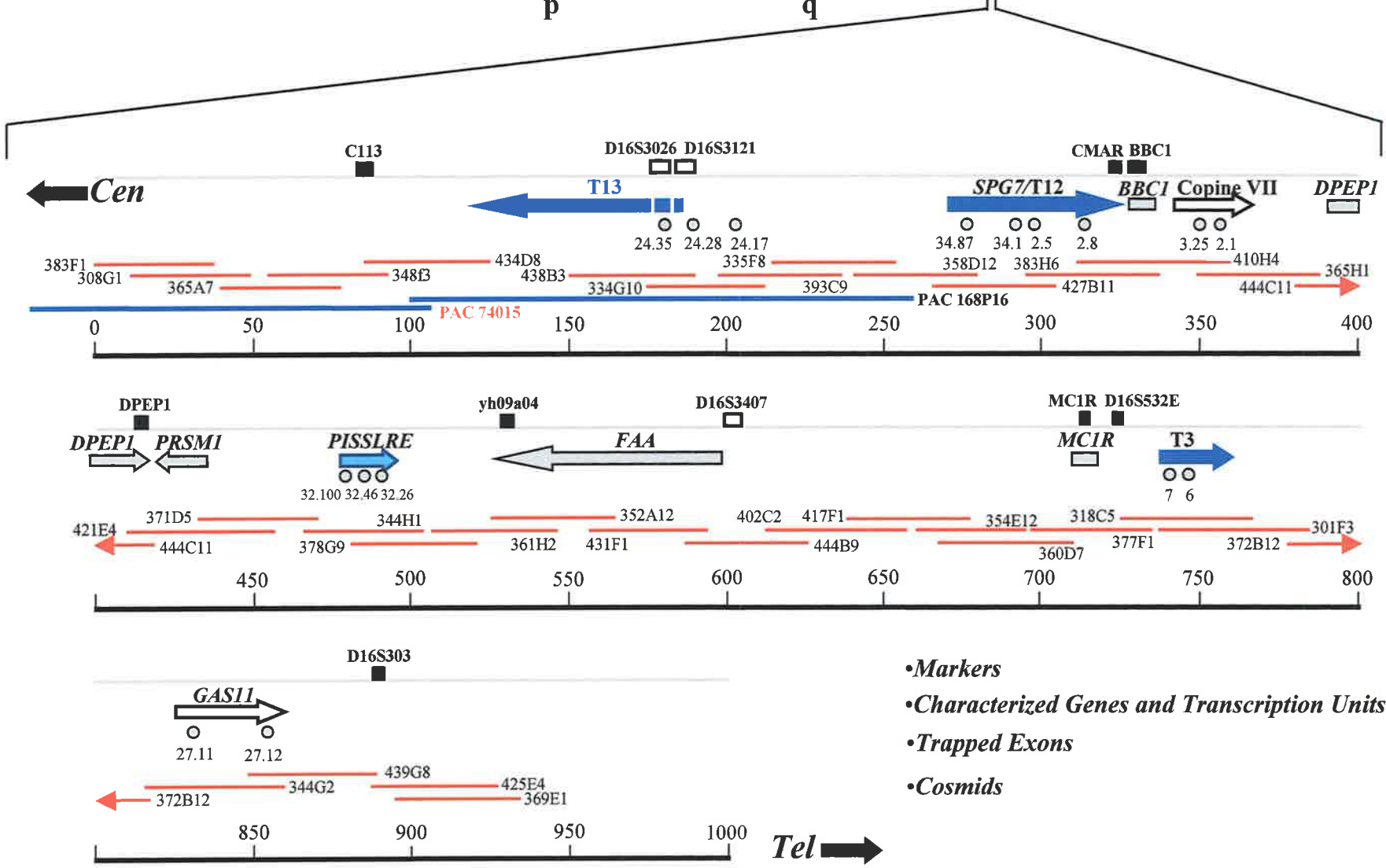
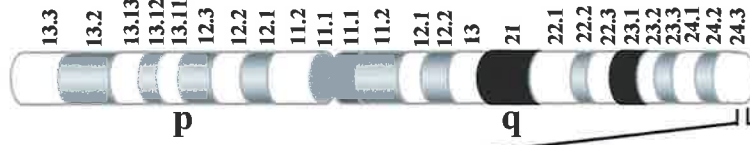
The physical map that was established in this chromosomal region provided a genomic framework for subsequent gene identification. The trapped exons were mapped back to the original genomic clones to form exon clusters, each represented a possible transcript with a known physical location. This integrated physical and transcription map of 16q24.3 LOH region (Whitmore *et al.*, 1998a, and summarized in *Figure 1.3*) provides the basis for the identification of the tumor suppressor gene related to breast cancer development and progression. Furthermore, this map can also be used for the isolation of genes that may be involved in other genetic diseases. One such example is the gene for fanconi anemia complimentary group A (*FAA*). The *FAA* gene was isolated during the initial stage of physical map construction utilizing both exon trapping and cDNA selection procedures (The Fanconi anaemia/Breast cancer consortium, 1996).

### 1.6.2 Studies of the Genes within the 16q24.3 minimal LOH region

At the early stage of this study a number of known genes had been located to this region, these included *BBC1*, *CMAR*, *DPEP1*, and *MC1R*. These genes were possible candidates as a tumor suppressor and therefore required further investigation.

The breast basic conserved gene (*BBC1*) encodes a protein showing sequence homology to the human ribosomal protein L13 and decreased expression of *BBC1* mRNA has been detected in malignant breast tumors compared to that in benign tumors (Adams *et al.*, 1992). The L13 protein has been shown to contain a DNA binding motif characteristic of transcription factors (Tsurugi and Mitsui, 1991), suggesting *BBC1* is unlikely to be a candidate tumor suppressor gene targeted by LOH. Mutation analysis of this gene, however,

**Chromosome 16**



- Markers
- Characterized Genes and Transcription Units
- Trapped Exons
- Cosmids

has been undertaken in breast cancer samples exhibiting LOH restricted to 16q24.3, but no evidence of mutations were detected in these samples (Moerland *et al.*, 1997).

The cellular matrix adhesion regulator (*CMAR*) gene was originally isolated from a colorectal cell line cDNA library and has been shown by gene transfection to enhance the binding of the cell to the extracellular matrix component, collagen type 1 (Pullman and Bodmer, 1992). Since *CMAR* is involved in cell adhesion, then loss of *CMAR* function may be associated with the initial step of tumor invasion and metastasis. However, mutation analysis in breast cancer has excluded *CMAR* as a candidate breast tumor suppressor (Moerland *et al.*, 1997). Furthermore, *CMAR* sequence has now been shown to be a part of the 3' untranslated region of the gene involved in spastic paraplegia, *SPG7*, isolated from this chromosome region (discussed in *chapter 4*).

The renal dipeptidase gene (*DPEP1*) was originally isolated from a kidney cell cDNA library by subtractive hybridization with Wilms' tumor mRNA (Austruy *et al.*, 1993). Reduced expression of *DPEP1* mRNA was shown in Wilms' tumor compared with its expression in mature kidney. However no further evidence of its involvement in tumorigenesis has been reported, it is therefore not a target gene for breast cancer LOH. The melanocyte stimulating hormone receptor (*MC1R*) gene may not be considered as a candidate gene, since its function is associated with the regulation of pigmentation in keratinocyte (Valverde *et al.*, 1995).

## **1.7 Aims and Scope of The Study**

The strong association of LOH at the chromosome region 16q24.3 and sporadic breast cancer lead to the establishment of chromosome 16-breast cancer project with the ultimate goal for cloning of the breast cancer-associated tumor suppressor gene. Following the successful positional cloning of the *FAA* gene, the detail physical map at this region has

been expanded encompassing <sup>a</sup>the region of approximately 950 kb. This physical map is bounded proximally by D16S3026 and distally by D16S303. Subsequently, the transcript map has been established on this physical framework and provided the basis for cloning of candidate gene(s).

The aim of the project presented in this study was to clone and characterize the genes from this physical region for the identification of the candidate tumor suppressor associated with sporadic breast cancer. This will primarily involve the isolation of full-length transcript sequence of the genes. Expression profile of each gene will also be established. Characterization of the gene sequence and its protein product by sequence homology as well as functional motif searching may provide a possible function of the corresponding protein. The candidate genes based on the possible function compatible with a tumor suppressor will be further characterized by mutation screening of DNA isolated from tumors defining the critical LOH region at 16q24.3.

Other characterized genes that are not involved in breast cancer but show protein sequence homology to known gene with significant physiological function will also be studied in detail.

## **Chapter 2**

### **Materials and Methods**

---

## Table of Contents

	Page
<b>2.1 Materials</b>	54
2.1.1 Fine Chemicals and Compounds and Their Suppliers	54
2.1.2 Enzymes	55
2.1.3 Electrophoresis	56
2.1.4 Antibiotics	56
2.1.5 Bacterial Strains	56
2.1.6 Media	57
2.1.6.1 <i>Liquid Media</i>	57
2.1.6.2 <i>Solid Media</i>	57
2.1.7 DNA vectors	58
2.1.8 DNA Size Markers	58
2.1.9 Radioactive Chemicals	58
2.1.10 Miscellaneous Materials and Reagent Kits	59
2.1.11 Buffers and Solutions	59
<b>2.2 Methods</b>	62
2.2.1 Isolation of DNA	62
2.2.1.1 <i>Cosmid DNA</i>	62
2.2.1.2 <i>BAC and PAC DNA</i>	63
2.2.1.3 <i>Plasmid DNA</i>	64
2.2.1.4 <i>Genomic DNA</i>	64
2.2.2 Isolation of Total RNA	65
2.2.2.1 <i>RNA Isolation from Fresh Tissues</i>	65
2.2.2.2 <i>RNA isolation from Cell Lines</i>	66
2.2.3 Oligonucleotide Primers	67
2.2.4 Quantitation of DNA, RNA and Oligonucleotide Primers	67
2.2.5 Preparation of Bacterial Glycerol Stocks	67
2.2.6 Preparation of Sheared Human Placenta and Salmon Sperm DNA	68
2.2.7 Restriction Endonuclease Digestion	68
2.2.8 Agarose Gel Electrophoresis	69
2.2.9 Southern Blotting	69
2.2.9.1 <i>Transfer in 10X SSC</i>	69
2.2.9.2 <i>Transfer in 0.4 M NaOH</i>	70

2.2.10 Radio-Isotope Labelling of DNA	70
2.2.10.1 Labelling DNA fragments in Solution	70
2.2.10.2 Labelling DNA fragments in Agarose	71
2.2.10.3 Pre-reassociation of Labelled Probes	71
2.2.11 Hybridization of Nucleic acid bound membranes	72
2.2.11.1. Southern Hybridization	72
2.2.11.2 Northern Hybridization	72
2.2.11.3 Washing of Membranes	72
2.2.11.4 Stripping of Membranes for Re-Use	73
2.2.12 Polymerase Chain Reaction (PCR)	73
2.2.12.1 Standard PCR Reactions	74
2.2.12.2 Colony PCR Reactions	74
2.2.13 Reverse Transcription PCR (RT-PCR)	75
2.2.13.1 Reverse Transcription of cDNA	75
2.2.13.2 PCR amplification of cDNA	76
2.2.14 Purification of DNA Fragments	76
2.2.14.1 Purification of PCR Products and Radio-Labelled Probes	76
2.2.14.2 Purification of DNA Fragments from Agarose Gels	77
2.2.15 Cloning of DNA Fragments	78
2.2.15.1 DNA Ligation into Plasmid Vector	78
2.2.15.2 Ligation of PCR Products into the pGEM-T Vector	78
2.2.15.3 Preparation of Competent Bacterial Cells	78
2.2.15.4 Transformation of Competent Bacterial Cells	79
2.2.15.5 Preparation of Colony Master Plates and Colony Lifting	79
2.2.16 DNA Sequencing	80
2.2.16.1 Dye Terminator Cycle Sequencing	80
2.2.16.2 Dye Primer Cycle Sequencing	81
2.2.17 Rapid Amplification of cDNA Ends (RACE)	82
2.2.17.1 5' RACE	82
2.2.17.2 3' RACE	83
2.2.18 Screening of cDNA library	83
2.2.18.1 PCR screening of Formatted Fetal Brain cDNA Library	84
2.2.18.2 Screening cDNA Library by Hybridization	84
2.2.19 Bioinformatics	85



## **2.1 Materials**

### **2.1.1 Fine Chemicals and Compounds and Their Suppliers**

Ammonium sulphate	Ajax Chemicals, Australia
Bacto-agar	Sigma Chemical Co., USA
Bacto-tryptone	Difco Laboratories, USA
Boric acid	Ajax
BSA (bovine serum albumin-pentax fraction V)	Sigma
Chloroform	BDH Lab Supplies
DAPI (diamidino phenylindole dihydrochloride)	Sigma
DEPC (diethylpyrocarbonate)	BDH
Deoxy-nucleotide triphosphate (dNTPs)	Amersham Biosciences
Dextran sulphate	Amersham Biosciences or Promega
<i>N,N</i> -dimethyl formamide	Sigma
DMEM (Dulbecco's modified Eagles medium)	Trace Biosciences
DMSO (dimethylsulphoxide)	Sigma
DTT (dithiothreitol)	Sigma
EDTA (ethylenediamine tetraacetic acid; Na <sub>2</sub> EDTA.2H <sub>2</sub> O)	Ajax
Ethanol (99.5% v/v)	BDH
FCS (fetal calf serum)	Trace Biosciences
Ficoll (Type 400)	Sigma
Formamide (deionised before use)	Fluka Chemika
Glucose	Ajax
Glutamine	Trace Biosciences
Glycerol	Ajax
Human placental DNA	Sigma
IPTG (isopropylthio- $\beta$ -D-galactosidase)	Progen
Isoamyl alcohol (IAA)	Ajax
Isopropanol	Ajax
<i>N</i> -lauroylsarcosine (Sarkosyl)	Sigma
LipofectACE	Gibco-BRL (Life Technology)
Magnesium chloride (MgCl <sub>2</sub> .6H <sub>2</sub> O)	Ajax
Magnesium sulphate (MgSO <sub>4</sub> .7H <sub>2</sub> O)	Ajax

Maltose	BDH
$\beta$ ME ( $\beta$ -mercaptoethanol)	BDH
Mixed bed resin (20-50 mesh)	BioRad, USA
MOPS (3-[N-morpholino]propane sulphonic acid)	Sigma
Opti MEM-1 (powder sachet)	Gibco-BRL
Paraffin oil	Ajax
PEG (polyethylene glycol) 3350	Sigma
Phenol	Wako Ind., Japan
Polyvinylpyrrolidone (PVP-40)	Sigma
Potassium dihydrogen orthophosphate ( $\text{KH}_2\text{PO}_4$ )	Ajax
Propidium iodide	Sigma
Salmon sperm DNA	Calbiochem
Sodium acetate	Ajax
Sodium carbonate ( $\text{NaHCO}_3$ )	Astra Pharm., Australia
Sodium chloride	Ajax
tri-Sodium citrate	Ajax
Sodium dihydrogen orthophosphate ( $\text{NaH}_2\text{PO}_4 \cdot 2\text{H}_2\text{O}$ )	Ajax
Sodium dodecyl sulphate (SDS)	BDH
di-Sodium hydrogen orthophosphate ( $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$ )	Ajax
Sodium hydroxide (NaOH)	Ajax
Sucrose	Ajax
Tris-base	Boehringer Mannheim
Tris-HCl	Boehringer Mannheim
Triton X-100	Ajax
Yeast extract	Gibco-BRL

### 2.1.2 Enzymes

All restriction enzymes were purchased from either New England Biolabs (Beverly, Massachusetts, USA) or Progen (Brisbane, Queensland, Australia). Each enzyme was supplied with the appropriate digestion buffer and bovine serum albumin (BSA) if required. All other enzymes not part of kits are listed below along with their suppliers.

CIAP (Calf intestinal alkaline phosphatase)	Boehringer Mannheim
---	---------------------

<i>E. coli</i> DNA polymerase (Klenow fragment)	New England Biolabs
Proteinase K	Merck
RNaseH	Promega
RNasin	Promega
Superscript™ RNase H <sup>-</sup> reverse transcriptase	Gibco-BRL
T4 DNA ligase	Progen
T4 Polynucleotide kinase	Amersham Biosciences
<i>Taq</i> DNA polymerase	Gibco-BRL
X-gal (5-Bromo-4-chloro-3-indoyl- $\beta$ -D galactosidase)	Progen

### 2.1.3 Electrophoresis

Acrylamide (40%)	BioRad
Agarose: Molecular Biology grade	Progen or Amersham Biosciences
APS (Ammonium persulphate)	BioRad
Bis (2% Bis solution)	BioRad
Bromophenol blue	BDH
Ethidium bromide	Sigma
Formaldehyde (37%)	Ajax
MDE (2X gel solution)	FMC bioproducts, USA
TEMED ( <i>N,N,N',N'</i> -tetramethylethylenediamine)	BioRad
Xylene cyanol	BDH

### 2.1.4 Antibiotics

Ampicillin	Sigma
Benzyl penicillin	CSL
Chloramphenicol	Sigma
Kanamycin	Progen
Tetracycline	Sigma

### 2.1.5 Bacterial Strains

All bacterial strains were kept at -70 °C in LB-Broth + 15% glycerol. Each strain was

streaked for single colonies on LB-Agar plates containing an appropriate antibiotic (see below) from stock.

*E. coli* XL1-Blue MRF':  $\Delta(mcrA)183 \Delta(mcrCB-hsdSMR-mrr)173 \text{ endA1 supE44 thi-1 relA1 lac}$  [F' *proAB lacI<sup>f</sup>ZAM15 Tn10 (Tet<sup>r</sup>)*]<sup>c</sup>. This strain was a host for recombinant plasmids and was purchased from Stratagene. XL1-Blue cells were streaked for single colonies on LB-tetracycline plates then propagated overnight at 37 °C in LB-Broth supplemented with tetracycline (15 µg/ml).

*E. coli* Y1090r-:  $\Delta(lac)U169 \Delta(lon)? AraD 139 \text{ strA supF mcrA trpC22::Tn10 (Tet<sup>r</sup>)}$  [pMC9 Amp<sup>r</sup> Tet<sup>r</sup>] *mcrB hsdR*. (Note: pMC9 is pBR322 with *lacI<sup>f</sup>* inserted). This strain was a host for recombinant  $\lambda$ gt 11 bacteriophage and was supplied with the Clontech 5' Stretch fetal brain cDNA library. These cells were streaked for single colonies on LB-tetracycline plates and subsequently propagated overnight at 37 °C in LB-Broth supplemented with 10 mM MgSO<sub>4</sub> and 0.2% (v/v) maltose.

## 2.1.6 Media

All liquid media was prepared with double distilled water and was sterilised by autoclaving. For the preparation of solid media plates, antibiotics were added once the autoclaved media had cooled to a temperature of approximately 50 °C.

### 2.1.6.1 Liquid Media

LB-Broth (Luria-Bertaini Broth): 1% (w/v) Bactotryptone; 0.5% (w/v) yeast extract; 1% (w/v) NaCl; adjust pH to 7.5 with 5 N NaOH before autoclave.

OMI (pH 7.1): 1 sachet of Opti MEM-1; 28 ml NaHCO<sub>3</sub>; 80 ml FCS; 55 nM βMe; 1 ml benzyl penicillin, *per* litre.

Supplemented DMEM (pH 7.0): 13 g DMEM; 44 ml NaHCO<sub>3</sub>; 4 mM glutamine; 1 ml benzyl penicillin; 100 ml FCS, *per* litre.

TSS: LB-Broth; 10% (w/v) PEG 3350; 5% (v/v) DMSO; 50 mM MgCl<sub>2</sub>.

### 2.1.6.2 Solid Media

LB-Agar: LB-Broth; 1.5% (w/v) Bacto-agar.

LB-Ampicillin: LB-Broth; 1.5% (w/v) Bacto-agar; 100 µg/ml ampicillin.

LB-Chloramphenicol: LB-Broth; 1.5% (w/v) Bacto-agar; 34 µg/ml chloramphenicol.

LB-Kanamycin: LB-Broth; 1.5% (w/v) Bacto-agar; 50 µg/ml kanamycin.

LB-Tetracycline: LB-Broth; 1.5% (w/v) Bacto-agar; 15 µg/ml tetracycline.

Top Agar: LB-Broth; 0.7% Bacto-agar.

### 2.1.7 DNA Vectors

pSPL3B was used for exon trapping from BAC and PAC DNA. This vector was kindly provided by Dr. Tim Connors and Dr. Greg Landes (USA).

pUC19 was used for the subcloning of cosmid DNA fragments and was purchased from New England Biolabs.

### 2.1.8 DNA Size Markers

λPPI (EcoRI digested )	Progen or Bresatec
pUC19 (HpaII digested)	Bresatec
Digest III (λcI857 Sam 7 HindIII/ φX-174 HaeIII restricted)	Amersham Biosciences

### 2.1.9 Radioactive Chemicals

α- <sup>32</sup> P]dCTP (3,000 Ci/mmol)	Amersham Biosciences
[γ- <sup>32</sup> P]dATP (5,000 Ci/mmol)	Amersham Biosciences

### 2.1.10 Miscellaneous Materials and Reagent Kits

ABI Prism™ DyePrimer Cycle sequencing kit	Perkin Elmer
ABI Prism™ DyeTerminator Cycle sequencing kit	Perkin Elmer
BigDye Terminator Cycle sequencing kit	Perkin Elmer
Breast cancer cell lines	ATCC, USA
CloneAMP pAMP10 kit	Gibco-BRL
ExpressHYB solution	Clontech
Genescreen Plus Nylon membranes	Dupont
<i>HalfTERM</i> DyeTerminator sequencing reagent	GenPack
Labelling mix-dCTP	Amersham Biosciences
Multiple tissue Northern Blot (Catalogue # 7760-1)	Clontech
Human RNA Master Blot (Catalogue # 7770-1)	Clontech
Oligo-dT <sub>12-18</sub>	Gibco-BRL
pGEM-T cloning kit	Promega
Prep-A-Gene purification kit	BioRad
Qiagen Plasmid mini kit	Qiagen
QIAquick purification kit	Qiagen
SMART™ PCR cDNA Library Construction kit (Catalogue # K1051-1)	Clontech
Random hexamers	Amersham Biosciences or Perkin Elmer
<i>RTth</i> reverse transcriptase RNA PCR kit	Perkin Elmer
Trizol	Gibco-BRL
3MM Whatman paper	Lab Supply

### 2.1.11 Buffers and Solutions

All solutions were prepared with sterile double distilled water followed by autoclaving at 120 °C for 15 to 30 minutes. Solutions provided with kits are not mentioned. Buffers and solutions routinely used in this study were as follows:

10X Agarose gel loading buffer: 100 mM Tris-HCl (pH 8.0); 200 mM EDTA (pH 8.0); 2.0% (w/v) sarkosyl; 15% (w/v) ficoll 400; 0.1% bromophenol blue; 0.1% xylene cyanol.

1X pAMP10 annealing buffer: 20 mM Tris-HCl (pH 8.4); 50 mM KCl; 1.5 mM MgCl<sub>2</sub>.

Cell lysis buffer: 0.32 M sucrose; 10 mM Tris-HCl; 5 mM MgCl<sub>2</sub>; 1% (v/v) Triton X-100 (pH 7.5).

Colony denaturing solution: 1.5 mM NaCl; 0.5 M NaOH.

Colony neutralising solution: 3 M NaCl; 0.5 M Tris-HCl.

De-ionised formamide: 2 g of mixed bed resin/50 ml formamide. Mix for ~1 hour then filter.

100X Denhart's solution: 2% (w/v) ficoll 400; 2% (w/v) polyvinylpyrrolidone; 2% (w/v) BSA.

10X dNTP labelling solution: 1 mM labelling mix -dCTP.

Filter denaturing solution: 0.5 M NaOH.

Filter neutralising solution: 0.2 M Tris-HCl; 2X SSC.

5X First strand buffer (supplied with Superscript reverse transcriptase): 250 mM Tris-HCl (pH 8.3); 375 mM KCl; 15 mM MgCl<sub>2</sub>.

Formamide loading buffer: 50% (v/v) glycerol; 1 mM EDTA (pH 8.0); 0.25% (w/v) bromophenol blue; 0.25% (w/v) xylene cyanol.

Gel denaturing solution: 2.5 M NaCl; 0.5 M NaOH.

Gel neutralising solution: 1.5 M NaCl; 0.5 M Tris-HCl (pH 7.5).

10X Labelling buffer: 0.5 M Tris-HCl (pH 7.5); 0.1 M MgCl<sub>2</sub>; 10 mM DTT; 0.5 mg/ml BSA; 0.05 A<sub>260</sub>/μl random hexamers (Amersham Pharmacia Biotech).

10X MOPS buffer: 0.2 M MOPS; 50 mM NaAc; 1 mM EDTA.

Phenol: buffered with Tris-HCl (pH 7.4).

PBS (Phosphate buffered saline): 138 mM NaCl; 2.7 mM KCl; 8.1 mM Na<sub>2</sub>HPO<sub>4</sub>·7H<sub>2</sub>O; 1.2 mM KH<sub>2</sub>PO<sub>4</sub> (pH 7.4).

10X PCR buffer (supplied with *Taq* DNA polymerase): 200 mM Tris-HCl (pH 8.0); 500 mM KCl.

2X PCR mix: 33 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>; 133 mM Tris-HCl (pH 8.8); 20 mM βME (added fresh each time); 0.013 mM EDTA; 0.34 mg/ml BSA; 20% (v/v) DMSO; 0.4 mM dNTPs.

3X Proteinase K buffer: 10 mM NaCl; 10 mM Tris-HCl; 10 mM EDTA (pH 8.0).

RNA hydrolysing solution: 50 mM NaOH; 1.5 M NaCl.

10X RNA loading dye: 50% glycerol; 1mM EDTA (pH 8.0); 0.25% bromophenol blue; 0.25% xylene cyanol.

RNA neutralising solution: 0.5 M Tris-HCl (pH 7.4); 1.5 mM NaCl.

SM buffer: 50 mM Tris-HCl; 100 mM NaCl; 8 Mm MgSO<sub>4</sub>; 0.01% (v/v) gelatin (pH 7.5).

3 M Sodium acetate pH 5.2 (NaOAc): 24.6 g sodium acetate/100 ml; pH 5.2 with acetic acid.

Southern hybridization solution: 50% (v/v) de-ionised formamide; 5X SSPE; 2% SDS; 1X Denhart's; 10% (w/v) dextran sulphate; 100 µg/ml denatured salmon sperm DNA.

20X SSC: 3 M NaCl; 0.3 M tri-sodium citrate (pH 7.0).

20X SSPE: 3.6 M NaCl; 0.2 M NaH<sub>2</sub>PO<sub>4</sub>·2H<sub>2</sub>O; 0.02 M EDTA.



1X TBE: 90 mM Tris-base; 90 mM boric acid; 2.5 mM EDTA (pH 8.0).

TE: 10 mM Tris-HCl (pH 7.5); 0.1 mM EDTA.

X-gal: 50 mM (2% w/v) solution, using N,N-dimethyl-formamide as diluent.

## **2.2 Methods**

The methods described in this chapter are generally used in common throughout this thesis and most are based on the standard procedures presented in *Molecular Cloning: A Laboratory Manual*, 2nd Ed (Sambrook 1989) unless otherwise indicated. Those procedures specific for each chapter are presented in detail in the corresponding method section. ✕

### **2.2.1 Isolation of DNA**

All DNA isolations except for genomic DNA were based on the alkaline lysis method adapted from the procedure described in Sambrook *et al.* (1989). The DNA was subsequently purified utilizing Qiagen column technique according to the manufacturer conditions (Qiagen Plasmid Mini Handbook, 1995) with slight modifications. All buffers and columns were supplied with Qiagen kit.

#### **2.2.1.1 Cosmid DNA**

Each cosmid was received as bacterial stab in nutrient agar and was streaked and grown on LB-Kanamycin plate. A single colony was picked out and grown overnight at 37 °C in 200 ml of LB-broth with kanamycin (50 µg/ml). Culture was spun down at 4,000 rpm for 15 minutes and bacterial pellet was resuspended in 2 ml of resuspending buffer (P1). Two ml of lysis buffer (P2) was then added and carefully mixed, and allowed completely lysis by standing at room temperature for 5 minutes. Two ml of buffer P3 was added to cell lysate with gently mixed and followed by 10 minute incubation on ice. This mixture was then spun

at 13,000 rpm for 15 min, the supernatant was transferred to a new tube and spun for another 10 minutes. During the second spin a Qiagen Tip-20 column was equilibrated with 1 ml of QBT buffer. The supernatant from second spin was applied to the equilibrated column and allowed to drain completely through. The column was subsequently washed with 4 ml of buffer QC and the DNA was eluted into a 1.5 ml eppendorf tube with 0.8 ml of buffer QF. To precipitate the DNA, 0.6 ml isopropanol was added, mixed and spun at 13,000 rpm for 30 minutes. The DNA pellet was washed with 1 ml of 70% ethanol, spun for a further 5 minutes, and allowed to air-dried for 20 minutes. The DNA pellet was dissolved in 15 ul of sterile Milli-Q water and quantified by spectrophotometry (2.2.4)

#### **2.2.1.2 BAC and PAC DNA**

Bacterial stabs of BAC and PAC clones were streaked for the isolation of single colonies on LB-Chloramphenicol and LB-Kanamycin plates respectively. Single colony of BAC clone was grown overnight at 37 °C in 200 ml of LB-broth containing chloramphenicol (34 µg/ml). Cells were pelleted the next day by spinning at 4,000 rpm and were subjected to DNA isolation. PAC clone was initially grown overnight at 37 °C in 10 ml of LB-broth with kanamycin (50 µg/ml). The next day, 7 ml of this culture was transferred to 200 ml of new LB-kanamycin broth and grown for 1.5 hours at 37 °C. After the addition of IPTG to a final concentration of 0.5 mM, the culture was grown for further 5 hours and then spun down at 4,000 rpm. Supernatant was removed and cell pellet was left overnight at -20 °C. Isolation of both BAC and PAC DNA was performed by using Qiagen Tip-100 column with the same procedure used for cosmid DNA isolation while the volume of buffer being used was adjusted. These included the use of 4 ml of buffer P1, P2 and P3 per DNA isolation, with the Tip-100 columns were equilibrated with 4 ml QBT buffer, and washed with 10 ml buffer QC. DNA was eluted into 10 ml sterile tube with 5 ml of QF buffer, then

aliquoted into 6 eppendorf tubes. DNA was precipitated with 0.7 volumes of isopropanol, and pellets were washed, as per cosmid DNA preparation. Individual dried DNA pellets were re-dissolved in 5 µl of sterile Milli-Q water, using wide bore pipette tips to prevent shearing of the DNA, then pooled into one tube and quantified by spectrophotometry (2.2.4).

### **2.2.1.3 Plasmid DNA**

For each plasmid clone, a single colony was grown overnight at 37 °C in 20 ml LB-broth containing 100 µg/ml ampicillin. Plasmid DNA was isolated using Qiagen Tip-20 column with all procedures followed the manufacturer conditions (Qiagen Plasmid Mini Handbook, 1995). In the final step, plasmid DNA pellet was re-dissolved in 15 µl of sterile Milli-Q water before spectrophotometric quantification.

### **2.2.1.4 Genomic DNA**

DNA was isolated from peripheral blood lymphocytes, and breast cancer cell lines by the methods adapted from Wyman and White (1980). All DNA preparations were kindly performed by either Shirley Richardson or Jean Spence (Department of Cytogenetics and Molecular Genetics, WCH). All cell lines were maintained by Sharon Lane and Cathy Derwas (Department of Cytogenetics and Molecular Genetics, WCH).

Cell lysis buffer was added either to a blood sample or to a cell pellet obtained from a cell line culture to a volume of 30 ml. The tube of suspended cells was left on ice for 30 minutes, then spun at 3,500 rpm for 15 minutes at 4 °C. Part of supernatant, 20 ml, was removed by aspiration, followed by the addition of a further 20 ml of cell lysis buffer, resuspended and a repeat centrifugation. The supernatant was aspirated and discarded, and 3.25 ml of 3X Proteinase K buffer, 500 µl of 10% (W/V) SDS and 200 µl Proteinase K (at 10 mg/ml) was added and mixed with the pellet. After an overnight incubation at 37 °C with

mixing, 5 ml of phenol was added and followed by a 15-minute incubation at room temperature on a rotator. The solution was spun at 3,000 rpm for 10 minutes, the aqueous layer was removed and transferred to a new clean tube, and phenol was added to a volume of 10 ml. After a 10 minute-incubation at room temperature, the mixture was then centrifuged for 10 minutes, and the aqueous layer was transferred to a clean tube. Chloroform was added to a final volume of 10 ml, the sample was mixed and re-centrifuged, and the aqueous layer was transferred to a new clean tube. DNA was precipitated by mixing with 300  $\mu$ l of 3 M NaOAc (pH 5.2) and 10 ml of absolute ethanol. After centrifugation, DNA pellet was washed with 70% ethanol, air-dried, and dissolved in 100  $\mu$ l of TE.

### 2.2.2 Isolation of Total RNA

The isolation of RNA from all tissues and cell lines was carried out under strict RNase-free conditions. All solutions were treated with 0.2% (v/v) DEPC, overnight at room temperature with shaking before <sup>being</sup> autoclaved. All pipettors were cleaned with RNaseZAP (Ambion) before used in combination with the use of filtered pipette tips. RNA isolation was based on the procedure of Chomczynski and Sacchi (1987) utilizing the Trizol reagent. All isolated RNA samples were stored at -70 °C.

#### 2.2.2.1 RNA Isolation from Fresh Tissues

RNA was isolated from a number of tissues obtained from a 20 week-old male human fetus. Access to these samples was kindly granted by Dr. Roger Byard (Department of Histopathology, WCH). Selected tissues were frozen in liquid nitrogen and stored at -70 °C until RNA preparation was required. In each preparation, 1 ml of Trizol reagent was used for every 100 mg of frozen tissue. Approximately 100 to 500 mg of tissue was removed from the frozen stock and placed in a sterile mortar dish containing a small amount of liquid

nitrogen to keep the sample frozen. The tissue was ground to a fine powder with a sterile pestle. Liquid nitrogen was periodically added to the ground tissue to prevent thawing. The mortar and pestle had previously been washed thoroughly with ethanol, rinsed with DEPC-treated water, then heat treated for 4 hours at 160-180 °C to inactivate any residual RNAses. The ground tissue was then transferred to a 10 ml tube containing the appropriate amount of Trizol reagent. The powder was resuspended by gentle mixing and transferred to a sterile 5 ml <sup>glass?</sup> grass homogenizer (Lab Supply). The solution was homogenized until lumps of powder were no longer visible, then transferred to a 1.5 ml eppendorf tube, and left at room temperature for 5 minutes. One-tenth the volume of chloroform was added and the tube was shaken for 25 seconds and placed on ice for a further 5 minutes. The sample was then spun at 12,000 rpm at 4 °C for 15 minutes, and the supernatant was transferred to a new tube. An equal volume of isopropanol was added to this tube, and followed by incubation at 4 °C for 15 minutes. After centrifugation at 4 °C for 15 minutes, the RNA pellet was washed in 800 µl of 75% ethanol and allowed to air-dry for 15 minutes. The RNA pellet was dissolved in 50 µl of DEPC-treated water, and pooled into one tube if the original tissue homogenate was transferred into more than one eppendorf tube. A second precipitation step was performed involving the addition of one-twentieth the volume of 4 M NaCl and 2 volumes of 100%(absolute) ethanol to the dissolved RNA. This was followed by incubation at -20 °C for 1 hour then spun for 15 minutes at 4 °C. The RNA pellet was again washed with 800 µl of 75% ethanol, allowed air-drying for 20 minutes, and dissolved in DEPC-treated water. RNA concentration was measured by spectrophotometry (2.2.4). The integrity of the RNA preparation was tested by running a small aliquot on a 0.8% agarose gel electrophoresis.

#### **2.2.2.2 RNA Isolation from Cell Lines**

RNA isolation from cell lines followed the same procedure as that for RNA isolation

from tissue sources. Typically,  $1 \times 10^7$  cells obtained from cell culture were washed in PBS and spun down at 1,200 rpm for 10 minutes. The cells were then resuspended in 1 ml of Trizol reagent. This was subsequently followed by the procedures identical to those described in 2.2.2.1.

### 2.2.3 Oligonucleotide Primers

The oligonucleotide primers for both PCR and sequencing were generally designed to contain approximately 50% G-C content, and an annealing temperature of 60 °C, calculated by  $2x(A+T)+4x(G+C)$  as per Suggs *et al.* (1981). Ideally, primers were not to possess runs of identical bases and not to contain four contiguous base pairs of inter-strand or intra-strand complementary. In addition, primers were designed such that G-C, C-C, C-G, or G-G bases were present at their terminal 3' ends wherever possible. All oligonucleotides were purchased commercially from either Gibco-BRL or Bresatec (Adelaide, Australia).

### 2.2.4 Quantitation of DNA, RNA and Oligonucleotide Primers

Samples were diluted in water and their absorption of UV light (at wavelength of 260nm) was measured. The absorbance was multiplied by the dilution factor and a conversion factor (absorption co-efficient) for particular type of nucleic acid being quantitated. Relevant conversion factors were 50 for double stranded DNA, 40 for RNA, and 33 for oligonucleotides, which gave a concentration in  $\mu\text{g/ml}$ . The spectrophotometers, either Pharmacia Biotech Ultrospec 3000 or CECIL CE-2020 were used throughout this experiment

### 2.2.5 Preparation of Bacterial Glycerol Stocks

All bacterial clones were maintained as glycerol stocks. These were prepared from a 10

ml aliquot of overnight culture that had been used for DNA isolation. The 10 ml culture aliquot was spun down at 3,000 rpm. After the supernatant was removed, the pellet was resuspended in 1 ml of LB-Broth containing 15% glycerol and store at -70 °C until DNA preparation was required, whereby a 5 µl aliquot was used to streak plate for single colonies.

### **2.2.6 Preparation of Sheared Human Placenta and Salmon Sperm DNA**

Desiccated human placental DNA and salmon sperm DNA was purchased commercially and reconstituted in sterile TE to a concentration of 5 mg/ml and 20 mg/ml respectively. Aliquots of 1 ml were transferred to 1.5 ml eppendorf tubes and incubated at 100 °C for 6 to 20 hours. Periodically, 1-µl aliquots were taken and analysed on 1.5% agarose gel electrophoresis until the average sized fragments were below 700 bp. The concentration of the samples were then analysed by spectrophotometry and adjusted to their starting values (2.2.4)

### **2.2.7 Restriction Endonuclease Digestion**

Digestion of DNA with restriction endonuclease was performed using appropriate reaction buffer that was supplied and recommended by the manufacturer. Digestion was generally carried out in 10 µl volume, overnight at 37 °C unless otherwise specified. If recommended, BSA at a final concentration of 100 µg/ml was included in the digestion reaction.

For 10 µg of genomic DNA isolated from tissue, blood lymphocyte or cell line, digestion was performed in 50 µl reaction volume with 20 units of the restriction enzyme. To confirm the completion of digestion, 5 µl aliquot was run by electrophoresis on 0.8 % agarose gel (2.2.8).

Digestion of cosmid, BAC, or PAC DNA was undertaken in a 15 µl volume with 5

units of each restriction enzyme, for standard 250 ng of cosmid or 500 ng of BAC or PAC. The digestion reaction was scaled up when desired.

Plasmid DNA digest was generally desired for the isolation of cloned cDNA insert. The cDNA clones were either purchased from Genome Systems or obtained by screening cDNA library. For each 5 µg of plasmid, DNA was digested with 20 unites of the appropriate restriction enzyme in a 50 µl reaction volume.

### 2.2.8 Agarose Gel Electrophoresis

Electrophoresis of DNA materials was generally performed in agarose gel at gel percentage ranging from 0.8% to 2.5 % (w/v) in 1x TBE buffer. The analysis of plasmid, cosmid, BAC or PAC DNA digests were usually performed in 0.8% gels, and run at 100 volts. The PCR products were analysed in 1.5 to 2.5% agarose gels depending on the size of the expected products, and run at 110-120 volts. All electrophoresis was carried out in 1x TBE buffer. Each DNA sample was mixed with agarose gel loading buffer (to 1x final concentration) prior to loading. After electrophoresis, gels were stained in 0.02% ethidium bromide solution for 30 minutes, and DNA was visualised under UV light.

### 2.2.9 Southern Blotting

#### 2.2.9.1 Transfer in 10X SSC

This method was adapted from that originally described by Southern (1975) and was used for the transfer of restriction digested genomic DNA. After electrophoresis, the agarose gel was soaked in 500 ml of gel denaturing solution for 60 minutes, followed by 500 ml of gel neutralizing solution for a further 60 minutes. Prior to the completion of gel neutralization step, a piece of GeneScreen Plus nylon membrane was cut to size and soaked in water for 10 minutes followed by 15 minutes soaked in 10X SSC. The gel was then



placed face down on a blotting tray containing 10X SSC, and overlaid with the pre-treated membrane. After air bubbles were carefully removed, 2 pieces of 3MM Whatman paper soaked in 10x SSC were placed on the membrane, followed by three dry sheets. A wad of paper towels was then placed over this and allowed to transfer overnight. After overnight transfer, the position of gel wells was marked on the membrane, which was then soaked for 1 minute in filter denaturing solution followed by a 2 minutes soak in filter neutralizing solution. The membrane was then left to dry overnight or baked in the microwave oven for 5 minutes at half energy power.

#### **2.2.9.2 Transfer in 0.4 M NaOH**

This method was adapted from that described by Reed and Mann (1985) and was used for the transfer of digested plasmid, cosmid, BAC or PAC DNA. After the electrophoresis, the gel was stained in ethidium bromide, <sup>a</sup> photograph taken, and the DNA size marker positions on gel were stabbed with a needle and ink. GeneScreen Plus membrane cut to size was soaked in water for 5 minutes, followed by 0.4 N NaOH for another 5 minutes. The gel was blotted in 0.4 N NaOH overnight, without any pre-treatment, as described above except that the blotting tray contained 0.4 N NaOH. After overnight blotting, the position of the wells and DNA size markers was marked and the membrane was treated in filter neutralizing solution for 2 minutes. The membrane was then air-dried or dried in the microwave oven as above.

#### **2.2.10 Radio-Isotope Labelling of DNA**

Double-stranded DNA fragments were labelled with [ $\alpha$ -<sup>32</sup>P]dCTP using a modified procedure to that described by Feinberg and Vogelstein (1983).

##### **2.2.10.1 Labelling DNA Fragments in Solution**

In a standard labelling reaction, 50 ng of double stranded DNA was made up to 34  $\mu$ l with sterile water, mixed, and incubated at 100 °C for 5 minutes. After spinning down the contents, 5  $\mu$ l of 10X dNTP labelling solution, 5  $\mu$ l of 10X labelling buffer, 5  $\mu$ l of [ $\alpha$ - $^{32}$ P]dCTP (50  $\mu$ Ci), and 5 units of *E. coli* DNA polymerase I were added. This was incubated at 37 °C for 30 minutes. The reaction was stopped with the addition of 1  $\mu$ l of 0.5 M EDTA, and if required, un-incorporated [ $^{32}$ P] dCTP was removed using QIAquick columns (2.2.14.1). DNA probes that required blocking of repetitive sequences were then pre-reassociated (2.2.10.3). Probes not requiring pre-blocking were heat denatured at 100 °C for 5 minutes before adding to pre-hybridized filters (2.2.11.1).

#### 2.2.10.2 Labelling DNA Fragments in Agarose

The DNA fragment contained in agarose was heated at 100 °C for 2 minutes. An aliquot of approximately 50 ng was then removed and transferred to a tube with sterile water such that the volume was 34  $\mu$ l. This solution was heated at 100 °C for 5 minutes and the same steps as in 2.2.10.1 were subsequently followed.

50 ng of  
DNA?  
? how  
amount

#### 2.2.10.3 Pre-reassociation of Labelled Probes

Probes suspected to contain human repetitive elements or probes to be used for Northern hybridization had their repetitive elements blocked, before hybridization with the membrane, with sheared denatured human placental DNA (2.2.6). The procedure used was based on that of Sealy *et al.* (1985). To the labelled probe (50  $\mu$ l reaction), 100  $\mu$ l of human placental DNA (5 mg/ml), and 50  $\mu$ l of 20X SSC were added. This was incubated at 100 °C for 10 minutes, on ice for 1 minute, followed by an incubation at 65 °C for at least 1 hour. The probe could then be added directly to the pre-hybridized membranes (2.2.11.1).

### 2.2.11 Hybridization of Nucleic Acid Bound Membranes

All hybridizations, including Northern Blot and RNA Master Blot hybridizations were performed in a HYBRID orbital midi oven.

#### 2.2.11.1. Southern Hybridizations

Southern filters were hybridized based on a modification of Brown (1993). Membranes were placed in glass bottles and pre-wet in 5X SSC for 1 minute. This solution was replaced with either 10 ml of Southern hybridization solution (for small bottles) or 15 ml (for large bottles) and filters were pre-hybridized at 42 °C for at least 2 hours. Once the DNA probe had been labelled and either denatured (2.2.10.1) or pre-reassociated (2.2.10.3), it was added directly to the pre-hybridized filters.

not in ref list

#### 2.2.11.2 Northern Hybridization

Northern membrane, commercially purchased was hybridized according to protocols supplied with the Clontech multiple tissue Northern blots (User manual, PT 1200-1). Membranes were pre-soaked in 5X SSC for 1 minute, then pre-hybridized at 65 °C for at least 2 hours in 10 ml of ExpressHyb solution containing denatured salmon sperm DNA (100 µg/ml). After the DNA probe had been labelled (2.2.10), cleaned (2.2.14.1), and pre-reassociated (2.2.10.3), the pre-hybridization solution was removed from the membrane, and replaced with a fresh 10 ml of ExpressHyb solution containing the probe and salmon sperm DNA (100 µg/ml). Hybridization proceeded overnight at 65 °C. Commercial Northern blots were hybridized with the control  $\beta$ -Actin probe supplied with the membrane to test for the integrity and loading of the RNA samples.

#### 2.2.11.3 Washing of Membranes

Southern membranes were washed based on procedures presented in Sambrook *et al.* (1989). Filters were treated sequentially at 42 °C for 10 minutes in solutions containing 2X SSC and 1% SDS. If high background was still evident, the filters were washed for another 20 minutes at 42 °C in 2X SSC and 1% SDS (heated to 65 °C first), followed by another 20 minutes in 0.1X SSC and 1% SDS at 65 °C if needed.

Northern membranes were washed according to manufacturer specifications (User manual, PT1200-1). Filters were first rinsed in 200 ml of a solution containing 2X SSC and 0.05% SDS at room temperature for 1 minute at a time. Counts were monitored after each rinse. Following this, four 10 minute washes at room temperature in the same solution were performed with constant monitoring after each wash. If the counts were still high (>50 cpm), the membranes were washed in a solution containing 0.1X SSC and 0.1% SDS for 5 minutes at a time at 65 °C. All washed membranes were exposed for the appropriate time to X-OmatK XK-1 Kodak diagnostic film either at room temperature or -70 °C.

#### **2.2.11.4 Stripping of Membranes for Re-Use**

Southern membranes were stripped of radio-labelled probe by incubating the washed filters at 45 °C for 30 minutes in a solution containing 0.4 M NaOH. The filters were then transferred to a solution containing 0.1% SDS; 0.1X SSC; 0.2 M Tris-HCl (pH 7.5), and incubated a further 15 minutes. Northern membranes were stripped of radio-labelled probe by the addition of a solution of 0.5% (w/v) SDS that had been heated to 100 °C, followed by an incubation at room temperature until cool.

#### **2.2.12 Polymerase Chain Reaction (PCR)**

All PCR reactions were performed using *Taq* DNA polymerase in one of two buffer systems; 2X PCR mix (modification of Kogan *et al.* 1987) or 10X PCR buffer (supplied by

Gibco-BRL). In each case 1.5 mM MgCl<sub>2</sub> was used unless high background or no PCR products were obtained with a control template. If this occurred, the PCR conditions were optimised by varying the concentration of MgCl<sub>2</sub> used in the reaction. All PCR reactions were set up on ice, and were performed in 0.5 ml PCR tubes using a thermal cycler 480 (Perkin Elmer Cetus) or a PCR express thermal cycler (Hybaid). All colony PCRs were performed in 0.2 ml PCR tubes using an FTS-960 thermal sequencer (Corbett Research).

### **2.2.12.1 Standard PCR Reactions**

Routine PCR reactions were performed in 10 µl volumes using 1 µl of 10X PCR buffer, 0.2 µl of 10 mM dNTPs, 0.3 µl of 50 mM MgCl<sub>2</sub>, 75 ng of each primer, 0.25 units of *Taq* DNA polymerase, and template DNA. Template DNA amounts were 100 ng for genomic and cell line DNA, 10 ng for BAC and PAC DNA, and 1 ng for plasmid DNA. All reactions were overlaid with a drop of paraffin oil and incubated at 94 °C for 2 minutes, followed by 35 cycles of 94 °C for 30 seconds; 60 °C for 1 minute; 72 °C for 2 minutes, followed by a final elongation step of 72 °C for 7 minutes. Each reaction was then run on an agarose gel of the appropriate percentage (2.2.8). When the cloning of PCR products was anticipated, the original reaction was scaled up to 20 µl. Half was then examined on an agarose gel while the remainder was kept for cloning into the pGEM-T vector (2.2.15.2).

### **2.2.12.2 Colony PCR Reactions**

A single colony was picked with a sterile disposable pipette tip, streaked onto a grid position on a fresh L-Agar plate containing the appropriate antibiotic (master plate), and the tip placed into a PCR tube containing a drop of paraffin oil. The tubes were left for 10 minutes at room temperature during which the PCR reaction mix was prepared. This mix consisted of 5 µl of 2X PCR mix, 0.3 µl of 50 mM MgCl<sub>2</sub>, 0.2 µl of each primer (30 ng of

each), 3.6  $\mu$ l of sterile water, and 0.2  $\mu$ l of *Taq* DNA polymerase (0.25 units). The pipette tips were removed from the PCR tubes and the tubes were incubated at 99 °C for 10 minutes. The samples were then held at 80 °C, and 10  $\mu$ l of the prepared PCR reaction mix was added to each tube below the level of the oil. The tubes were incubated at 95 °C for 2 minutes followed by 35 cycles of 95 °C for 15 seconds; 60 °C for 30 seconds; 72 °C for 2 minutes, and a final extension step of 7 minutes at 72 °C. In general, 5  $\mu$ l of each reaction was examined on a 2.5% agarose gel.

### **2.2.13 Reverse Transcription PCR (RT-PCR)**

#### **2.2.13.1 Reverse Transcription of cDNA**

The procedures were adapted from those supplied with the Superscript enzyme. All reactions were carried out on ice at all times. Three micrograms of total RNA or 100 ng of polyA<sup>+</sup> mRNA was added to either 50 pmoles of random hexamers (Perkin Elmer) or 100 pmoles of oligo dT<sub>12-18</sub>. The volume was made up to 11.5  $\mu$ l with the addition of DEPC-treated water, mixed, and incubated at 65 °C for 5 minutes. Following a 1 minute incubation on ice, the contents of the tube were spun down briefly, and 4  $\mu$ l of 5X 1<sup>st</sup> strand buffer, 2  $\mu$ l of 0.1 M DTT, 1  $\mu$ l of 10 mM dNTPs, and 0.5  $\mu$ l (20 units) of RNAsin were added. This was incubated at 42 °C for 2 minutes, followed by the addition of 1  $\mu$ l (200 units) of Superscript reverse transcriptase, and a further incubation at 42 °C for 30 minutes. The reaction was terminated by incubation at 70 °C for 10 minutes and samples were kept at -20 °C until needed. A control reaction was included for each individual experiment where the reverse transcriptase enzyme was omitted. This was to test for genomic contamination present within the RNA template that may be seen following the PCR step (2.2.13.2). All polyA<sup>+</sup> mRNA samples were kindly provided by Dr. Jozef Gecz (Department of Cytogenetics and Molecular Genetics, WCH).

### 2.2.13.2 PCR Amplification of cDNA

Two microlitres of the reverse transcription reaction was combined with 2  $\mu$ l of 10X PCR buffer, 0.4  $\mu$ l of 10 mM dNTPs, 0.6  $\mu$ l of 50 mM MgCl<sub>2</sub>, 1  $\mu$ l of each primer (150 ng of each), and the volume made up to 19  $\mu$ l with sterile water. After the addition of 1  $\mu$ l *Taq* DNA polymerase (0.5 units) to all tubes, including those tubes without reverse transcriptase, one drop of paraffin oil was added, and the tubes were incubated at 94 °C for 2 minutes. The samples were then incubated at 94 °C for 30 seconds; 60 °C for 1 minute; 72 °C for 2 minutes for 35 cycles, followed by a final elongation step at 72 °C for 7 minutes. Ten microlitre aliquots of all RT-PCR products were analysed on 2.5% (w/v) agarose gels while the remaining aliquot for many reactions was kept for subcloning into the pGEM-T vector (2.2.15.2) or for direct purification using QIAquick columns (2.2.14.1) for sequencing purposes. A positive control PCR to test the success of the cDNA synthesis was done for each reverse transcription reaction with primers to the esterase D (*ESTD*) housekeeping gene (GenBank Accession number M13450). These primers give rise to a 452 bp amplicon from cDNA template only. Primer sequences of *ESTD* are *forward*, 5' GGA GCT TCC CCA ACT CAT AAA TGC C 3' (nt 423-447) and *reverse*, 5' GCA TGA TGT CTG ATG TGG TCA GTA A 3' (nt 875-851).

### 2.2.14 Purification of DNA Fragments

#### 2.2.14.1 Purification of PCR Products and Radio-Labelled Probes

PCR products were cleaned according to manufacturer protocols using QIAquick columns (Qiagen) and buffered <sup>s</sup> which were supplied with the kit. Briefly, the remaining PCR reaction was removed from the paraffin oil and transferred to a clean tube. Each sample was mixed with a 5X volume of buffer PB. This mixture was added to a purification

column, and spun for 1 minute at 13,000 rpm. The collection tube was drained and 750  $\mu$ l of buffer PE was then added, followed by a 1 minute spin at 13,000 rpm. The collection tube was again drained, and the column spun again. The DNA was eluted by the addition of 50  $\mu$ l of sterile water to the column, followed by centrifugation at 13,000 rpm for 1 minute with a clean collection tube. The purified DNA was quantitated by spectrophotometry (2.2.4).

Radio-labelled probes were purified from un-incorporated nucleotides using the same procedure as described above. However, only 700  $\mu$ l of buffer PE was used to reduce the chance of contaminating the centrifuge from overflowing of the spin column.

#### **2.2.14.2 Purification of DNA Fragments from Agarose Gel**

Restriction fragments excised from agarose gels were purified from the agarose using the Prep-A-Gene purification system (Bio-Rad) according to manufacturer conditions. Where possible, fragments to be isolated were run in 0.8% (w/v) gels, and trimmed of excess agarose once cut from the gel. Five microlitres of Prep-Agene matrix was used per microgram of DNA excised. Three gel slice volumes of binding buffer were added to the excised band, followed by a 10 minute incubation at 50 °C to dissolve the agarose. After addition of the required amount of matrix, the tube was vortexed briefly, and incubated at room temperature for 10 minutes on a rotating wheel. The tube was then spun for 30 seconds, and the supernatant removed. The pellet was rinsed and resuspended in binding buffer equivalent to 25X the amount of added matrix, and spun a further 30 seconds. The pellet was then washed twice with a 25X matrix volume of wash buffer. The pellet was then air-dried for 10 minutes at room temperature, resuspended in at least 1 pellet volume of sterile water, and incubated at 42 °C for 5 minutes. The tube was then spun for 30 seconds and the supernatant transferred to a fresh tube. The above step was repeated and the supernatants pooled. The concentration of the DNA eluted was then determined by

*with clear  
wash  
just good*



spectrophotometry (2.2.4).

## 2.2.15 Cloning of DNA Fragments

### 2.2.15.1 DNA Ligation into Plasmid Vector

<sup>Each</sup> All ligations <sup>c. 10-3</sup> were performed in a volume of 10  $\mu$ l at 15 °C for 3 to 16 hours. Typical ligations involved 1:1 molar ratios of the vector and the DNA fragment to be cloned based on methods described in Sambrook *et al.* (1989). In most reactions 50 ng of vector DNA was used. Ligations included 5 units of T4 DNA ligase (MBI) using the buffer supplied with the enzyme.

### 2.2.15.2 Ligation of PCR Products into the pGEM-T Vector

The pGEM-T vector system (Promega) was employed to clone DNA fragments amplified by PCR using *Taq* DNA polymerase. This system takes advantage of the addition of a single adenosine residue by thermostable DNA polymerase to the 3' end of all duplex molecules (except for the PCR products generated by *Pfu* DNA polymerase). The A overhangs are base-paired with the single thymidine overhangs present in the pGEM-T vector, facilitating the ligation. In most cases, 1  $\mu$ l of the completed PCR reaction was added to 1  $\mu$ l of the pGEM-T vector, 1  $\mu$ l of 10X ligation buffer, and 1  $\mu$ l of T4 DNA ligase in a 10  $\mu$ l volume. Ligations were performed overnight at 15 °C.

### 2.2.15.3 Preparation of Competent Bacterial Cells

Bacterial cells were made competent based on a modification of Chung *et al.* (1989). XL-1 Blue cells were streaked for single colonies on an LB-Tetracycline plate from a 5  $\mu$ l aliquot of a glycerol stock. A single colony was grown overnight in 10 ml of LB-Broth plus tetracycline (15  $\mu$ g/ml) at 37 °C. The next day, 50 ml of LB-Broth plus tetracycline (15

$\mu\text{g/ml}$ ) was seeded with 1 ml of the overnight culture and grown until an  $A_{600}$  of 0.3 was reached (approximately 1.5 to 2 hours). The cells were then transferred to a 50 ml tube and spun at 2,000 rpm for 10 minutes. The supernatant was removed and the cells were carefully resuspended in 5 ml of ice-cold TSS media. The cells were left on ice for 10 minutes before being transformed (2.2.15.4). Any left over cells were transferred to ice-cold 1.5 ml eppendorf tubes, as 500  $\mu\text{l}$  aliquot, and frozen in liquid nitrogen. Frozen cells were then stored at  $-70\text{ }^{\circ}\text{C}$  until required.

#### 2.2.15.4 Transformation of Competent Bacterial Cells

The procedure utilized was modified from the technique described in Chung *et al.* (1989). For each reaction, 100  $\mu\text{l}$  of competent XL-1 Blue cells were thawed on ice and 5  $\mu\text{l}$  of the ligation reaction was added. Following gentle mixing, the cells were placed at  $4\text{ }^{\circ}\text{C}$  for 30 minutes. During this step, 880  $\mu\text{l}$  of LB-Broth was placed in a 10 ml tube and glucose was added to a concentration of 20 mM. The cells and DNA mix were then added, and the tube was incubated at  $37\text{ }^{\circ}\text{C}$  for 1 hour. For vectors with blue/white colour selection, an aliquot of 200  $\mu\text{l}$  of transformed cells, 40  $\mu\text{l}$  of X-Gal (20 mg/ml) and 20  $\mu\text{l}$  of 200 mM IPTG were then spread on an LB-Ampicillin plate, and incubated overnight at  $37\text{ }^{\circ}\text{C}$ . Otherwise, 200  $\mu\text{l}$  of transformed cells alone was spread onto an LB-Ampicillin plate.

Handwritten notes:  
The above  
primary  
transformant  
colony  
XL-1 Blue

#### 2.2.15.5 Preparation of Colony Master Plates and Colony Lifting

Colonies representing potential recombinant molecules were picked and streaked onto a gridded master plate. Following an overnight incubation at  $37\text{ }^{\circ}\text{C}$ , inserts of gridded clones were subsequently amplified by colony PCR (2.2.12.2) or transferred to nylon membranes for screening purposes. The procedure to transfer bacterial clones to membranes was based on a modified version to that described by Grunstein and Hogness (1975). Master plates

containing colonies for transfer were first placed at 4 °C until chilled. Agar plates containing individually gridded bacterial colonies were overlaid with a membrane and markers used to orient the filters were stabbed into the edge of the membrane using a needle and ink. This was left for 5 minutes. The membrane was removed and placed colony side up on a sheet of Whatman 3MM paper soaked in colony denaturing solution and left for 7 minutes. The membrane was then transferred to a sheet of Whatman 3MM paper soaked in colony neutralisation solution and left for 5 minutes. This step was repeated on a separate sheet of Whatman paper, and the membranes were then soaked in 500 ml of 2X SSC for 5 minutes, air-dried for 10 minutes, and baked at 65 °C for 1 hour in an oven. In some instances duplicate colony lifts were done in which a second membrane was placed onto the agar plate once the first one had been removed. All steps subsequent to this were as stated above.

#### **2.2.16 DNA Sequencing**

The sequencing of both double stranded plasmid DNA and PCR product DNA was achieved with cycle sequencing using reagents supplied with the ABI Prism Dye Terminator or Dye Primer (-21M13 forward and M13RP1) Cycle Sequencing kit. Electrophoresis was performed on a DNA sequencer, Applied Biosystems Model 373A, initially by Dr. Josef Gecz or Dr. Julie Nancarrow (Department of Cytogenetics and Molecular genetics, WCH), or later on the same machine when it was relocated to the Australian Genome Research Facility in Brisbane. Custom electrophoresis for DNA sequencing was also provided by molecular pathology sequencing service, Institute of Medical and Veterinary Science (IMVS), Frome Road, Adelaide.

##### **2.2.16.1 Dye Terminator Cycle Sequencing**

For each sequencing reaction, either 180 ng of purified PCR product of DNA or 500 ng

of plasmid DNA was mixed with 20 ng of the desired primer, 4  $\mu$ l of the Dye Terminator mix, 4  $\mu$ l of halfTERM Dye Terminator Sequencing Reagent, and made up to a volume of 20  $\mu$ l with sterile water. After <sup>being</sup> overlaid with a drop of paraffin oil, the tubes were incubated for 25 cycles of 96 °C for 30 seconds; 50 °C for 15 seconds; 60 °C for 4 minutes. The sample was then purified using procedures provided with the kit (Revision A, 1995). This involved removal of the sample from the oil and transfer to an 1.5 ml eppendorf tube containing 2  $\mu$ l of 3 M NaOAc (pH 5.2) and 50  $\mu$ l of 95% ethanol. Following a 10 minute incubation on ice, the tubes were spun at 13,000 rpm for 30 minutes, and the supernatant was discarded. The pellet was then washed with 250  $\mu$ l of 70% ethanol, spun for a further 5 minutes and the ethanol removed. The pellet was allowed to dry for 10 minutes at room temperature, and then kept at 4 °C in the dark until run on a sequencing gel.

### 2.2.16.2 Dye Primer Cycle Sequencing

For the sequencing of each template, four separate sequencing reactions were needed. To the first two tubes, one containing 4  $\mu$ l of dd-ATP mix and the other containing 4  $\mu$ l of dd-CTP mix, 1  $\mu$ l (200 ng) of DNA template was added. To the remaining two tubes, one containing 8  $\mu$ l of dd-GTP mix and the other containing 8  $\mu$ l of dd-TTP mix, 2  $\mu$ l (400 ng) of DNA template was added. After the addition of a drop of paraffin oil, the tubes were incubated for 15 cycles of 95 °C for 30 seconds; 55 °C for 30 seconds; 70 °C for 1 minute, followed by a further 15 cycles of 95 °C for 30 seconds; 70 °C for 1 minute. Samples were then purified using procedures provided with the kit (Revision B, 1995). All four samples from each template were removed from the oil and combined in a tube containing 80  $\mu$ l of 95% ethanol. Following a 10 minute incubation on ice, the tubes were spun at 13,000 rpm for 30 minutes then the supernatant was removed. The pellet was washed with 250  $\mu$ l of 70% ethanol, spun for 5 minutes, and air-dried for 10 minutes after removal of the

supernatant. Samples were subsequently kept at 4 °C in the dark until run on a sequencing gel.

### **2.2.17 Rapid Amplification of cDNA Ends (RACE)**

#### **2.2.17.1 5'RACE**

A procedure for the identification of the 5' ends of transcripts was adapted from the "SMART" PCR method (Clontech, USA) that has been developed for the selective amplification of full-length cDNAs. According to the manufacturer, this method is based on the use of SMART oligonucleotide to capture the 5' end of the mRNA and serves as a short, extended template at the 5' end for the reverse transcription reaction. This allows synthesis of the first strand cDNA that contains the complete 5' end of the mRNA as well as the sequence complementary to the SMART oligonucleotide, which then serves as a universal PCR priming site (SMART anchor) in the subsequent amplification.

Most of the components were supplied with the kit (PT3000-1, Clontech) except the gene specific primers (GSP) which are listed in the relevant chapters that used this technique.

In the initial step, synthesis of the first strand cDNA, 1 µg of total RNA or 100 ng of polyA<sup>+</sup> mRNA was incubated with 50 pmoles of random hexamers (Perkin Elmer) and 1 µM of SMART oligonucleotide (5' TACGGCTGCGAGAAGACGACAGAAGGG 3', Clontech). This reaction was carried out in the first strand reaction buffer containing 2 mM DTT, 1 mM dNTP, and 200 units of MMLV reverse transcriptase (Superscript, Gibco-BRL). The reaction was allowed at 42 °C for 1 hr and was stopped by placing the tube on ice. The cDNA amplification step was carried out using 5' PCR anchoring primer and the gene specific primer (GSP). Two microliters of the first strand cDNA was added to 80 µl of RNase-free deionized water, 10 µl of 10x Klen Taq PCR buffer, 2 µl of dNTP mix, 2 µl of

5' PCR primer, 2  $\mu$ l of gene specific primer (GSP1), and 2  $\mu$ l 50x Advantage KlenTaq Polymerase Mix. After <sup>mixing</sup> <sup>centrifugation</sup> mixed and spun to collect the contents at the bottom of the tube, the reaction mixture was overlaid with 2 drops of mineral oil, <sup>the</sup> lid closed, and the tube <sup>was</sup> <sup>then</sup> placed in a 95 °C preheated thermal cycler. PCR was carried out by incubation at 95 °C for 1 min and followed by 20 cycle of 95 °C, 15 sec; 68 °C, 15 sec; 60 °C, 30 sec; and 72 °C for 5 min. This was followed by two rounds of the nested PCR. The first PCR was undertaken using a GSP2 and 5' PCR primers, and the second was GSP3 and 5' PCR primer in the standard PCR reaction buffer and *Taq* DNA polymerase (2.2.13). Ten microlitres of the PCR products were analysed on 1.5% (w/v) agarose gels, and the remaining sample was cloned into the pGEM-T vector (2.2.15.2) and subsequently sequenced (2.2.16) using either vector primers or gene specific primers.

#### 2.2.17.2 3' RACE

The 3'RACE procedure adapted from Gecz et al. (1997) was used to determine the 3' sequence of the transcript. First strand cDNA was reverse transcribed from poly A+ mRNA utilizing an anchored oligo-dT primer, 5'-CCATCCTAATACGACTCACTATAGGGCT-CGAGCGGC(T)18VN-3' (V is any of the four nucleotide, A, C, G, T, and N is C, G, or A ). Second strand cDNA was synthesized by a linear PCR using a biotin labeled (biotinylated) gene specific primer. Gene specific biotinylated cDNAs were then captured on streptavidin-coated magnetic beads (DYNAL). The 3' end was subsequently PCR amplified using the nested gene specific primers and the primers specific for anchoring sequence of the oligo-dT primer. The amplified fragment was ligated into pGEM-T vector (2.2.15.2) for subsequent cloning and sequencing.

#### 2.2.18 Screening of cDNA library

### **2.2.18.1 PCR screening of Formatted Fetal Brain cDNA Library**

A phage library of random-primed and oligo-dT-primed (5'-Stretch Plus) cDNA from human fetal brain (Clontech) was previously amplified and formatted for subsequent rapid screening (Whitmore, personal communication). The formatting procedure involved plating a 1.5 fold representation of all clones in the library on one hundred 15 cm LB-Agar plate (15,000 phage per plate). The amplified clones from each plate were then transferred to a 10 ml tube, a total of 100 tubes. An aliquot was taken from each tube to produce a combination of 40 pools. These pools, representing the library, were subsequently used for templates in PCR screening. The phage vector,  $\lambda$ -gt11, contains the priming sites for gt11 *forward* and *reverse* primers flanking the cloning region. This allows amplification of the insert.

When using a combination of vector and gene specific primers PCR screening of 40 pools allows amplification of the gene specific 5' cDNA fragment for further analysis. PCR screening of these pools was also used to identify a specific fraction of the library from the original 100 tubes (fractions), for subsequent screening by hybridization (2.2.18.2).

### **2.2.18.2 Screening cDNA Library by Hybridization**

Screening of the library by hybridisation was performed for the isolation of specific cDNA clone. The competent host for phage entry, Y1090 E.coli cells, were first grown in 50 ml of LB-Broth, 100  $\mu$ l of 1 M MgSO<sub>4</sub>, and 100  $\mu$ l of 20% maltose, overnight at 37 °C. The cells were precipitated by spinning at 3000 rpm for 5 minutes. The pellet was resuspended in 10 ml of 10mM MgSO<sub>4</sub> and kept on ice. This stock was use to inoculate another 10 ml of 10 mM MgSO<sub>4</sub> to the cell density of approximately OD<sub>600</sub> = 1 and placed on ice for use as host cell. During this period, 15 cm LB-Agar plates were pre-warmed at 37 °C for subsequent phage plating.

A selected fraction of phage library, as screened by PCR, was diluted into 2 to 3 serial

dilutions in SM buffer. Each diluted library was then added to 500  $\mu$ l of the host cells that was resuspended in 10mM MgSO<sub>4</sub>. The tubes were incubated at 37 °C for 30 minutes. During this period, 8 ml of melted Top Agar was aliquoted into 10 ml tubes and left at 50 °C. For each of diluted phage, the mixture of cells and phage were then added to the Top Agar, mixed by inversion, then poured onto a pre-warmed 15 cm LB-Agar plate. After the Top Agar had set, the plates were incubated overnight at 37 °C. After overnight incubation, phage plaques were transferred to nylon membranes as “phage library filters” for subsequent phage-hybridization screening. Phage library filters were prepared in duplicate from each plate utilizing the same procedure used for colony lifting, described in section 2.2.15.5. Filter hybridization and washing was carried out as described in section 2.2.11.1.

### 2.2.19 Bioinformatics

The URLs of tools and database, which were used for the analysis of DNA and protein sequences in this thesis are given in *Table 2.1*.



***Table 2.1***

**URL of tools and database used in analysis of nucleotide and protein sequences**

<b><i>Tools /Database</i></b>	<b><i>URL and Description of program usage</i></b>	<b><i>Reference</i></b>
BLAST:	<a href="http://www.ncbi.nlm.nih.gov/BLAST/">http://www.ncbi.nlm.nih.gov/BLAST/</a> Programs for rapid sequence similarity searches for protein and DNA sequences. Also provides searches for nucleotide sequences translated in all six frames.	Altschul <i>et al.</i> , 1997
PSI-BLAST:	<a href="http://www.ncbi.nlm.nih.gov/blast/psiblast.cgi">http://www.ncbi.nlm.nih.gov/blast/psiblast.cgi</a> An implementation of BLAST capable of detecting distantly related proteins by creating a PSSM from the significant matches detected in one round and using that PSSM as the query in the next iteration.	Altschul <i>et al.</i> , 1997
CDD:	<a href="http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml">http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml</a> Conserved Domain Database: Collection of protein domains with links to structures if available. CDD uses a library of PSSMs that is searched by Reverse Position-Specific BLAST to match a protein query.	Wheeler <i>et al.</i> , 2001
Pfam:	<a href="http://www.sanger.ac.uk/Pfam/">http://www.sanger.ac.uk/Pfam/</a> Collection of protein domains and families consisting of multiple alignments and hidden Markov models generated in a semi-automatic manner.	Bateman <i>et al.</i> , 2000
SMART:	<a href="http://smart.embl-heidelberg.de/">http://smart.embl-heidelberg.de/</a> Searchable collection of protein domains for detecting and analysing domain architecture.	Schultz <i>et al.</i> , 2000
PRINTS:	<a href="http://bmbsgi11.leeds.ac.uk/bmb5dp/prints.html">http://bmbsgi11.leeds.ac.uk/bmb5dp/prints.html</a> Database of protein fingerprints (group of motifs distributed in the sequence characteristic of a family).	Attwood <i>et al.</i> , 2000
PROSITE:	<a href="http://www.expasy.ch/prosite/">http://www.expasy.ch/prosite/</a> Database of biologically significant sites, patterns and profiles in proteins that helps in function prediction.	Hofmann <i>et al.</i> , 1999
PSORT	<a href="http://psort.ims.u-tokyo.ac.jp/">http://psort.ims.u-tokyo.ac.jp/</a> A program for detecting sorting signals in proteins and predicting their subcellular localization	Nakai and Horton, 1999
CLUSTALW	<a href="http://clustalw.genome.ad.jp/">http://clustalw.genome.ad.jp/</a> For multiple sequence alignment	Higgins <i>et al.</i> , 1996

## **Chapter 3**

# Characterization of Transcription Unit

3 (T3)

---

## Table of Contents

	Page
<b>3.1 Introduction</b>	86
<b>3.2 Methods</b>	88
3.2.1 Identification of the Transcript sequence	89
3.2.1.1 Database searching	89
3.2.1.2 cDNA library screening	89
3.2.1.3 Identification of the transcript using genomic sequence	89
3.2.1.4 PCR and RT-PCR	90
3.2.2 Expression Analysis	90
3.2.3 <i>In silico</i> Analysis of Protein function	91
3.2.4 Whole-body <i>in situ</i> hybridization of the mouse embryos	91
3.2.4.1 Prehybridization of embryos	91
3.2.4.2 Preparation of RNA probes and hybridization of embryos	92
3.2.4.3 Post-hybridization washes and embryo blocking	93
3.2.4.4 Antibody binding	94
3.2.4.5 Post-antibody washes and histochemistry	94
3.2.5 Study for T3 mutations in Breast Cancer	95
<b>3.3 Results</b>	96
3.3.1 Determination of Transcription Unit 3 (T3) Transcribed Sequence	96
3.3.1.1 Database analysis and cDNA sequencing	96
3.3.1.2 Transcript Sequence Analysis	97
3.3.2 Expression analysis of T3	98
3.3.3 Gene Structure of T3-S	98
3.3.4 Determination of T3-L cDNA Sequence	99
3.3.4.1 Database Analysis	99
3.3.4.2 Evidence of a Mouse transcript homologous to T3-L	100
3.3.4.3 T3-L Sequence Assembly	101
3.3.5 Transcript Sequence Analysis of T3-L and mouse orthologue	103
3.3.6 Protein Sequence and Function Analysis	104
3.3.7 Genomic Organization and Alternative Splicing	105
3.3.8 Expression Study of T3	108
3.3.8.1 Differential Expression	108
3.3.8.2 Whole-body <i>in situ</i> hybridization of mouse embryo	108

3.3.9 Mutation Analysis of <i>T3</i> in Breast Cancer	109
<b>3.4 Discussion</b>	<b>110</b>

### 3.1 Introduction

The refined LOH studies of 16q24.3 in breast cancer have revealed that the likely minimal LOH region is defined by the markers D16S303 and D16S2306. In addition, detailed physical and transcription maps were also established in this region (Whitmore *et al.*, 1998a), to form a framework for the identification of candidate tumor suppressor gene(s). Two known genes, *FAA* and *PISSLRE*, mapped to this region (FAB consortium, 1996; Whitmore *et al.*, 1998a), and based on their functions were previously suggested to be possible candidate genes. However, none of these genes showed mutations in breast cancer samples showing either restricted 16q24.3 LOH or complex LOH of 16q (Cleton-Jansen *et al.*, 1999; Crawford *et al.*, 1999). Other recently identified genes in this region, *GAS11*, *C16orf1* (Whitmore *et al.*, 1998b), and *Copine VII* (Savino *et al.*, 1999) also failed to demonstrate gene mutations in breast cancer. Additional genes and transcripts were, therefore identified in the region for consideration as a candidate gene involved in breast cancer.

In the physical region between the *MC1R* and *GAS11* genes (**Figure 1.3**), a cluster of two trapped exons, ET6 and ET7, isolated from the overlapping cosmids 31<sup>8</sup>/C5 and 301F3 (Whitmore *et al.*, 1998a) were selected for gene identification. These trapped products were less than 10 kb apart. They were grouped and assigned, based on their order in the physical map, as the transcription unit 3 (**T3**). Based on database analysis, ET6 was shown to have a sequence homology to a database entry MMDEF85, a partial 5' sequence of the mouse def-8 gene. This is a developmentally regulated gene identified from cells of the haemopoietic system.

X  
(on error  
with 1.3)

Def-8 is a novel mouse gene, which was isolated by utilizing the retroviral gene-trap vector carrying a  $\beta$ -galactosidase-neomycin fusion gene to infect myeloid progenitor cell lines (Hotfilder *et al.*, 1999). This gene-trap vector was constructed such that a splice

acceptor signal derived from *c-fos* was inserted at the 5' end of the *lacZ/neo* fusion gene, which itself lacked an endogenous ATG start codon. Upon integration into actively transcribed loci, the infected cells then gave rise to the clones that express a fusion of the endogenous, trapped gene with the *lacZ/neo* reporter. The gene-trap integration (GTI) clones were subsequently isolated by G418 selection and a  $\beta$ -galactosidase positive phenotype. To study the expression of developmentally regulated genes during myeloid differentiation, GTI clones were induced in order to allow differentiation *in vitro* into either macrophages or granulocytes. Gene expression was monitored by histochemical determination of  $\beta$ -galactosidase activity. From this expression analysis, 8 GTI clones each representing different trapped genes were isolated and were designated *def* (differentially expressed in FDCP-Mix) gene, as *def-1* to *def-8*. Expression of all *def* genes in undifferentiated progenitor cells has been shown to be down-regulated upon differentiation into either macrophages or granulocytes (Hotfilder *et al.*, 1999). Whereas *def-1* to *def-4* expression has been maintained upon macrophage differentiation, their expression is down-regulated during granulocyte differentiation. Expression of *def-5* to *def-8* appears to be down-regulated upon differentiation into both macrophages and granulocytes.

Down-regulation of both endogenous *def-6* and *def-8*, representing the second group, has also been observed during cell differentiation by Northern hybridization, using isolated trapped gene products as probes (Hotfilder *et al.*, 1999). Whereas down-regulation of *def-8* expression is restricted to macrophages and granulocytes, *def-6* expression also appears to be down-regulated upon erythrocyte differentiation. Further, Northern analysis of both *def-6* and *def-8* has also been undertaken during mouse development and in adult tissue. Expression of both genes is observed in mouse embryonic stages from day 7 to day 17. Among adult tissues examined, expression of *def-6* appears to be prominent in spleen, which is a haemopoietic organ, and minor in lung, with a very low signal in heart, muscle

and kidney. In contrast, def-8 appears to be expressed in all adult tissues.

The expression profiles of def-8 in non-haemopoietic tissue suggest the importance of this gene during embryonic development as well as a requirement for fully differentiated tissue. Despite the determination of gene expression, the gene sequence of def-8 is only available in databases as a partial 5' sequence of mRNA. The homology of the trapped exon, ET6 to the def-8 5' sequence indicates that there is likely to be a human homologue located at the 16q24.3 LOH region. Identification of the full-length transcript of the human gene will provide more detail of the gene sequence as well as its predicted coding protein. Based on the developmentally regulated expression, this gene was selected for characterization and considered to be a potential candidate gene involved in breast cancer pathogenesis. To achieve this primary aim, the full-length transcript sequence of human gene will be identified and its expression determined in various tissues. A predicted protein sequence will also be characterized to identify its possible function through protein database homology searches and functional motif prediction tools. Characterization of the genomic structure will also be undertaken to provide information on the exon/intron boundaries for gene mutation screening in breast cancer samples. Finally, the expression study in mouse embryos will be determined. This will require the isolation of a mouse cDNA clone from cDNA libraries which will be used for RNA probe preparation for hybridization to whole body mouse embryo.

### **3.2 Methods**

The following methods outline the general procedure used in this chapter. Most of the detailed technical procedures are given in "Materials and Methods" (chapter 2). Only the methods that are specific to this chapter are given in detail.

### 3.2.1 Identification of the Transcript sequence

#### 3.2.1.1 Database searching

BLAST (Basic Local Alignment Search Tool) programs (Altschul *et al.*, 1997) were used for sequence homology searches through public sequence databases. For the transcript sequence identification, both ET6 and ET7 sequences were initially used, by BLASTN (for nucleotide sequence homology searches), to search for homologous cDNAs from both EST and non-redundant databases. The homologous cDNA sequences, MMDEF85 and N35521, were first obtained and were subsequently used to compare with other cDNAs for further sequence analysis. The cDNA sequences of clones isolated from cDNA libraries (described in 3.2.1.2) were also used to search for homologous sequences.

#### 3.2.1.2 cDNA library screening

A human fetal brain cDNA library (formatted from the random primed, for 5' stretched, and the oligo-dT primed libraries, Clontech) and an oligo-dT primed cDNA library from human fetal liver (Stratagene) were screened utilizing trapped exon ET6 as a probe. ET6 of 134 bp was excised from the pAMP10 vector with *Not I/Sal I* double digestion,  $\alpha\text{P}^{32}$ -labeled by random priming, and hybridized to cDNA library filters (section 2.2.18.2). Duplicate plaques that gave hybridized signal were picked, resuspended in SM buffer, and eluted overnight. The eluates were re-plated, and plaques were re-hybridized until plaque-purified. Positive plaques were isolated and cDNA inserts were directly amplified from purified phages using vector primers and gene-specific primers (*Table 3.1*), and sequenced by the Dye-terminator method (Perkin-Elmer) on ABI 373 DNA sequencer using the same set of primers above.

#### 3.2.1.3 Identification of the transcript using genomic sequence



Genomic sequence of the cosmid clone 301F3 (Kremmidiotis, G. and Gardner, A., personal communication, 1999) was used to probe express<sup>ed</sup> sequences from both human and mouse EST databases. The homologous EST sequences were analyzed and primers designed for subsequent RT-PCR experiment. To optimize for both human and mouse sequences, primers were designed from the homologous regions that are human-mouse identical (**Table 3.1**). For the identification of human transcript, RT-PCR was carried out on the total RNA isolated from human fetal heart. The cDNA products were separated by agarose gel electrophoresis. The cDNA bands were purified from the gel and sequenced. The cDNA clone (qo68a07) corresponding to EST AI309221 that matched the 3' region of the transcript was also purchased, DNA isolated and sequenced. The corresponding mouse cDNA was determined by PCR amplification from mouse cDNA libraries derived from two mouse tissues, fetal brain (Gibco, Life Technology) and embryo at 10.5 day post coitus (d.p.c.) (Gibco, Life Technology). <sup>A/line?</sup> Mouse cDNA clone (courtesy from Dr Timothy Cox, Department of Genetics, Adelaide University) was also isolated from mouse fetal brain cDNA library (Stratagene) utilizing<sup>a</sup> cDNA probe derived from RT-PCR product mentioned above.

#### 3.2.1.4 PCR and RT-PCR

All PCR and RT-PCR procedures were performed as described in section 2.2.12.1 and section 2.2.13. The sequences of all oligonucleotide primers used for the identification of the transcript sequence are presented in **Table 3.1**. Those of primers designed for determination of the exon-intron boundaries are given in **Table 3.2**.

#### 3.2.2 Expression Analysis

The transcription products were analysed by Northern hybridization procedure ✓

(2.2.11.2). Northern blot filter containing 2 µg of poly(A)<sup>+</sup> mRNA from six different human tissues were hybridized according to manufacturer's instructions in "ExpressHyb" hybridization solution (Clontech). Two cDNA probes were used in separated hybridization reactions. The first was a cDNA insert of clone yx61a06 that showed a sequence homology to that of ET7-ET6. The second probe derived from RT-PCR products that were generated according to the homologous ESTs of the genomic region down stream from the first probe (see **Figure 3.4**).

Differential expression of the transcript was determined by hybridization of the second probe to human RNA dot blot filter (Human RNA Master Blot , Catalogue # 7770-1, Clontech) that contains mRNAs from various human tissues of both fetal and adult stages. Hybridization was also performed in "ExpressHyb" and followed the instruction recommended by manufacturer.

### **3.2.3 *In silico* Analysis of Protein function**

The BLASTP program was used to compare the T3 protein with the sequences of known, functional, and novel proteins in the public protein databases (NCBI). The functional domains of the protein were predicted using three software programs, Profile Scan, Pfam, and SMART. Protein localization analysis was by PSORT.

### **3.2.4 Whole-body *in situ* hybridization of the mouse embryos**

This part was performed at Department of Genetics, Adelaide University, under kind supervision of Dr. Timothy Cox, and with the assistance of Miss Sonia Donati.

#### **3.2.4.1 *Prehybridization of embryos***

Embryos at mid-gestation period (average 10.5 d.p.c) had been previously harvested,

fixed, and stored in 100% methanol. They were re-hydrated by sequential rinsing in (i) 75% methanol / 25% PBS, (ii) 50% methanol / 50% PBS, and (iii) 25% methanol / 75% PBS at 4 °C, each for 5 minutes. Embryos were washed a further three times in PBT [0.1% Tween-20 (polyoxyethylene-sorbitan monolaurate) in PBS] for 5 minutes each at room temperature, then treated with 10 µg/ml Proteinase K in PBS for an appropriate amount of time. For a general guideline, 9.5 d.p.c. embryos were treated for 10 minutes, with each additional day of development requiring an additional 2 minutes. After Proteinase K treatment, embryos were washed twice with PBT for 5 minutes each and re-fixed in PG-PBT [0.2% glutaraldehyde (EM-grade, Sigma), 4% paraformaldehyde (PFA) in PBT] at room temperature for 20 minutes. After fixation, embryos were washed five times in PBT, each for 5 minutes at room temperature. The embryos were transferred to 5 ml capped tubes and washed at room temperature, once with a 1:1 mix of HB and PBT for 15 minutes and once with HB for 10 minutes [HB (hybridization buffer): 50% formamide; 5x SSC (pH 4.5); 50 µg/ml heparin; 0.1% Tween-20]. The tube was then filled with HB containing 100 µg/ml tRNA and 100 µg/ml sheared denatured herring sperm DNA, and prehybridization was allowed for 4 hours at 70 °C.

#### 3.2.4.2 Preparation of RNA probes and hybridization of embryos

Both sense and anti-sense RNA probes were prepared from cloned cDNA fragment in pBluescriptSK (pBSK, Strategene) vector. The mouse *mt3* cDNA fragment of 576 bp was previously cloned into pBSK, in the 5'→3' direction between *Hind*III and *Bam*HI sites of the multiple cloning site region. The recombinant plasmid was linearized either by *Hind*III or *Bam*HI digestion to obtain the DNA templates for synthesis of anti-sense and sense RNA probes respectively. The linearized plasmid was purified by phenol/chloroform extraction, followed by ethanol precipitation. To 1 µl of linearized plasmid (1 µg/µl), 2 µl

of 10x transcription buffer containing 0.1 M DTT, 2  $\mu$ l of 10x NTP labelling mixture containing DIG-UTP (digoxigenin-UTP), 1  $\mu$ l of RNase inhibitor (20 units/ $\mu$ l), RNase-free water to a final volume of 18  $\mu$ l, and 2  $\mu$ l of either T3 or T7 RNA polymerase (20 units/ $\mu$ l) were added depending on the linearization of the plasmid and the RNA probe being synthesized. The reaction was allowed at 37 °C for 2 hours. The RNA transcript was analyzed by electrophoresis of 1 $\mu$ l aliquot of the reaction mixture on 1% agarose gel in TBE, followed by ethidium bromide staining and UV illumination. The amount of RNA synthesized was estimated from the band intensity by assuming that a 10 fold increase in intensity of the RNA band to that of the plasmid represented approximately 10  $\mu$ g of probe. To remove the DNA template, 2  $\mu$ l of DNase I (10 units/ $\mu$ l, RNase-free) was added to the reaction mixture and incubated for a further 15 minutes at 37 °C. After 15-minute incubation, 2.5  $\mu$ l of 4 M LiCl and 75  $\mu$ l prechilled 100% ethanol were added and mixed. The mixture was left at -70 °C for 30 minutes and followed by centrifugation at 12,000xg for 15 minutes. The pellet was washed with 50  $\mu$ l of 70% cold ethanol (v/v), dried, and dissolved in RNase-free water to a final concentration of approximately 1  $\mu$ g/ $\mu$ l. The RNA probe, DIG-labelled, was denatured in 1 ml HB containing 100  $\mu$ g/ml tRNA and 100  $\mu$ g/ml sheared denatured herring sperm DNA, at 95 °C for 3 minutes then placed at 70 °C to equilibrate. The denatured RNA probe was then added to the embryos in hybridization mix (3.2.4.1) to a final concentration of 1  $\mu$ g/ml and hybridization was allowed overnight at 70 °C, with agitation.

#### **3.2.4.3 Post-hybridization washes and embryo blocking**

After the completion of embryo hybridization the following washes were carried out in HB, SSC-FT (2x SSC pH 4.5; 50% formamide; 0.1% Tween-20), and 1x TBST [10x

stock: 8% (w/v) NaCl; 0.2% (w/v) KCl; 1% Tween-20; 25 ml 1M Tris pH 7.5 in the final volume 100 ml]. Embryos were washed twice with HB at 70 °C, each for 30 minutes, then twice with SSC-FT for 5 minutes each at room temperature. This followed by three washes with SSC-FT at 65 °C for 30 minutes each. Embryos were allowed to cool down to room temperature and washed three times, for 5 minutes each, with 1x TBST at room temperature. The embryos were then blocked by incubating in 10% heat-inactivated sheep serum in TBST, for 1-3 hours at room temperature.

#### 3.2.4.4 Antibody binding

The anti-DIG antibody, alkaline phosphatase conjugated (Boehringer Mannheim) was used to <sup>for the detection of mRNA (target) in</sup> determine the hybridized embryos. The antibody was first pre-adsorbed by incubating 0.5 µl of anti-DIG antibody in 1 ml TBST containing 1% heat-inactivated sheep serum and 2 mg heat-inactivated mouse powder. The reaction was allowed for 1 hour at 4 °C, with agitation. After completion, the reaction tube was centrifuged at 15,000 for 5 minutes and the supernatant containing pre-adsorbed antibody was collected. To allow antibody binding, the solution was removed from the pre-blocked embryos (3.2.4.3) and was replaced with pre-adsorbed antibody and incubated overnight at 4 °C.

#### 3.2.4.5 Post-antibody washes and histochemistry

After the completion of overnight incubation with pre-adsorb antibody, the excess antibody conjugate was removed and the embryos were washed three times with TBST at room temperature, each for 5 minutes. The embryos were then transferred to the new vials and washed a further five times, for 1 hour each, at room temperature with TBST containing 2 mM Levamisole. They were left overnight at 4 °C in the new TBST/Levamisole, with agitation.

Prior to color development, the embryos were washed three times, for 10 minutes each, at room temperature in alkaline phosphate buffer (APB) containing 2 mM Levamisole (APB: 100 mM NaCl; 50 mM MgCl<sub>2</sub>; 0.1% Tween-20; 100 mM Tris, pH 9.5). To develop the color, the embryos were incubated in APB/2mM Levamisole containing 4.5 µl/ml NBT and 3.5 µl/ml BCIP. This was carried out at room temperature under dark conditions on a shaker table. The reaction was allowed <sup>to progress</sup> for a period of time until the color, dark purple, was visible, by observing the progress of the reaction for brief intervals under a dissecting microscope. The color reaction was stopped by removing and rinsing the embryos three times at room temperature in PBT containing 1 mM EDTA and 1% glacial acetic acid. The embryos were subsequently washed a further three times, each for 5 minutes, in PBT / 1 mM EDTA / 1% glacial acetic acid. The embryos were then stored, at 4 °C, in 80% glycerol and 1% acetic acid in PBT. Each embryo was examined under the dissecting microscope and the photographs were taken for further analysis.

### 3.2.5 Study for T3 mutations in Breast Cancer

Mutation study was undertaken on the genomic DNA isolated from two panels of total <sup>of</sup> 48 breast cancer tissue samples and from 24 breast cancer cell lines. The breast cancer tissue samples were taken from <sup>the</sup> patients, with informed consent, from two medical centers: the Flinders Medical Center, South Australia, and the Guy's Hospital, London, UK. Screening for gene mutation was by SSCP analysis of the genomic PCR fragment utilizing the primer pairs designed for the amplification of each exon, including part of its flanking introns. All primers were custom-labeled with a fluorescence dye (HEX). The HEX labeled genomic PCR products were separated by electrophoresis on the non-denaturing polyacrylamide gel, 4.5<sup>g</sup> and 6 % gel with 2% glycerol in TBE buffer, using the Gene Scan 2000 (Corbett Research). PCR products showing a conformational change were

**Table 3.1**  
**Sequences and Positions of the Primers used for T3 Transcript Sequence Determination**

<i>No</i>	<i>Primer</i>	<i>Primer Sequence (5'→3')</i>	<i>Nucleotide Position in the Transcript</i>
<b><i>Common Primers for Human and Mouse homologous cDNA<sup>1</sup></i></b>			<b><i>Positions in the human transcript</i></b>
1	35521-P	cagccgggagacagatgcgag	29-49
2	35521-F	caggtgggatgctatggaatag	58-80
3	35521-U	gcttggtgaaggggtgagg	131-112
4	35521-R	catcacgcggttcagggcagcg	253-233
5	N355-D	gtgcaagcaggtgattctgg	343-362
6	35521-8	gtgtaacaccatcatctggg	535-554
7	T3EA-F	gtgttattaccgctgtcacag	586-606
8	T3EA-R	ctcggcacagcggtaatcc	727-709
9	T3EB1-F	actgcagccactgccactg	798-816
10	T3EB1-R	caccagctcctccacgtag	982-964
11	T3EB2-F	cttcatcacctgcagggag	1027-1045
12	T3EB2-R	ctccctgcaggtgatgaag	1045-1027
13	T3EC-F	ggcagcattttgtggagaacg	1083-1103
14	T3ED-F	cagccacacgtctgtgtgc	1279-1297
15	T3ED-R	gacgtgtggctgtcgaacg	1290-1272
16	T3XE-F	accacttgtcccaagtgtgc	1346-1365
17	AI309-R	tgcaggatcccacgcctaac	2275-2255
18	T3EF-R	cccacactgttgatccaacc	3052-3033
<b><i>Primers Specific for Mouse transcript</i></b>			<b><i>Positions in the mouse transcript</i></b>
19	MT3-A	tccttctgcaccatagagatg	1765-1785
20	MT3-B	gcaaggccatctctatgggtg	1792-1773
21	MT3-C	gaggccaatgataactgacag	2273-2253
22	MT3-D	atcacatgtctgtcctgagtc	3016-3036
<b><i>Vector Primers</i></b>			
23	T3	attaaccctcactaaagga	
24	T7	taatacgactcactataggg	
25	M13F	tgtaaaacgacggccagt	
26	M13R	caggaaacagctatgacc	
27	gt11F	tggcgacgactcctggagcccg	
28	gt11R	tgacaccagaccaactggtaatgg	
29	SP6	gatttaggtgacactatag	

<sup>1</sup>All except primers no. 1,5,17,18 were also used for the identification of the mouse transcript sequences.

Primers 35521-R and 35521-U were also utilized in the 5'RACE experiment in combination with SMART 5' primer (2.2.17.1).

reamplified from 100 ng of genomic DNA with corresponding unlabelled primers. The products were purified and sequenced.

### **3.3 Results**

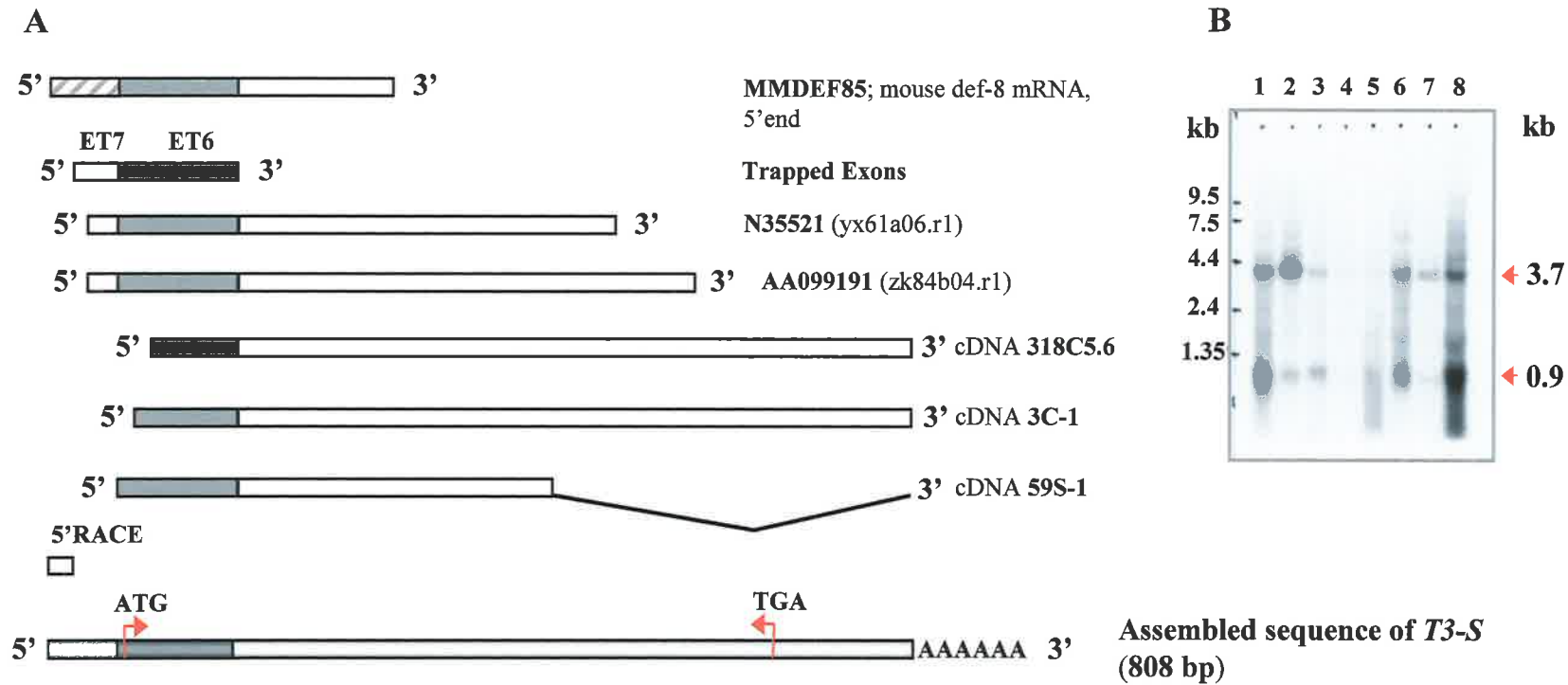
#### **3.3.1 Determination of Transcription Unit 3 (T3) Transcribed Sequence**

##### ***3.3.1.1 Database analysis and cDNA sequencing***

Based on the 16q24.3 physical map (*Figure 1.3*), the trapped exons ET6 and ET7 are less than 10 kb apart and were considered likely to be part of the same gene. Non-redundant database searches initially identified that ET6 showed sequence homology to the transcript MMDEF85, the 5' sequence of the mouse developmentally regulated gene, def-8. However, results from analyses of human EST database indicated that both ET6 and ET7 belonged to the same transcript as shown by their sequence homology to the same ESTs, yx61a06.r1 (entry N35521) and zk84b04.r1 (entry AA099191). N35521 and AA099191 are the 5' sequence of cDNA clones, yx61a06 and zk84b04 respectively, which were directionally cloned into the vectors. Therefore, the orientation of these trapped products is ET7-ET6, corresponding with the 5'→3' direction of the available sequences of both ESTs (*Figure 3.1A*). Furthermore, sequence alignment of N35521 and MMDEF85 showed that continuation of homology between human and mouse cDNA <sup>was</sup> were also observed in the region beyond the 3' end of ET6. These results suggested that ET7/ET6 sequence was part of the human gene, designated "transcription unit 3" (T3), which is homologous to the mouse def-8 gene.

To extend the transcribed sequence of T3, ET6 was excised from the pAMP10 vector and used as a probe for the screening of cDNA libraries (3.2.1.2). The cDNA clone 318C5.6 was obtained from screening a fetal brain cDNA library cloned into the λgt11 vector (Clontech). ET6 was also used to screen a fetal liver cDNA library inserted in λZAP





**Figure 3.1** : **A**) Alignment of Transcription Unit 3, “Short isoform” (T3-S), with the cDNA clones (318C5.6, 3C-1, and 59S-1), ESTs (N35521 and AA099191), trapped exons (ET6 and ET7), 5’ RACE product and the 5’ sequence of mouse def-8 mRNA. *Hatched box* indicates part of the mouse cDNA that is not homologous to the human sequence (ET7). *Single line* represents the intronic sequence found within the 3’half of clone 59S-1. Translation start (ATG) and termination (TGA) sites are indicated by forward and reverse arrows respectively. The 5’ RACE product extends the transcript sequence from ET7. **B**) Expression analysis of T3. A commercial Northern membrane was hybridized with cDNA insert of clone yx61a06. The tissue sources of mRNA are: **1 heart, 2 brain, 3 placenta, 4 lung, 5 liver, 6 skeletal muscle, 7 kidney, 8 pancreas**. Size markers, in kb, are indicated on the left hand side. Two major bands of 3.7 kb and 0.9 kb represent the *long* (L) and *short* (S) isoform of T3 mRNA transcripts respectively.

vector (Stratagene), from which cDNA clones 3C-1 and 3C-2 were isolated. PCR analysis of 3C-1 and 3C-2 using the vector primers (T3 and T7) as well as a combination of vector primer and gene specific primer (35521-R) indicated that both contained the same insert size. Therefore, in subsequent studies only the 3C-1 clone was analyzed. In addition, the PCR product amplified from the region close to the 5' end of cDNA clone yx61a06 (primers 35521-F and -R) was also used to screen the same fetal brain cDNA library (Clontech) and the cDNA clone 59S-1 was isolated. The DNA sequence of 318C5.6, 3C-1, and 59S-1 was determined by using the vector primers as well as the primers designed to the sequence available from ESTs yx61a06.r1 (N35521) and zk84b04.r1 (AA099191). The results showed that both 318C5.6 and 3C-1 shared the same sequence with those of N35521 and AA099191 but also extended the sequence down-stream from their 3' end. The extended sequence revealed the 3' polyadenylation signal (AATTAA) characteristic of 3' end of the transcript. The cDNA 59S-1, although containing an insert of 1.2 kb, larger than those of 318C5.6 or 3C-1, showed the identical sequence only from the 5' end to a point later found to be the exon 4 boundary. The remainder of the sequence was intronic indicating an incomplete splicing product. The 5' ends of all isolated cDNA clones did not, however, extend beyond the ET7 sequence. Therefore, 5' RACE was also performed (described in section 2.2.17.1) using gene specific primers and the SMART 5' primer (*Table 3.1*). This allowed an additional 24 bp sequence to be determined upstream from the 5' end of ET7. The relationship between the various cDNA clones is shown diagrammatically in *Figure 3.1A*.

### **3.3.1.2 Transcript Sequence Analysis**

The sequence of cDNA clones, trapped exons, and 5' RACE product were assembled and identified a 808 bp transcript with an open reading frame (ORF) of 591 bp extending


from the start codon, followed by a 3' untranslated region (3'UTR) of 131 bp (**Figure 3.2**). The methionine start codon, ATG, and its surrounding nucleotides conform with the Kozak consensus sequence characteristic of translation start site (Kozak, 1989) and is preceded by an in-frame stop codon, TAG, 66 bp upstream from the start site. The *T3* ORF was predicted to code for a protein of 197 amino acids. Two possible polyadenylation signals were identified within the 3'UTR of the transcript, the first (AATTAA) at position 715-717, and the second (ATTAAA) at position 777-782. From an examination of the DNA sequence of the cDNA clones isolated from the cDNA libraries and those identified from the EST database only the second site appeared to be functional.

### 3.3.2 Expression analysis of *T3*

The cDNA insert of clone yx61a06 was used as a probe to hybridize to a commercial multiple tissue Northern membrane (Clontech). Two prominent bands with different intensity of approximately 0.9 kb and 3.7 kb were detected in all tissues examined (**Figure 3.1B**). The result indicated that at least two transcription isoforms were expressed but the expression of the two isoforms varied in different tissue. The transcript that had been sequenced was equivalent to the short isoform (0.9 kb), designated "*T3-S*", and appeared to be highly expressed in heart, skeletal muscle, and pancreas. Expression of *T3-S* was relatively low in brain, placenta, and liver, with very low expression presented in lung and kidney. The 3.7-kb isoform, assigned as "*T3-L*", was strongly expressed in brain which was contrast to the expression of the short isoform. Medium level of *T3-L* expression was seen in heart, skeletal muscle, and pancreas. In kidney tissue, expression of *T3-L* was stronger than that of *T3-S*. The remaining tissues showed a low level of expression of *T3-L*.

### 3.3.3 Gene Structure of *T3-S*

To explore the gene organization for *T3-S*, primers designed to both strands of *T3-S*

gtagcggggcgccggcgccggatccagcgcagccgggagacagatgcgagggcggtcaggtgggatgcATCGAATATG 80  
 M E Y 3  
**ET7 (34 bp) 1**  **2**  
**ET6 (134 bp)**  
 ATGAGAAGCTGGCCCGTTTCCGGCAGGCCACCTCAACCCCTTCAACAAGCAGTCTGGGCCGAGACAGCATGAGCAGGGC 160  
 D E K L A R F R Q A H L N P F N K Q S G P R Q H E Q G 30  
 CCTGGGGAGGAGGTCCCGACGTCACTCCTGAAGAGGCCCTGCCTGAGCTGCCCCCTGGGGAGCCGGAATTCCGCTGCC 240  
 P G E E V P D V T P E E A L P E L P P G E P E F R C P 57  
 TGAACGCGTGATGGATCTCGGCCTGTCTGAGGACCACTTCTCCCGCCCTGTGGTCTGTTCTGGCCTCTGACGTCACG 320  
 E R V M D L G L S E D H F S R P V G L F L A S D V Q 83  
 AGCTGCGGCAGGCGATCGAGGAGTGCAAGCAGGTGATTCTGGAGCTGCCCGAGCAGTCGGAGAAGCAGAAGGATGCCGTG 400  
 Q L R Q A I E E C K Q V I L E L P E Q S E K Q K D A V 110  
 GTGCGACTCATCCACCTCCGGCTGAAGCTCCAGGAGCTGAAGGACCCCAATGAGGATGAGCCAAACATCCGAGTGCTCCT 480  
 V R L I H L R L K L Q E L K D P N E D E P N I R V L L 137  
 TGAGCACCGCTTTTACAAGGAGAAGAGCAAGAGCGTCAAGCAGACCTGTGACAAGTGTAAACACCATCATCTGGGGGCTCA 560  
 E H R F Y K E K S K S V K Q T C D K C N T I I W G L 163  
 TTCAGACCTGGTACACCTGCACAGGTGGGCCGAGACCCAGACGAGGAGTGAAGGAATGAGAGAGACCAAAGTTCTCGCCTC 640  
 I Q T W Y T C T G G P R P R R G V R N E R D Q S S C L 190  
 CGCTGGGCTCACATTCAGATGTGAGCTATCCAGTGTGGTGGCCACTAGCCAAACATGGCCAGTCAAAGTTAAATTAAC 720  
 R W A H I Q M \* 197  
 AATATGCATTTCTCGTTGTACTGTCTACATTGGACAGGGCTGTTCTGGATGAGAAAATTAAACTAGTGAATTACAAAAA 800  
 AAAAAAA 808

**Figure 3.2** : Nucleotide and deduced amino acid sequence of the “short transcription isoform” of *T3 (T3-S)*. Numbers in the right hand column indicate the nucleotide and amino acid positions. The translation start site and the termination codon are indicated at positions 71-73 and 662-664 respectively. The nucleotides surrounding the ATG start site that conform with the Kozak consensus sequence are underlined. An upstream stop codon preceding the translation start site is located at positions 2-4. A polyadenylation signal beginning at nucleotide 777 is underlined. The exon boundaries are indicated by the vertical lines with double-head arrows and the given numbers of exons. The trapped exon ET7 (underline) is part of exon 1 whereas ET6 was trapped from exon 2. Vertical line with forward arrow at nucleotide position 585 indicates the boundary of exon 5 utilized by *T3-L* isoform (*Figure 3.6*) and the flanking donor splice signal (GT) is underlined.

***Table 3.2***

**Sequence of Primers used for the Identification of T3 Gene structure**

No	Primers	Nucleotide Sequences (5' → 3')	Methods utilized
1	35521-P	cagccgggagacagatgcgag	Direct sequencing of cosmid 318C5
2	35521-U	gcttggtgaaggggtgagg	
3	35521-1	tcaacccttcaacaagcag	PCR amplification and Direct Sequencing of cosmid 318C5
4	35521-R	catcacgcgttcagggcagcg	
5	35521-2	aattccgctgccctgaacgc	
6	35521-4	ctgcttgactcctcgatcg	
7	35521-3	gatcgaggagtgcaagcagg	
8	35521-5	tcattggggtccttcagctc	
9	35521-6	agctgaaggacccaatgag	
10	35521-7	cccagatgatggtgttacac	
11	35521-8	gtgtaacaccatcatctggg	
12	35521-9	acatctgaatgtgagcccag	

sequence (**Table 3.2**) were used to PCR amplified the genomic DNA from cosmid clone 318C5, and the amplified products were sequenced from both ends. The same primers were also used to perform direct sequencing of cosmid 318C5. Alignment of genomic sequence with that of cDNA allowed the identification of *T3-S* exon/intron boundaries. As shown in **Figure 3.3**, the gene for *T3-S* consists of 5 exons ranging in size from 60 bp to 353 bp. The splice junction boundaries (**Table 3.3**) conform to the consensus sequences for donor and acceptor splicing sites provided by Shapiro and Senapathy (1987). The translation initiation codon, ATG, is situated within exon 2 (**Figure 3.2** and **3.3A**). The 5' UTR constitutes entire exon 1 and 10 bp 5' part of exon 2. The last exon, exon 5, contains part of coding sequence, stop codon (TGA), and the 134 bp 3'UTR. The trapped product ET7 (35 bp) appeared to be trapped from 3' part of exon 1, whereas ET6 was completely trapped from exon 2.

### 3.3.4 Determination *T3-L* cDNA Sequence

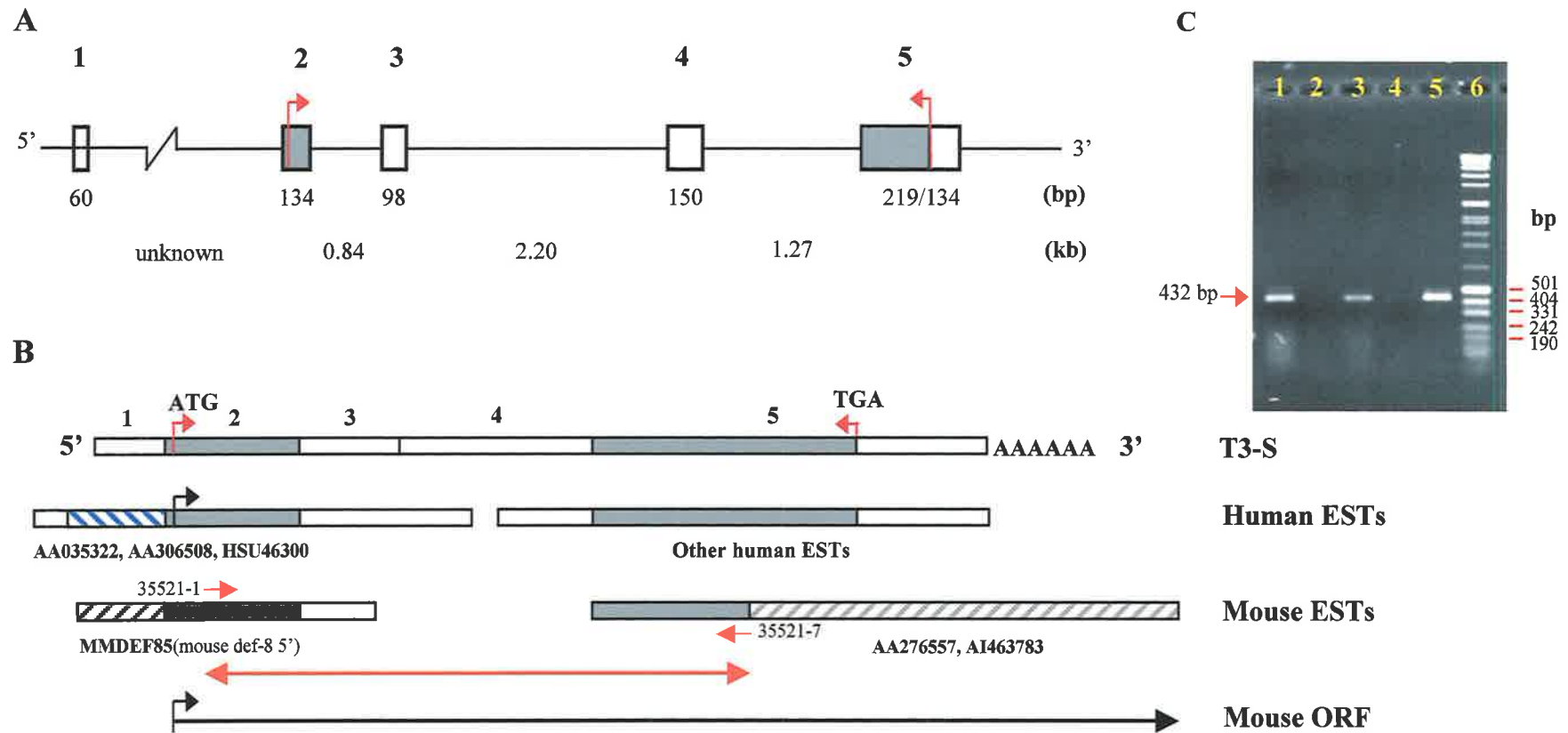
#### 3.3.4.1 Database Analysis

The result of Northern analysis indicated that part of *T3-L* shared the same sequence with *T3-S*. However, it was not known whether they shared the same ORF or only shared the sequence but with different ORF. Only with the identification of a complete cDNA sequence of *T3-L* would the relationship of both forms be addressed. A complete sequence of *T3-S* was used as a query in an attempt to identify more extended sequence from the EST database. Search results showed that most of the human ESTs homologous to *T3-S*, from either the 5' or 3' direction (**Figure 3.3B**), failed to provide more extended sequences. Some human ESTs appeared to include an extra exon between exons 1 and 2. This extra exon, however, did not create any new ORF indicating that this is likely to be an alternatively spliced form of 5' UTR. The same extra exon was also detected in the 5'

RACE products (*Figure 3.9*), therefore confirming this possibility.

### 3.3.4.2 Evidence of a Mouse transcript homologous to *T3-L*

In addition to the homology initially observed between the 5' sequence of mouse *def-8* and human *T3-S* exon 2 and part of exon 3, database searches also identified two mouse ESTs, AA276557 (va24d01.r1) and AI463783 (va24d01.y1), showing a high degree of sequence homology (89%) with the 3' part of *T3-S*. This homology, however, included only 137 bp from the 5' end of both ESTs, corresponding with the region between position 448 and 584 within the *T3-S* exon 5 (*Figure 3.2* and *3.3B*). No homology to human *T3-S* was observed in the remaining mouse sequence that extended 3' down stream a further 297 bp. To determine if *def-8* and AA276557/AI463783 were part of the same gene, a pair of primers were selected from the regions showing 100% human/mouse homology (35521-1 and -7) within exon 2 and exon 5 (*Figure 3.3B*). They were successfully used to PCR amplify the cDNAs (*Figure 3.3C*) from mouse brain and embryo cDNA libraries (Gibco). Sequencing of these PCR products identified the 432 bp mouse cDNA with high degree of homology with that of *T3-S*. Sequence assembly of this 432 bp product, the *def-8* 5' sequence, and the sequence of AA276557 and AI463783 resulted in a mouse cDNA sequence of 899 bp. From this assembled sequence, an ORF with no termination was obtained, sharing the equivalent translation start codon as well as the amino sequence homology with that of human *T3-S*. This homology, however, deviated immediately after the nucleotide position 584 of *T3-S*, mentioned above. Continuation of the ORF was also extended by the mouse EST AI661425, the 3' sequence of the same cDNA clone (va24d01) origin of AI463783, and again showed no termination. This assembled mouse cDNA sequence with a longer ORF, compared with that of *T3-S*, might represent the mouse homologue of *T3-L*.



**Figure 3.3 :** A) Exon and Intron Organization of *T3-S*. The start (forward arrow) and stop (reverse arrow) codons are situated within exon 2 and exon 5 respectively. The size of each exon and intron (estimated from the PCR product) are indicated by the numbers underneath, as “bp” for exon and “kb” for intron. The size of intron 1 was unknown at the time where genomic sequence was not available B) Alignment of *T3-S* cDNA with the human and mouse ESTs. The reverse hatched box represents an extra exon spliced in between exon1 and exon 2, designated exon 1a, seen in at least three human ESTs. Sequence homology between mouse ESTs and *T3-S* is indicated by corresponding filled and opened boxes, with hatched boxes are non-homologous. Line with double arrow-heads represents the RT-PCR product (also indicates in C), using primer 35521-1 and 35521-7, linking mouse def-8 and AA276557/AI463783. The mouse ORF created by assembled sequence is also indicated utilizing the equivalent start site with that of *T3-S*. C) Agarose gel electrophoresis of RT-PCR products amplified from mouse cDNA libraries: lane 1 fetal brain, lane 2 thymus, lane 3 embryo (10.5 dpc), lane 4 liver, and the product, lane 5 amplified from human cDNA 3C-1. The DNA size marker, lane 6, and PCR products of 432 bp are indicated.



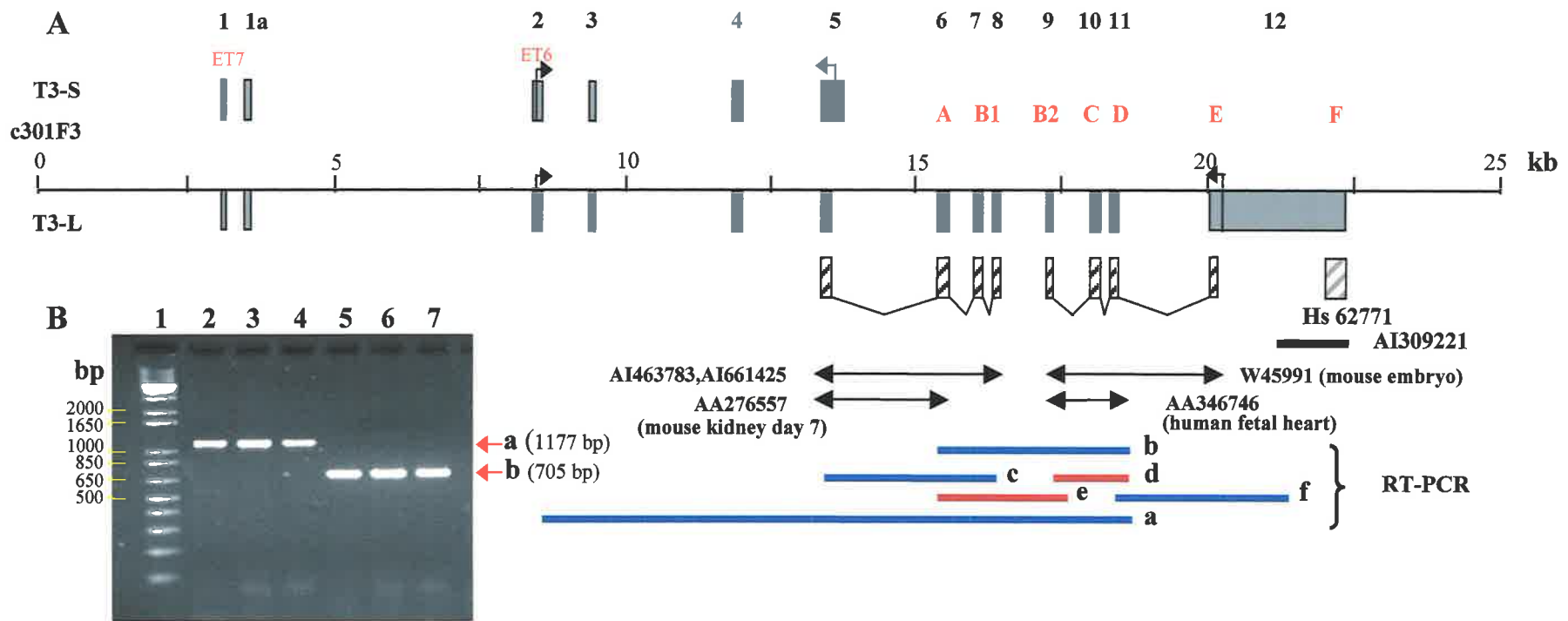
### 3.3.4.3 *T3-L* Sequence Assembly

Cosmid 301F3 was part of the cosmid contig within the 16q24.3 physical map from which trapped exons ET6 and ET7 were isolated. As the large scale sequencing project has been established, the sequence of cosmid 301F3 was determined (Kremmidiotis, G. and Gardner, A., personal communication, 1999). To identify additional exons within close proximity to *T3-S*, the sequence of 301F3 was used to search database for homology to the partially sequenced transcripts. As indicated in **Figure 3.4A**, 301F3 showed regions of homology to three mouse ESTs, AA276557, AI463783, and AI661425, which were linked to *T3-S* exon 5, previously mentioned in section 3.3.4.2. In addition, genomic sequence homologous to the mouse EST W45991 and human EST AA346746 was also identified in the adjacent region approximately 1 kb down stream. Altogether, the regions of homology between EST sequences and genomic sequence established additional exons initially designated exons A, B1, B2, C, D, and E, which were possibly part of *T3-L*. A set of primers (**Table 3.1**) designed from the regions showing 100% human/mouse sequence homology within some exons were successfully used to RT-PCR amplify the human cDNA from human fetal tissue RNA (**Figure 3.4**). After sequencing, these cDNA fragments were assembled and formed part of *T3-L* cDNA by linking to the position 584 within *T3-S* exon 5. This linked sequence extended another 822 bp and also created a continuation of ORF from exon 5 with no termination similar to that seen in mouse cDNA (section 3.3.4.2). Two RT-PCR products spanning from exon A to exon D (**Figure 3.4A**) were also used to hybridize to a multiple tissue Northern membrane (ClonTech). As expected, a single band of 3.7 kb corresponding to *T3-L* was detected (**Figure 3.5A**). The variation in band intensity between the different tissue was similar to the results seen when the membrane was probed with *T3-S* cDNA (section 3.3.2).

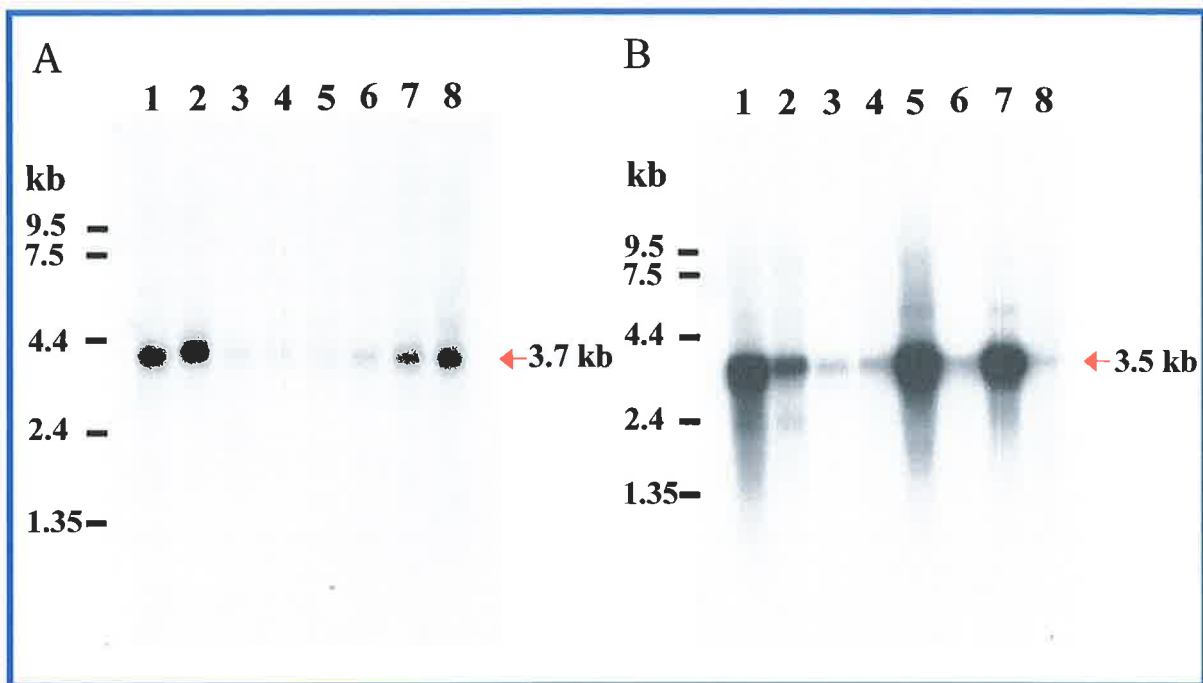
The joined cDNA sequence from exon 1 to exon E was only 1,406 bp in length with no evidence of ORF termination nor 3' UTR. Therefore, additional 3' sequence was sought by further analysis of the EST database. The genomic region, designated exon F, situated approximately 2.4 kb down stream from exon E (*Figure 3.4A*) was matched with a group of human ESTs that belonged to the UniGene cluster Hs.62771. Sequencing of a cDNA clone (qo68a07) of AI309221, one of these ESTs, identified an insert sequence of 1274 bp. This sequence showed a polyadenylation signal (ATTAAA) and a polyadenylation site 13 bp down stream from this signal. This sequence was then successfully joined with exon E by RT-PCR using a primer from exon D (ED-F) and the reverse primer (AI309-R) generated from this cDNA sequence (*Table 3.1*). The final sequence assembly identified a complete *T3-L* cDNA sequence of 3,525 bp (*Figure 3.6*), consistent with the size of the band on Northern analysis.

At this stage the gene designed "*T3*" codes for at least two transcription isoforms, **T3-S** and **T3-L**, based on the Northern analysis results, the transcribed sequences, and the evidence of shared ORF.

In addition to the assembled mouse cDNA sequence (section 3.3.4.2), identification of the complete mouse transcript sequence homologous to human *T3-L* was also undertaken. The same set of primers, which were used in the human *T3-L* transcribed sequence assembly (*Table 3.1*), were also successfully used to amplify and confirm the mouse cDNA sequence (*Figure 3.4*). Since the mouse EST W45991 provided the most 3' region of the available cDNA sequence, the clone mc81g01 that was used to generate this EST sequence was isolated and DNA sequenced. The extended sequence from W45991 was then used for additional dbEST analysis, and followed by the sequential EST searching for the overlapping cDNA sequences. A contig of mouse ESTs was finally obtained, and the sequences were assembled. This extended sequence formed mostly the 3' UTR of the



**Figure 3.4 :** Transcript sequence assembly and Gene structure of *T3-L*. **A)** Genomic structure of *T3-L* and alignment of the gene with ESTs. All exons of both isoforms of *T3* are indicated by filled boxes. The location of each exon and the length of introns are indicated by the relative distance on genomic DNA, cosmid 301F3. Exons A, B1, B2, C, D, E, and F were initially predicted from the sequence homology of genomic DNA with ESTs (hatched boxes). The single lines between each homology box are introduced for gaps. The extend of each EST showing homology to genomic sequence is indicated by double-headed arrow. The cDNA insert of clone represented by EST AI309221 was completely sequenced. **Colour bars** (a, b, c, d, e, and f) represent the extend of RT-PCR products successfully being amplified from both human and mouse cDNAs and were assembled to form a transcript sequence. The precise exon boundaries were confirmed by sequence alignment between genomic DNA and an assembled cDNA sequence. The *T3-L* was organized into 12 exons, exon 5 was from 137 bp 5' part of *T3-S* exon 5, exons 7 and 8 was from B1, and exons E and F formed part of exon 12. Translation start site and ORF termination are indicated by forward and reverse arrows respectively. **B)** Agarose gel electrophoresis of two major RT-PCR products amplified from human and mouse cDNAs: *lane 1* DNA size marker, *lanes 2 and 5* mouse fetal brain, *lanes 3 and 6* mouse embryo (10.5 dpc), *lanes 4 and 7* human fetal heart. The products size are: *band a*, 1177 bp, primer 35521-1 and T3ED-R; *band b*, 705 bp, primers T3EA-F and T3ED-R. The sequence of all primers are presented in *Table 3.1*.



**Figure 3.5** : Northern blot hybridization of human and mouse *T3*. **A)** Expression of *T3-L*. A commercial Northern membrane of human mRNA was hybridized with cDNA probe derived from RT-PCR product of 705 bp amplified using primers T3EA-F and T3ED-R (figure 3.3 B and Table 3.1). The tissue sources of mRNA are: **1** heart, **2** brain, **3** placenta, **4** lung, **5** liver, **6** skeletal muscle, **7** kidney, **8** pancreas. Size markers (kb) are indicated on the left side. **B)** Expression pattern of *mt3*. A Northern membrane of mouse mRNA was hybridized with two mouse cDNA probes, the PCR products of 432 bp (primers 35521-1 and 35521-7) and 750 bp (primers T3EA-F and T3ED-R) amplified from mouse cDNA library. The mouse tissue sources of mRNA are: **1** heart, **2** brain, **3** spleen, **4** lung, **5** liver, **6** skeletal muscle, **7** kidney, **8** testis.

mouse transcript showing no homology to that of *T3-L*. Finally all sequencing-confirmed mouse cDNA fragments were joined to form a 3,108 bp transcript sequence of the mouse orthologue of *T3-L* (**Figure 3.7**). Two mouse cDNA fragments corresponding with *T3-L* from exon 2-exon 5 and exon B1-exon E regions were also used to hybridize to the mouse multiple tissue Northern membrane (Clontech). In contrast to human *T3*, only a single band of approximately 3.4 kb was identified in all mouse tissue tested (**Figure 3.5 B**), with varying degree of expression. This indicated that the mouse orthologue of *T3* gene expressed only one transcript homologous to *T3-L*. Since the present mouse transcript was initially identified from MMDEF 85, the partial 5' sequence of mouse *def-8* cDNA (described in section 3.1, Hotfilder *et al.*, 1999), it is therefore the complete, full-length, transcribed sequence of *def-8* gene.

### 3.3.5 Transcript Sequence Analysis of *T3-L* and mouse orthologue

The transcript sequence of the "long isoform" of *T3* (*T3-L*) and its mouse orthologue, assigned as *mt3(def-8)*, consist of 3,525 bp and 3,108 bp respectively. *T3-L* was the extension of *T3-S* from the nucleotide 584 within exon 5 and constituted a 1353 bp ORF by utilizing the same initiation codon as that of *T3-S* (**Figure 3.6**). This transcript contains a large 3' UTR of 2,103 bp with a polyadenylation signal (ATTAAA) situated between nucleotides 3507 and 3512. This is 13 bp upstream from the polyadenylation site. The *T3-L* ORF is predicted to encode a protein of 451 amino acids, and the first 172 amino acids from the N-terminus is shared with that of *T3-S*. The *mt3(def-8)* transcript, by using the translation start codon equivalent to that of *T3*, encodes an ORF of 1344 bp (**Figure 3.7**) which shows a high degree of sequence homology (~86% identity) with the *T3-L* ORF. The *mt3(def-8)* ORF was predicted to code for a protein of 448 amino acids, which was ~95% identical with that of *T3-L*. It was followed by a 1,705 bp 3' UTR containing a signal

---

**Figure 3.6** : Nucleotide and deduced amino acid sequence of the “long transcription isoform “ of *T3 (T3-L)*. The nucleotide and amino acid positions are indicated by numbers in the right hand column . *T3-L* shares the first 584 nucleotides (exons 1-5) with that of *T3-S*. The same translation start site and upstream stop codon are also indicated. The termination codon is indicated at nucleotide positions 1424-1426. A polyadenylation signal beginning at nucleotide 3507 is underlined. The exon boundaries are indicated by the vertical lines with double-head arrows and the given numbers of exons. Two functional motifs, diacylglycerol/phorbol ester (GAG/PE) binding domain, residues 139-189 (consensus:H-CxxC-CxxC-HxxC-C) and plant homeodomain-like (PHD) zinc finger, residues 390-431 (consensus: CxxC-CxxC-HxxC-CxxC) are underlined and the amino acid residues corresponding with consensus are boxed, where “xx” indicates any two amino acids and “--” indicates various length of amino acid sequence. Detail information of these two domains can be found in “discussion” section.

gtagcggggcgccggcgatccagcgcagccgggagacagatgcgaggcgcggtcaggtgggatgct **ATGGAATATG** 80  
1 2  
↔ ↔  
M E Y 3  
 ATGAGAAGCTGGCCCGTTTCCGGCAGGCCACCTCAACCCCTTCAACAAGCAGTCTGGGCCGAGACAGCATGAGCAGGGC 160  
 D E K L A R F R Q A H L N P F N K Q S G P R Q H E Q G 30  
2 3  
↔ ↔  
 CCTGGGGAGGAGGTCCCGGACGTCACTCCTGAAGAGGCCCTGCCTGAGCTGCCCCCTGGGGAGCCGGAATTCCGCTGCC 240  
 P G E E V P D V T P E E A L P E L P P G E P E F R C P 57  
3 4  
↔ ↔  
 TGAACGCGTGATGGATCTCGGCCTGTCTGAGGACCACTTCTCCCGCCCTGTGGTCTGTTCTGGCCTCTGACGTCCAGC 320  
 E R V M D L G L S E D H F S R P V G L F L A S D V Q 83  
 AGCTGCGGCAGGCGATCGAGGAGTGAAGCAGGTGATTCTGGAGCTGCCCCGAGCAGTCGGAGAAGCAGAAGGATGCCGTG 400  
 Q L R Q A I E E C K Q V I L E L P E Q S E K Q K D A V 110  
4 5  
↔ ↔  
 GTGCGACTCATCCACCTCCGGCTGAAGCTCCAGGAGCTGAAGGACCCCAATGAGGATGAGCCAAACATCCGAGTGCTCCT 480  
 V R L I H L R L K L Q E L K D P N E D E P N I R V L L 137  
 TGAGCACCGCTTTTACAAGGAGAAGAGCAAGAGCGTCAAGCAGACCTGTGACAAGTGTAACACCATCATCTGGGGGCTCA 560  
 E H R F Y K E K S K S V K Q T C D K C N T I I W G L 163  
5 6  
↔ ↔  
 TTCAGACCTGGTACACCTGCACAGGGTGTATTACCGCTGTACAGTAAGTGCTTGAACCTCATCTCCAAGCCCTGTGTG 640  
 I Q T W Y T C T G C Y Y R C H S K C L N L I S K P C V 190  
 AGCTCCAAAGTCAGCCACCAAGCTGAATACGAACTGAACATCTGCCCTGAGACAGGGCTGGACAGCCAGGATTACCGCTG 720  
 S S K V S H Q A E Y E L N I C P E T G L D S Q D Y R C 217  
6 7  
↔ ↔  
 TGCCGAGTGCCGGGCGCCATCTCTCGGGGTGTGCCAGTGAGGCCAGGCAGTGCGACTACACCGGCCAGTACTACT 800  
 A E C R A P I S L R G V P S E A R Q C D Y T G Q Y Y 243  
7 8  
↔ ↔  
 GCAGCCACTGCCACTGGAACGACCTGGCTGTGATCCCTGCACGCGTTGTACACAACCTGGGACTTTGAGCCTCGAAGGTT 880  
 C S H C H W N D L A V I P A R V V H N W D F E P R K V 270  
 TCTCGTGCAGCATGCGCTACCTGGCGCTGATGGTGTCTCGGCCGTACTIONCAGGCTCCGGGAGATCAACCCTCTGCTGTT 960  
 S R C S M R Y L A L M V S R P V L R L R E I N P L L F 297  
8 9  
↔ ↔  
 CAGCTACGTGGAGGAGCTGGTGGAGATTGCAAGCTGCGCCAGGACATCCTGCTCATGAAGCCGTACTIONCAGGCTCC 1040  
 S Y V E E L V E I R K L R Q D I L L M K P Y F I T C 323  
9 10  
↔ ↔  
 GGGAGGCCATGGAGGCTCGTCTGCTGCTGCACTCCAGGATCGGCAGCATTTTGTGGAGAACGACGAGATGTACTCTGTC 1120  
 R E A M E A R L L L Q L Q D R Q H F V E N D E M Y S V 350  
 CAGGACCTCCTGGACGTGCATGCCGGCCGCTGGGCTGCTCGCTCACCGAGATCCACACGCTCTTCGCCAAGCACATCAA 1200  
 Q D L L D V H A G R L G C S L T E I H T L F A K H I K 377  
10 11  
↔ ↔  
 GCTGGACTGCGAGCGGTGCCAGGCCAAGGGCTTCGTGTGTGAGCTCTGCAGAGAGGGCGACGTGCTGTTCCCGTTCGACA 1280  
 L D C E R C Q A K G F V C E L C R E G D V L F P F D 403  
11 12  
↔ ↔  
 GCCACACGTCTGTGTGCGCCGACTGCTCCGCGGTCTTCCACAGGACTGCTACTACGACAACCTCCACCACTTGTCCCAAG 1360  
 S H T S V C A D C S A V F H R D C Y Y D N S T T C P K 430  
 TGTGCCCGGCTCAGCCTGAGGAAGCAGTCGCTCTTCCAGGAGCCAGGTCCCGATGTGGAGGCC **TAG**CGCCGAGGAACAGT 1440  
C A R L S L R K Q S L F Q E P G P D V E A \* 451  
 GCTGGGCACCCCGCTGGCCCGCAGGACCCACCCTGCCAACATCAAGTTGTTCTTCTGCTCCGGAGACCCCTGGGGTG 1520  
 CGGCCCTGGCCCCCTCCACCCCTGCTGGGCCAGAGCGGTGGCAGTGTCAAGGCCCGCTGTCTCCAGGTGCTTGCTGG 1600  
 GACTCGGGGCGGCTGCACCTGGCTGTACCTGGGTGTGCTGTGTGAGGGTCTTGTGCTGGCCCCCATCCTTCCCCCAA 1680  
 TGCAGAACTCCATGGGCAGGAGCTGGGGGACATCTCACCTCCCCATGGCACAGAGCCCTCCACACCCCTGGACCAGG 1760  
 GCATCCGGGCCCTAGAAATTCACAGCTCCCGTCTGGCCACCCTGGAAGCTCATCAGGCCAAGACCCGGACAGAGCTTC 1840

AGAGGAGTGTTGAGTGACACCTGAGGATGCGGCTGCACACACTCAGCCAAGGGCCGAGTCTCACCTGCGGTGGGGTTCG 1920  
GCTCTGCCTGGGGCTCCATCCCTTTCAGCCACTCGTGGCCTTGGGGATTCTGGTTGTCCCCAGCTGGGACTGTTCA 2000  
GTTGTACCTGCAGACCTGCCTCTCCCTGGCCTGAGGTTCAAAGGCCTCATCGGATGGTCAGTACAGTGGGGTACCTGT 2080  
TGTTTCTATACAACAGCAGGGAAGGGCCATGGAGCTTTTCCCTGCTGGGTGCTCCTGCTTTGGCCCAGCCACCTTCC 2160  
TGGTGCTCCAAGCTAGGAGGCTGTGGCCCCAGCCTGAGGAGGGTGTCTGGCCTCCAGGTGTGCAGCAGGGGCTGTGTGC 2240  
TGGGGGAGGTTCCAGTTAGGCGATGGGATCCTGCAGTGGTCTGGTGGCATTCTTGGAAACCAGATTTACCTGAGGAGCTC 2320  
TGTCTGCTCCCTGTGGAGGGCTCCAGATAGCTCAGAAATGACCAGCCAATGGCCTTTTGTGGGGGCTGAGGTCAAG 2400  
AGAGCTGAGAGTATTCGCTCGACTGAGCACATTAGGAAGATCAGGGCAGGCGTGTGGGAGGTCCCTCACTCCACGGGAC 2480  
AGAGGGCCCTGGACAGCAGAGGAAACCTACAGCTCTGGGTGAGGGGACACTTGGCTTTGGTGTGTTGCACTTTACAGATCC 2560  
TGGGTTCCACGAGGGCCCTCAGGAGAGGACGTGTGAGGACGTGGCTTCCCAGCCTTCTGCCTTGGGCAGTGGGGGTGCTC 2640  
CTGTCTGTCTTTTCCCCACACCCCTGGACTGTGCTTGGCTGTGGTGCACATGGTTGGCACACGGTGGGCAGAGGGCA 2720  
GAGAATGCCACTGCTTGGTTATTGGTCCCCTTTGACCAGGAAACCAAGAGGAGACACCTCAGTCAGCAGAAAGGCCACC 2800  
TGGCTCACTGGCTCATTCCAGGAGTGGGAGAGACGGCAGGGTCTCCTCTTTGTCTCCGGCATCAGGAAGGGGATGGTGT 2880  
CCACTCCCCACTGTGGTGGCTTTAGGCAAGGTTCTTATTGTCTGCTCTGCCTCGGTTTCCCCTATCTGGAAAATGGGGGCA 2960  
GGGGTCTGACCTACCTCAGGTGGAACGGTGAGCAGGGAACATGTGGAGTCTTTCAGAGAATGTGATGTGAGGTTGGAT 3040  
CAACAGTGTGGGTTCTGTCTGTCTTCCCCTTCTCTTTGGGGCTGAGGAGGAGGTTAAAGGCCAAATGCTGTTTCCCAA 3120  
CACCCCAAAGTCTGCACACGTCTCATGAATGCATCACATTTCTGTGATATGGATATTAGCCATTCCGAAATCTGTGTAAT 3200  
CAACTTCACATTAATCAAGTTACAAATCACTGTGTCCATAGAAAACTGTGCTGGTATTTGCTGGACAAAGGGTTGGGCC 3280  
CCTTTTATTTTACCTGCCACCCAGCATCTCCCCACATGCCCCCTTCTGGGTGACACAGCCGGTAAACGGAATCAACGTA 3360  
TGGTTCTTTCTGTGGGTCTGTGGCACAGCAGGAAGAGCCCGGTGCCGCCAGCACCTTGTGGAAGACCACACATGGGTGGT 3440  
CCCACAGCATGGGACCAGGCTGGCCTGAGGGATGCCAGTTGTAACAATGCTGCTGTCACTGTCTCATTAAATATACATC 3520  
CTTTA 3525



---

***Figure 3.7*** : Nucleotide and deduced amino acid sequence of *mt3*, the mouse orthologue of human *T3*. The nucleotide and amino acid positions are indicated by the numbers in the right hand column. The translation start site at ATG positions 58-60 and termination site at TAG positions 1402-1404 are highlighted. A polyadenylation signal beginning at nucleotide 3084 is underlined. The exon boundaries are indicated by the vertical lines with double-head arrows and the given numbers of exons.

gaactacagtgtagccagccgggtccgggtacaatgtagcggcagccaggtgggatgct**ATG**GAATATGATG 70  
 M E Y D 4

AGAAGCTGGTTCCGGTCCGGCAGGCCACCTCAACCCTTCAACAAGCAGCTTGGTCCGAGGCATCATGA 140  
 E K L V R F R Q A H L N P F N K Q L G P R H H E 28

ACAGGAACCCAGTGAGAAGGTCACCTTCTGAAGACACTTTGCCTGAGCTGCCCGCTGGGGAGCCTGAATC 210  
 Q E P S E K V T S E D T L P E L P A G E P E F 51

CACTACTCGGAGCGCATGATGGATCTCGGCCTGTCTGAGGACCACCTTTTCCCGCCCTGTGGGTCTCTTCC 280  
 H Y S E R M M D L G L S E D H F S R P V G L F 74

TGGCCTCTGATGTCCAGCAGCTGCCGACGGCCATCGAAGAATGCAAACAGGTGATCCTGGAGCTGCCCGA 350  
 L A S D V Q Q L R Q A I E E C K Q V I L E L P E 98

GCAGTCAGAGAAGCAGAAGGACGCTGTGGTGGCGGTGATCCACCTCCGGCTGAAGCTCCAGGAGCTGAAG 420  
 Q S E K Q K D A V V R L I H L R L K L Q E L K 121

GACCCCAATGAGGAGGAACCCAACATCCGGGTCTCTCTGGAACATCGCTTCTACAAGGAGAAGAGCAAGA 490  
 D P N E E E P N I R V L L E H R F Y K E K S K 144

GCGTTAAACAGACCTGCGATAAGTGAACACCATCATCTGGGGGCTCATCCAGACCTGGTACACCTGCAC 560  
 S V K Q T C D K C N T I I W G L I Q T W Y T C T 168

AGGGTGTGTTACCCTGTACAGCAAGTGCCTGAACCTCATCTCCAACCATGTGTGACGCTCCAAGGTC 630  
 G C C Y R C H S K C L N L I S K P C V S S K V 191

AGTCACCAGGCTGAGTATGAGCTGAACATCTGCCCTGAGACCGGGCTGGACAGCCAGGACTACCGCTGTG 700  
 S H Q A E Y E L N I C P E T G L D S Q D Y R C 214

CGGAGTCCCGTGTCTCCATCTCTGAGAGGTGTGCCTAGTGAGGCCCGGCAAGTGTGACTACACGGGCCA 770  
 A E C R A P I S L R G V P S E A R Q C D Y T G Q 238

ATACTACTGCAGCCACTGCCACTGGAACGACCTGGCTGTCTATCCCCGCGAGAGTGGTGCACAACCTGGGAC 840  
 Y Y C S H C H W N D L A V I P A R V V H N W D 261

TTTGAGCCACGCAAGGTGTCCCGTTCGAGCATGCCCTACCTGGCACTGATGGTGTCTCGGCCAGTGTCTCC 910  
 F E P R K V S R C S M R Y L A L M V S R P V L 284

GGCTCCGGGAGATTAACCCTTTGTCTGTTAACTACGTGGAAGAGCTGGTGGAGATCCGGAACTGGCCCA 980  
 R L R E I N P L L F N Y V E E L V E I R K L R Q 308

GGACATCCTGCTCATGAAGCCATACTTATCACCTGCAAGGAGGCCATGGAGGCGGACTACTGCTGCAG 1050  
 D I L L M K P Y F I T C K E A M E A R L L L Q 331

CTCCAAGACAGACAGCATTGTTGTGGAGAACGATGAGATGTACTCTATCCAGGACCTTTGGAAAGTGACA 1120  
 L Q D R Q H F V E N D E M Y S I Q D L L E V H 354

TGGGCCGCTCAGCTGCTCGCTCACTGAGATCCACACGCTCTTCGCCAAGCACATCAAGTTGGACTGTGA 1190  
 M G R L S C S L T E I H T L F A K H I K L D C E 378

GCGGTGCCAAGCCAAGGGTTCGTCTGTGAACTCTGCAAAGAAGGGGACGTGCTCTTCCCGTTCCGACAGC 1260  
 R C Q A K G F V C E L C K E G D V L F P F D S 401

CACACGTCTGTGTGCAATGACTGTTCCGCTGTCTTCCACAGGACTGTTACTACGACAACCTCGACCAGT 1330  
 H T S V C N D C S A V F H R D C Y Y D N S T T 424

GCCCAAGTGTGCCCGCTCACATTGAGGAAGCAGTCACTATTCCAGGAACCTGGCCTAGACATGGATGC 1400  
 C P K C A R L T L R K Q S L F Q E P G L D M D A 448

CTAGAACTGGAGAAGCCAGGCACCGTATCCACCCTCCGGCTGGTGCCTGCTGGCCTTGCCACCAG 1470

\*

ATGTGCATTCTACTCTGGAGACGACCCCCCCCCCCCAGTATATCCTCCAGACCTCTTCGTCTCGGGCCA 1540  
GAAGGAAGTGACTAGTGGCCACCGGGACTCATTCTCAGGTGCTTGTGGAGACTTCGAGTGTGTATACCT 1610  
GGCTGTTGATTGGGTGTGCTGCTATCGGGGGTCAAAAATACTGCTGTGCCCCATTGTAGGACTCCTTAG 1680  
GCAGAGGAGCCTGGGTGTGACTCTCCAGGAAGTGGGGGAGTGCCTCTCTTATGGCACAGAGCCCTCCAG 1750  
GCCTTTAAAACCAATCCTTCTGCACCATAGAGATGGCCTTGCTCCTGTCTAGTACACCCTTACAACCTGGT 1820  
CAGAACTTGAGAGGCCAGTCCAGTAAGACCAGAGGATGCCTCCTGACTGAGTTTTATCTCTGTGAGGA 1890  
AGATTTAGGTTCTACTCAGAGGCACCCAACCCTTTACACCTTGTCTCCTGAGGGACGTCCGGTTTTCT 1960  
CTGACTGGGGCTTTCACTACTGTGTCCCGTGGATCTTCCCCTGCATCCTGAGAGTTGTCATCGTCATCC 2030  
TTAATCTGTGCTCAGTACGTGTGCTTCCCCACCCCTTTGTACAATAGGGAGGGGAAGGGTCTGTGGAAC 2100  
TCCTTGGCAGATACGCCTGCTTCTGCCACAGCCTAGTCCACTTTATCCTGGTGTCTGAGCCGGGCAGC 2170  
TAGGAACCCACCCATAAGAGGCTTAGCATCCAGTTCATAGGTGGGACTGTGTCCGGGGAGGAGGTTCTA 2240  
GGTCAGGTGACCTGTGCTTATCATTGGCCTCGGTGGCATTCCGGGGATGAAGCTTTCCACAGAGGTC 2310  
ACTGTCTGTGACGCAGTGCCTGCTTGAAGATGACCAGTCCGTGGGTGTTAACGTAGGAGCCTGAGGT 2380  
CAAGAGCTAGGAATATTTGCTAGACCCACAGCAGGAGGCCACGGCCATACAGGAAGGTCTAAGTTTTG 2450  
AAGTGGGTGGGGCTACACCTTCCACCTCCATGCACACACCCATAAATCCTACGATGGAAGATGTGT 2520  
GGAGACGAGGGGAATCTCACAATAATGTGGCTCCTGTCTTTCTTCCCTGGGAGGCAGGAGTTGGTTGTG 2590  
TGCCCTTTGCTCCTTGTCTGTGAATTGCTGTTGACACACACAACTAGTGGGCAAAGGCCAGGGCATGTC 2660  
ACAACCTGGATTTCCCTTACCTTTAATGAGGCACAAGAAGGGACTTCTCAACTGGCCAGTTGAAGCTTAG 2730  
CTTACTGGCGTGTGGGGGAAGGACGGACACACTGCCACCTCCTCCTCACTGCTCTTTGCCATCAGGTG 2800  
GCTTTACTCATGCTGTTTTCTGTTTCAGTTTTCCCATCTGGAAGCATGGGGAGGGGGAGTTGACCTG 2870  
CCTCAGGGTGAAGATCAGGAGTCCCTTTGAGTGTGATGTGGGTCAACAATGGGTCCCTGTCCGCTTCT 2940  
CTTTGGCGTTAAAGAGGGTCTTAAAACCAATGCCATTTCCCATACCCCAAATCTGTGTGTTTCTGCC 3010  
ATTGATCACATGCTGTCTGAGTCTTGGCCCTTCTGAGCCTCCTGTGTAATCAACATCACGTTTCC 3080  
AACTTTAAATCATAGTGTCTAAAGGAAAAAAAAAAAAAAAA 3120

"TATAAA" between nucleotides 3088 and 3093 that is 16 bp upstream from the polyadenylation site. Homology between mouse [*mt3(def-8)*] and human (*T3-L*) transcripts is limited to the ORF sequence, no homology is observed between the sequences of either the 5' or the 3' UTR.

### 3.3.6 Protein Sequence and Function Analysis

A high degree of homology between the human (*T3-L*) and the mouse (*mt3*) transcripts, 86% identity between the ORF sequences and 95% identity between the predicted proteins, designated "T3-Lp" and "mt3p" respectively, suggest a functional conservation. The computerised protein localization prediction tool, PSORT, suggests that the T3-Lp/mt3p protein localizes in the nuclear compartment. Protein function was also estimated, using three functional motif-prediction programs, Profile Scan, Pfam, and SMART. Two functional motifs, a diacylglycerol/phorbol ester binding domain (**DAG/PE-bind**) and a plant homeodomain-like (**PHD**) zinc finger motif, were identified at a significant expectation (E-value) by all three programs. The DAG/PE-bind was located between amino acid residues 139 and 189 in T3-Lp, and between residues 136 and 186 in mt3p (*Figure 3.6* and *3.8*). The PHD domain was located close to the C-terminus between residues 389 and 432, and between residues 386 and 429, in T3-Lp and mt3p respectively (*Figure 3.6* and *3.8*).

Protein database searches, using BLASTP, identified the T3-L/mt3 amino acid sequence similarity to a number of proteins. The most significant similarity was to CG11534 gene product, a novel protein of 492 amino acids (AE003542) with unknown function, from *Drosophila melanogaster*. Sequence similarity between T3-Lp/mt3p and AE003542 were 37% identity and 57% positive (conservation of residues with similar physico-chemical properties). AE003542 also contains both a DAG/PE-binding domain at

---

**Figure 3.8** : T3 protein similarity. Multiple sequence alignment of the human (T3-Lp) and mouse (mt3p) protein, and novel proteins from *Drosophila melanogaster* (AE003542), and *Caenorhabditis elegans* (AF002197 and Y51H1A). Identical amino acids are highlighted in *blue*. Gaps are introduced to optimize alignment and are indicated by broken line. Scale bars above the aligned sequences indicate the order of characters as a whole. Numbers on the left-hand column denote the amino acid positions. Two functional domains, the diacylglycerol/phorbol ester binding domain (DAG/PE-bind) and the plant homeodomain-like zinc finger (PHD), are indicated with respect to T3-Lp between residues 139-189 and 389-432 respectively (▼: denote the consensus Cysteine, **C**, and Histidine, **H**, residues within these two domains). The consensus configuration of DAG/PE-bind is H-CxxC-CxxC-HxxC-C, and that of PHD finger is CxxC-CxxC-HxxC-CxxC, where “xx” indicates two residues, and “-” indicates various length of amino acids. A second region of similarity of 93 amino acids is also indicated (▶ ◀), between residues 204-296 of human protein, T3-Lp, and the homologous region of other proteins.

```

10      20      30      40      50      60
1 ME - T D K L A R F R Q A H L N P P F N K Q S G P R Q H E Q G P - G E V P D V T P E E A L B E L P P G E P E F S C P E
1 ME - D K K I V R F R Q A H L N P P F N K Q L C P R H H E Q E P - S E K V - - - T S E D T L P P L P A G F P E E H Y S K
1 M S S W K D S H T S I P G T V A Q L I N E S A S N L L H A S S T I C S T V G L G G S G S T G S G S E A G G S E S G P Q
1 M S S W - - - - - C S S E D A H N Q S S T P S T - R S R K S S E W L I S K I D R N E I D K E V Q M M D - - - D W R R S N
1 M N Y T F Y A R Y R Y R T D T I L - - - - - G T S L R D E L L E A L P S T H N D T F F N C T P I Y V Q S P

70      80      90      100     110     120
59 - - - R V M D I G L S E D H F S R P V - G D P L A S D V Q Q L R Q A I S S C K Q V I L E L P E Q S E K K Q K O A V V R L
56 - - - R M M D I G L S E D H F S R P V - G D P L A S D V Q Q L R Q A I S S C K Q V I L E L P E Q S E K K Q K O A V V R L
61 C A E Y R A L P I P A S I V R E Q W R L - I E T S D A N I G D L I Q A A T A H Q R D L V L L S E E L S E E R R W L V R H L
52 E A E N H V E C I K A A L E Y C R D K L K E L I D E E D T E M D G K E T K T P I N I V S N T E F P E D R L V L G N H
49 S - - - - L M G N S L A A I G A R H Q M - - - - W E S T S D R L S S S S D S G G G P V V S F G Q A - - - - -

130     140     150     160     170     180
114 I H L R I K L Q E L K D P N E D E P - - - - - N I R V L L S H R F Y K K K S K S V K Q T - - - - -
111 I H L R I K L Q E L K D P N E E E E - - - - - N I R V L L S H R F Y K K K S K S V K Q T - - - - -
120 V D L E Y S H Q E L E E A Q E Q H S L S S D M V V M N A L S A V V G H H E V P H H P H G R N R L Q A V - - - - -
112 V N - - - C L D R Q Q K A A R Q Q M Y F D K I V Q L G L E R Q A L Q D E A V V S T S S P L D E C R A E G H E P V M Q P V
90 - - - S A M E T R D E S L G T S S Q D L A D V D E - - - - - Q D V A D S S R K S V S A - - - - -

190     200     210     220     230     240
153 - - - - - C D K N T I I W G L I Q T W Y T C T G C Y Y R C H S K C L N L I S K P C V S S K V S H Q A E Y E L N - - -
150 - - - - - C D K N T I I W G L I Q T W Y T C T G C Y Y R C H S K C L N L I S K P C V S S K V S H Q A E Y E L N - - -
172 - A A K R N Y C D H C T I I W S V V Q N S V V C S D G G F L V H Q K C I D G V K R V C A H V L V S E R Q H P I S E - - -
169 R G G H N P C C E V C M H T I W R L V Q W W R R C G G M R A H D K C A E E V K R V C A - G V L S T R A K F S N T N
130 - E T A E K L C T - - - - -

250     260     270     280     290     300
204 I C P E T G L D S Q D Y R C A E C R A P I S L R G V P S - E A R Q C O D Y T G Q Y Y C S H C H W N D L A V - I P A R V V H
201 I C P E T G L D S Q D Y R C A E C R A P I S L R G V P S - E A R Q C O D Y T G Q Y Y C S H C H W N D L A V - I P A R V V H
229 I C B E I G L A S Q G Y K C A E C G T M L N I K N T W I - E P R L C O Y S G L Y Y C P R C N W N D S N F - I P A R I I H
228 L C E E R S L A E Q E Y Q C A E C T A P I C F D G Y A E Q E A R L O D Y S G E L F C P N C R W N D V W - S I P A R I V H
138 I P R E K G L D A Q D F R C A M C R K T I G - G S T F S - K F E T C A I D S K Y Y C T E C M K S G G K V S I P A R V V M

310     320     330     340     350     360
262 N W D F E P R K V S R C S M R Y L A L M V S R P V L R L R K I N P L L E S Y V E L V S I R K L R Q D I L L M K P Y F I
259 N W D F E P R K V S R C S M R Y L A L M V S R P V L R L R K I N P L L E S Y V E L V S I R K L R Q D I L L M K P Y F I
287 N W D F E P R K V S R T A L Q E I R F L N K P L I R L E D N E K L E N Y V E K L C A V K K L R Q N L V H M R H Y L A
287 N L I S T P R P V C H A V K L L L A I I D H R P L I D I N E S T L S H I K P H K E - - - - -
196 D W D W R E R A V S D R G R A W Y E A N Q E K A L I N I K T T N S R H Y A H A P A L E T R K L R E K Q V S M Y F

370     380     390     400     410     420
322 T C R E A M E A R L L - L Q L Q D R Q H F V N D E M Y S V Q D L I D V H A G R L G C S L T G Y A T T S A K H K L O C
319 Y C R E A M E A R L L - L Q L Q D R Q H F V N D E M Y S I Q D L I E V H M G R L S C S L T G Y A T T S A K H K L O C
347 A C N I A S E L K L V D Q Q L G V R R H L A Q S N E F Y S L S D L S Q W E S C A S E F E Q G V F K A E N D E R S C
328 - - - - -
256 T C R E S V S E D F - R R R L W P K E Y L R S E I D V Y S F A D L I C V K S G A L Q R R N S L L K H S I N V - M T C

440     450     460     470     480
381 E R C Q A K G F V C B L C R E - - - - - G D V L R F P P D S H T S - V C A D C S A V F R D C Y Y D N S T T C P K C A R
378 E R C Q A K G F V C B L C K E - - - - - G D V L R F P P D S H T S - V C N D C S A V F R D G Y Y D N S T T C P K C A R
406 P M C L A Q A Y I C S I G S N - - - - - N E V I F P P E D G C I - K C D Q C N S I F H R V C L T R K N M I C P K C I R
329 L T C A G N G E F C E L C D D I N V D Q K N K I M Y P P - T E N T R S C A T C L A V Y H K K C F E R K S L N C P R C E R
314 T L C K Q K G E C E L C T V - - - - - N E V I Y P E N T S T H R C L V C F S A F H V E C W - R T S G D C P K C V -

490     500     510
434 L S L R K - - - - - Q S L F Q E P G P S V E A 451
431 L T E R K - - - - - Q S L F Q E P G H D M D A 448
459 I Q E R R L L Q L D R M X S T E D D C D D D G V A T D D D V T A A E 492
388 - - - R R K R T E I P K T L S S S C E K E - - - - - A V R N 409
366 - - - R R Q N F S T R R A - - Q V D D P H N T - - - - - V L Q P 389

```

T3-Lp (451 aa)  
Mt3p (448 aa)  
AE003542 (CG11534 gene product, 492 aa)  
AF002197 (409 aa)  
Y51H1A.1 (389 aa)

T3-Lp (Human)  
mT3p (mouse)  
AE003542 (Drosophila)  
AF002197 (DAG/PE; C. elegans)  
Y51H1A.1 (C.elegans)

T3-Lp (Human)  
mT3p (mouse)  
AE003542 (Drosophila)  
AF002197 (DAG/PE; C. elegans)  
Y51H1A.1 (C.elegans)

T3-Lp (Human)  
mT3p (mouse)  
AE003542 (Drosophila)  
AF002197 (DAG/PE; C. elegans)  
Y51H1A.1 (C.elegans)

T3-Lp (Human)  
mT3p (mouse)  
AE003542 (Drosophila)  
AF002197 (DAG/PE; C. elegans)  
Y51H1A.1 (C.elegans)

T3-Lp (Human)  
mT3p (mouse)  
AE003542 (Drosophila)  
AF002197 (DAG/PE; C. elegans)  
Y51H1A.1 (C.elegans)

T3-Lp (Human)  
mT3p (mouse)  
AE003542 (Drosophila)  
AF002197 (DAG/PE; C. elegans)  
Y51H1A.1 (C.elegans)

T3-Lp (Human)  
mT3p (mouse)  
AE003542 (Drosophila)  
AF002197 (DAG/PE; C. elegans)  
Y51H1A.1 (C.elegans)

T3-Lp (Human, 451aa)  
mT3p (mouse, 448 aa)  
AE003542 (Drosophila, 492 aa)  
AF002197 (DAG/PE; C. elegans, 409 aa)  
Y51H1A.1 (C.elegans, 389 aa)

its central region and a PHD motif close to the C-terminus, similar to that of T3-Lp/mt3p. Similarity was also observed in the region of 93 amino acids close to the C-terminal part of DAG/PE binding domain in all proteins (**Figure 3.8**), with ~55% identity. These similarities suggest that AE003542 is likely to be a T3-Lp homologue. A protein AF002197 from *Caenorhabditis elegans*, containing a region similar to protein kinase C, was also similar to T3-Lp/mt3p with 33% identity and 53% positive. AF002197 also shows three domains: a DAG/PE-bind, a conserved 93 amino acid region, and a C-terminus PHD motif, found in T3-Lp/mt3p. The function of either AE003542 or AF002197 is unknown. However, the high degree of conservation among these proteins throughout evolution suggests a critical and basic cellular function.

### 3.3.7 Genomic Organization and Alternative Splicing

A complete genomic structure of *T3*, with respect to *T3-L*, was obtained from the alignment of the transcribed sequence of *T3-L* with the genomic sequence from cosmid 301F3. Overall, *T3* is organized into 12 exons encompassing a genomic sequence of approximately 20 kb (**Figure 3.4**). All splice junctions of exon/intron boundaries (**Table 3.3**) obey the consensus sequence for donor and acceptor splicing sites (Shapiro and Senapathy, 1987). The transcript of *T3-L* shares exons 1-4 with that of *T3-S*, and its exon 5 derives from the first 142 bp 5' part of *T3-S* exon 5. At this point there is a splicing donor site which is utilized for *T3-L* transcript (**Figure 3.2**). The extended sequence utilized for exon 5 of *T3-S* is part of intron 5 with respect to *T3-L* transcript. This sequence includes the continuation of ORF and the 3' UTR of *T3-S* (**Figure 3.2**). The DAG/PE binding domain is encoded by exons 5 and 6, whereas exons 11 and 12 codes for a PHD motif. Exon 12 is the largest exon spanning 2,202 bp of the genomic sequence. It contains part of ORF, stop codon (TGA), and a large 3' UTR (2,103 bp) (**Figure 3.6**).

To compare the gene structure of *mt3* with that of *T3*, the transcribed sequence of *mt3* (3,106 bp) was used to search for a mouse genomic sequence from a high throughput genomic sequence (htgs) database. A draft murine genomic sequence, AC023836 (as of October 2000), was retrieved and aligned with *mt3* cDNA. The alignment result indicated that *mt3* gene was also separated into 12 exons, with all the exon/intron junction boundaries also followed the gt/ag rule and surrounding consensus sequence of splicing site (Shapiro and Senapathy, 1987). When compared with human gene (*T3*), the mouse (*mt3*) exons appeared to use the equivalent boundaries with that of the human orthologue. The *mt3* exon 12 is also the largest exon, containing part of ORF, stop codon, and 3' UTR of 1705 bp (**Figure 3.7**).

In addition to two characterized transcriptional isoforms of *T3*, the *T3-S* and *T3-L*, the evidence of alternative splicing of *T3* was also observed in the 5' region. Initially in the transcript in at least three ESTs, AA035322, AA306508, and HSU46300, an extra exon of 97 bp was detected between exon 1 and exon 2. This extra exon, assigned as exon 1a, located within the first intron, 484 bp down stream from exon 1, possessed consensus splice donor and acceptor sites (**Figure 3.9C**). Inclusion of exon 1a was also observed in part of the 5' RACE products amplified from mammary gland and T-cell leukemia cell line (Jurkat) mRNA (**Figure 3.9**). Furthermore, some of these products also included an extra sequence of 125 bp, between exon 1a and exon 2. The location of this extra sequence in genomic DNA appeared to be the sequence that immediately extends from exon 1a (**Figure 3.9C**), which also showed the splice donor signal at its 3' boundary. Such alternative splicing patterns were also exhibited in the RT-PCR experiment on mRNA derived from breast cancer cell lines, normal mammary gland and fetal brain (**Figure 3.10B**), using 5' primer (35521-P) against primer from exon 4 (35521-5). Inclusion of exon

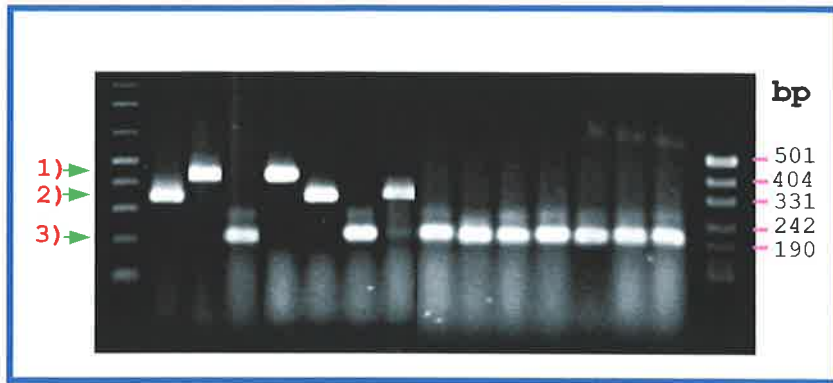
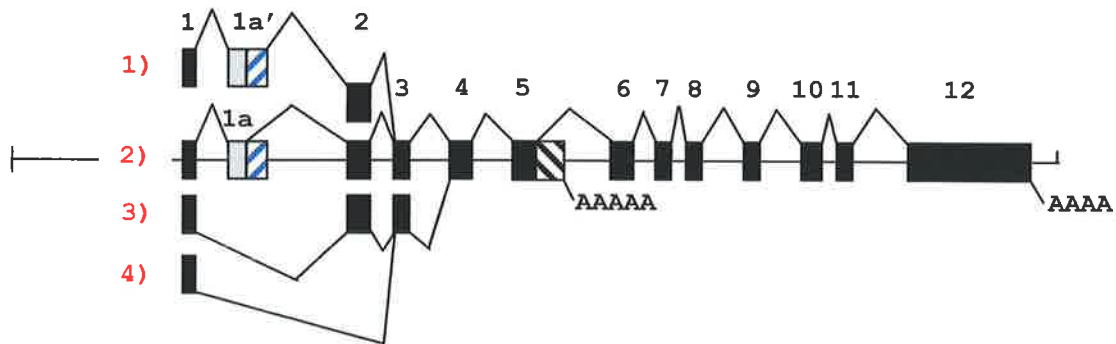


***Table 3.3***  
**Alignment of the T3 Gene Exon and Intron boundaries**

<i>Exon</i>	<i>3' Splice site (intron/exon)</i>	<i>Exon size (bp)</i>	<i>5' Splice site (exon/intron)</i>	<i>Intron size (kb)</i>
1	<b>5' UTR</b> -GACAGATGCGAGG	60	CGGCGGTCAG <b>gt</b> gagcggcg	5.260
2	tggccct <b>ag</b> GTGGGATGCT	134	ACTCCTGAAG <b>gt</b> gggtgctg	0.792
3	cttcctcc <b>ag</b> AGGCCCTGCC	98	CCGCCCTGTG <b>gt</b> aaggtttt	2.244
4	acctgggc <b>ag</b> GGTCTGTTCC	150	GGAGCTGAAG <b>gt</b> gggtgtgg	1.353
5	ccctctgc <b>ag</b> GACCCCAATG	142	ACCTGCACAG <b>gt</b> gggcccag	1.775
6	tccaccct <b>ag</b> GGTGTTATTA	165	ATCTCTCTGC <b>gt</b> gagtggtg	0.642
7	gaccctgc <b>ag</b> GGGGTGTGCC	128	GCCTCGAAAG <b>gt</b> gggggttg	0.147
8	tgccctggc <b>ag</b> GTTTCTCGCT	114	GGAGATTCGC <b>gt</b> gaggctgg	1.237
9	ttgtttgc <b>ag</b> AAGCTGCGCC	81	GCTGCTGCAG <b>gt</b> cagactgc	0.726
10	gacttggc <b>ag</b> CTCCAGGATC	141	GGACTGCGAG <b>gt</b> gggcctct	0.155
11	ttcccccc <b>ag</b> CGGTGCCAGG	110	TCTTCCACAG <b>gt</b> gggtgtgg	1.285
12	ctgcccc <b>ag</b> GGACTGCTAC	2,202	GGAGGCC <b>TAG</b> cgcc- <b>3' UTR</b>	-----
<b>Splicing consensus</b>	<u>ttctttncag</u> cctccc		AG <b>gt</b> aagt g	

---

**Figure 3.9** : 5' RACE and alternative splicing of the 5' region of *T3*. **A)** Agarose gel electrophoresis of the colony PCR analysis of the 5'RACE products (5'primer and 35521-U) that were cloned into the plasmid vector, pGEM-T. Each group of bands with the same size is indicated by a number with arrow on the left hand side of the gel. **B)** Exon organization of *T3* gene and the pattern of alternative splicing of the 5'region. Each splicing pattern corresponds to the 5' RACE product, shown in **A**, as indicated by the corresponding number 1, 2, or 3. Splicing pattern no. 4 with exon 2 skipping was detected by RT-PCR, the products of which shown in *Figure 3.10 B*, *band f*. **C)** The nucleotide sequences of *T3* exons 1-4, including exon 1a, and their flanking intron sequences. The boundaries of each exon are indicated by the consensus sequences for splice donor (**gt**)and acceptor (**ag**) sites. Exon 1a' is exon 1a that includes part of 3' flanking intronic sequence (underline), utilizing down stream splice donor site(**gt**). The location of gene specific primers used for 5' RACE (35521-U) and for RT-PCR analysis (35521-P and 35521-5) of the 5' region are also indicated.

**A****B****C****Exon 1 (ET7)**

35521-P

TAGCGGGGCGGCCGGGCGGATCCAGCGCAGCCGGGAGACAGATGCGAGGCGGCGGTTCAGgtgagcggcgcggggcccggcggggtcggggccggcggggacggggccggcggggacggggccggcggggacggggccggcggggacgggggtcggcggggtcggggctgggaggg

**Exon 1a**

ggtgggagagacactgccccgaggagccccgggagaccgggagctcccgcgggtgtggtgggtggggaggcacagtgtctgggtggcttctccctgagcagGTGCCAACCACGGCCAGGCTTCCGTGGCCAGCAGCCCTAGAGGAATGGCCATCCTGTCCCTGC  
GAGCCCTGGGCCCTGGCAGCGCATGCAGgtatggggcagacaggacgctgttgactccgcacaccgggggtgacttctcaggttc  
tcgggggtggcagctgcagtgatgacaccgcttctctggtggtcaGCCCGGGCAGGACCGGGAGGCCaggtccgtgcatcccg

Exon 1a'

**Exon 2 (ET6)**

agatcaaggctgtggtgtgaggactaccactttaaggaagtgaagaggccagcctcaccagacaccccagtggtggtggg  
gaaaggctggccctcagGTGGGATGCTATGGAATATGATGAGAAGCTGGCCCGTTCCGGCAGGCCACCTCAACCCCTTCAAC  
AAGCAGTCTGGGCCGAGACAGCATGAGCAGGGCCCTGGGGAGGAGGTCCCGGACGTCACCTCCTGAAGgtgggtgctggtgggag  
tcagggtgggagctgggagggtcTctgactgcttacgtggaccctccttctctctgcccgcgtctctgagcagccctggcttcc  
cc

35521-U

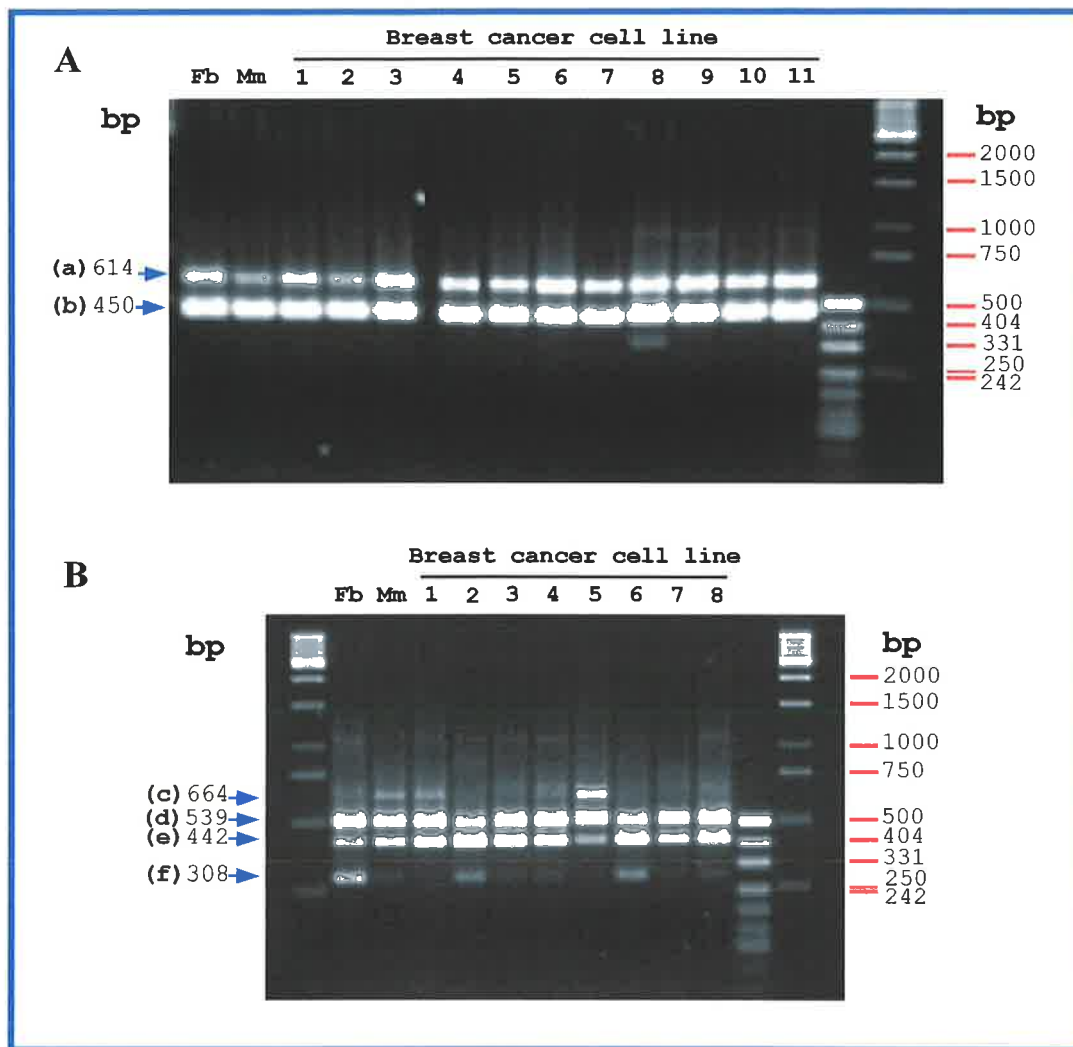
**Exon 3**

ctgaatccctttcccactcctggctatctcagctcctccctaggggatgggaggcaggaggtagcagctgacgtccacacctg  
tccccgtcttctccagAGGCCCTGCCTGAGCTGCCCCCTGGGGAGCCGGAATTCGCTGCCCTGAACCGGTGATGGATCTCGG  
CCTGTCTGAGGACCCTTCTCCCGCCCTGTGgttaagggttttagatctcggaggggagaggactgagggaaaccccaaggcagg  
aagggcCagggtttccttgtcactccctcccagggtggcagggcagtgctctgtgagctggggtaacagtc

**Exon 4**

tgagggtgtcggggctgctctgggctgtctcggcggatgggccttaacaggagctcctgcaggatgggcctttgactgcc  
cccggccccaaacctgggcagGGTCTGTTCCTGGCCTCTGACGCTCCAGCAGCTGCGGCAGGCGATCGAGGAGTGAAGCAGGTGA  
TTCTGGAGCTGCCCCAGCAGTCCGAGAGAAGCAGAAGGATGCCGTGGTGGACTCATCCACCTCCGGCTGAAGCTCCAGGAGCTGA  
AGgtgggtgtggggccgccccgcctggggcagcctggggccgaggccagaactgtggaactaggaggaccctgccagcctctg  
gctctctctcaggcgggtggtctctc

35521-5



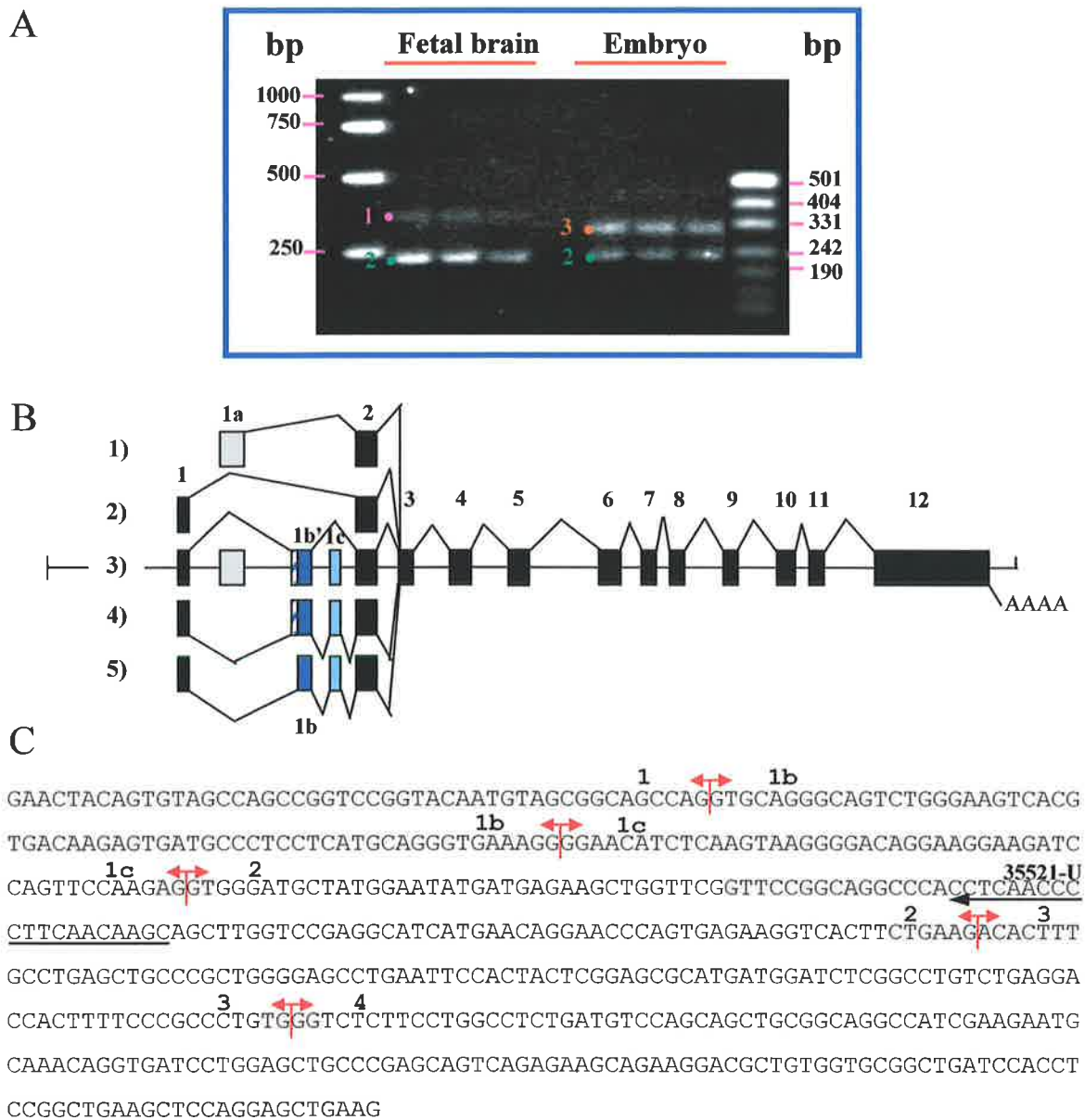
**Figure 3.10** : Agarose gel pattern of RT-PCR analysis of *T3* transcript in breast cancer cell line. **A**) *T3* expression in breast cancer cell lines as compared with those in normal mammary gland (**Mm**), and fetal brain (**Fb**). *T3* cDNA was co-amplified with *esterase D* utilizing two sets of primers in the same reaction: **band a**, *T3* primers, 35521-1 and T3EA-R (*Table 3.1*); and **band b**, primers for *esterase D* as internal control. **B**) RT-PCR products of the 5' region of *T3* using primers 35521-P (exon 1) and 35521-5 (exon 4) (*Figure 3.9C*, *Table 3.1*): **band c**, exons 1-1a'-2-3-4; **band d**, exons 1-1a-2-3-4; **band e**, exons 1-2-3-4; **band f**, exons 1-3-4. Size markers in **bp** are indicated on the right-hand side of each gel.

1a was observed in both *T3-S* and *T3-L* transcript in the RT-PCR product generated by using the exon 1a primer and either T3-S specific (35521-9) or T3-L specific (T3-EA-R) primer. Since the translation initiation codon was located within exon 2, all these splice forms maintained the same ORF.

Alternative splicing then generated a number of transcription variants of the existing transcription isoforms, T3-S and T3-L.

One interesting feature, the exon 2 skipping isoform, was observed in fetal brain and some breast cancer cell lines, but not in normal mammary gland (*Figure 3.10B*) nor fetal kidney. Skipping of exon 2 resulted in the removal of part of the transcript containing an ATG start codon at nucleotide positions 71-73. However, the sequence downstream from this skipped exon, has another inframe ATG codon within exon 3 at positions 251-253. The sequence immediately adjacent to this ATG codon conforms with the Kozak consensus for the translation start site. Therefore this provided a second start codon. The use of this second start codon will result in a protein that is 60 amino acids shorter from its N-terminus compared with the protein from the normal start codon. This protein product, however, still maintains the functional motifs, DAG/PE-bind (amino acids 139-189) and PHD zinc finger (amino acids 389-432). Since this alternative form of T3 was observed only in fetal brain and some breast cancer cell lines, it might represent a unique tissue-restricted isoform of this gene.

Alternative splicing of the 5' region was also observed in mt3, with the inclusion of various alternative exons between exons 1 and 2. As shown in *Figure 3.11*, inclusion of exons 1b and 1c were found in both EST AW763180 and 5' PCR products from mouse cDNA libraries. The use of exon 1a appeared to be independent of exon 1 and likely to be specific to fetal brain, since such product was not detected in a mouse embryo cDNA library. Utilization of exon 1a was also seen in MMDEF 85, the 5' sequence of def-8.



**Figure 3.11** :Alternative Splicing of the *mt3* 5' region. **A)** Agarose gel electrophoresis of PCR product amplified from mouse fetal brain and embryo cDNA libraries, utilizing vector primer (T7) and gene specific reverse primer (35521-U, exon 2). DNA bands were purified and sequenced: *band 1*, exons 1a-2; *band 2*, exons 1-2; *band 3*, exons 1-1b-2. DNA size markers in bp are indicated on both sides of the gel. **B)** Diagram showing exon organization of *mt3* and the pattern of 5' alternative splicing. Splicing pattern numbers 1, 2, and 3 correspond to the number of DNA bands shown in **A**. Splicing pattern numbers 4 and 5 were found in the partial cDNA sequences in dbEST database. **C)** Nucleotide sequence of *mt3* transcript from exons 1-4, including exons 1b and 1c. Also indicated is the location of primer 35521-U.

### 3.3.8 Expression Study of *T3*

#### 3.3.8.1 Differential Expression

The pattern of tissue distribution of *T3-L* mRNA was determined using two transcript specific probes constructed from nucleotides 586 to 1290 (*Figure 3.6*) to hybridize to a Human RNA Master Blot (Clontech). Expression of *T3-L* appeared to vary in different tissues and at different stages of development (*Figure 3.12*). High expression was observed in a series of tissues from the central nervous system, viz in cerebellum, occipital lobe, putamen, and substantia nigra. Intermediate expression was detected in the caudate nucleus, thalamus, and subthalamic nucleus, with relative low expression seen in cerebral cortex, frontal lobe, and hippocampus. Compared with heart, expression of *T3-L* was low in skeletal muscle, this may suggest the differential regulation in different types of muscular tissue. Low expression was also observed in aorta, colon, bladder, and uterus. Similar level of expression was seen in three hormone producing tissues: pituitary gland, adrenal gland, and thyroid gland. Among three lymphoid tissues, lower expression was shown in the thymus as compared with spleen and lymph node. Both mammary gland and prostate appeared to express *T3-L* at a similar low level. Compared with adult brain and heart, the corresponding fetal tissues showed a lower expression.

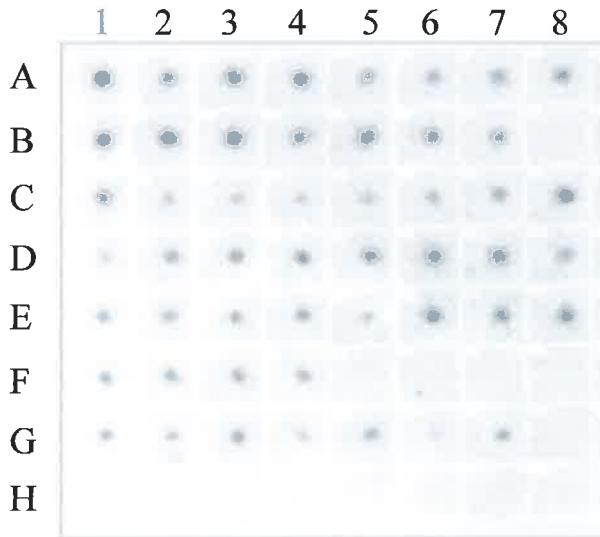
#### 3.3.8.2 Whole-body *in situ* hybridization of mouse embryo

The human (*T3*) and mouse (*mt3*) genes showed a high degree of homology of both the nucleotide sequence (ORF) and of the protein sequence. Study of gene expression at the early developmental stage can be determined in the mouse embryo, as an animal model. A mouse cDNA clone was isolated from a mouse fetal brain cDNA library, using a RT-PCR product of 432 bp, described in 3.3.4.2, as a screening probe. After the sequence

(i)

	1	2	3	4	5	6	7	8
A	whole brain	amygdala	caudate nucleus	cerebellum	cerebral cortex	frontal lobe	hippocampus	medulla oblongata
B	occipital lobe	putamen	substantia nigra	temporal lobe	thalamus	sub-thalamic nucleus	spinal cord	
C	heart	aorta	skeletal muscle	colon	bladder	uterus	prostate	stomach
D	testis	ovary	pancreas	pituitary gland	adrenal gland	thyroid gland	salivary gland	mammary gland
E	kidney	liver	small intestine	spleen	thymus	peripheral leukocyte	lymph node	bone marrow
F	appendix	lung	trachea	placenta				
G	fetal brain	fetal heart	fetal kidney	fetal liver	fetal spleen	fetal thymus	fetal lung	
H	yeast total RNA	yeast tRNA	E.coli rRNA	E.coli DNA	poly r(A)	Human Cot1 DNA	human DNA 100 ng	human DNA 500 ng

(ii)



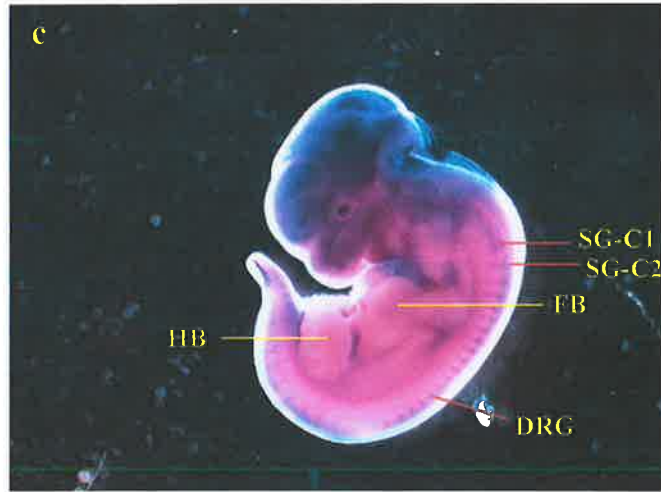
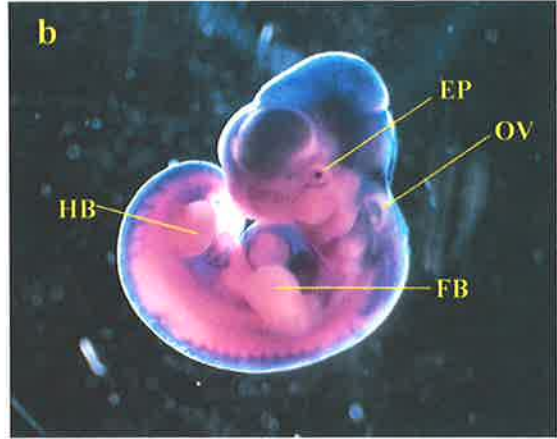
**Figure 3.12** : Differential Expression of T3. (i) Diagram of the tissue origin and position of human polyA<sup>+</sup> RNA dot-bloted on a nylon membrane. (ii) Hybridization with the [<sup>32</sup>P]dCTP-labelled T3 cDNA probe derived from the region between exons 6 and 11 (nt 586-1290). The amount of polyA<sup>+</sup> RNA in each dot was normalized by manufacturer using eight different housekeeping gene cDNA probes (Clontech)



---

**Figure 3.13** *T3* expression in developing mouse embryos. Mouse embryos at approximately 10.5-11 days postcoitum (dpc) were hybridized with *mt3* RNA probes: **a, b, c**, *antisense* probe; **d, e**, *sense* probe. The embryos show *mt3* expression in the developing peripheral nervous system and otic vesicle (ov). The first (*SG-C1*) and second (*SG-C2*) cervical ganglia were taken as reference points (Spörle and Schughart 1997). EP, eye pit; FB, forelimb bud; HB, hindlimb bud; DRG, dorsal root ganglia. Expression is also seen in the structure probably of the developing spinal cord at the tail. The anatomical nomenclature of the mouse embryo is based on Bard *et al.*, (1998).

[*Note*, the references are given: Bard, J.B.L., Kaufman, M.H., Dubreuil, C., Brune, R.M., Burger, A., Baldock, R.A., and Davidson, D.R. (1998) An internet-accessible database of mouse developmental anatomy based on a systematic nomenclature. *Mechanism of Development* 74: 111-120.; Spörle, R., and Schughart, K. (1997) System to identify individual somites and their derivatives in the developing mouse embryo. *Dev. Dyn.* 210: 216-226.]



**Table 3.4**  
**Oligonucleotide primers used for SSCP analysis of T3 coding sequence**

<i>Exon</i>	<i>Sequences of oligonucleotide primers (5'→3')</i>	<i>Product size (bp)</i>
2	<i>F</i> aag tga aag agg cca gcc tc <i>R</i> cac gta agc agt cag aga cc	253
3	<i>F</i> agc tga cgc tcc aca cct g <i>R</i> gag tga caa gga aac cct gg	212
4	<i>F</i> agg gag ctc ctg cag gat g <i>R</i> ctc cta gtt cca cag ttc tg	268
5	<i>F</i> agc cca gct cac ctg tga c <i>R</i> tgg cca cca cac tgg ata g	205
6	<i>F</i> ctt ttg aga agt gtg tgt cc <i>R</i> tcc ctc acc ctc ttc tct c	230
7	<i>F</i> gag gct gga aag ctg tgt gg <i>R</i> caa ccc cct tcc tgc tct c	249
8	<i>F</i> tgg gtg tga gag gtg tca c <i>R</i> cag agg aga aga ggg tac tg	235
9	<i>F</i> ctg tgt cct ggc tga gga g <i>R</i> aaa cct gag cct gag gtg c	200
10	<i>F</i> gac cac tgc agc gac cat g <i>R</i> gct tgg gtt gga gaa gag c	263
11	<i>F</i> aga gga gct ggc act gac g <i>R</i> cca gag gac agc gca gag	218
12	<i>F</i> atg ata gga gct gac cta gg <i>R</i> gtg ccc agc act gtt cct c	211

identity was confirmed, a restriction digested fragment of 576 bp corresponding with the region between exons 3 and 7 was excised and cloned into the pBluescript vector. The same restricted digested product was also used to hybridize to a mouse multiple tissue Northern blot. The RNA probes, both sense and anti-sense strands, were generated from this cloned 576 bp cDNA and used to hybridize to the mouse embryos at stage between 10.0-11.5 days post coitus (dpc). As shown in **Figure 3.13**, the hybridized signal was seen only in embryo hybridized with anti-sense RNA probe but not with sense probe. Expression signal was seen in the structure likely to be neural tube, otic vesicle, and eye pit. Expression was also seen in developing structure of peripheral nervous system especially in peripheral nerve ganglia .

### 3.3.9 Mutation Analysis of T3 in Breast Cancer

Single strand conformational polymorphism (SSCP) analysis was used for a mutation screening of T3 in two panels of 24 paired normal and tumor DNA samples from breast cancer patients and 24 breast cancer cell lines. Primers were designed to allow the amplification of each exon that included at least 50 bp of its flanking introns (**Table 3.4**). The primers were fluorescent labelled, the PCR product of each exon was denatured and analysed using a SSCP gel electrophoresis. A mobility band shift of exon 4 was observed in a number of breast cancer samples and cell line DNAs. This band shift, however, was also detected in the paired normal sample of each tumor as well as 10 out of 50 normal control samples. It was therefore considered to be a DNA polymorphism, which was present in the general population. Sequence analysis of shifted-band PCR product of exon 4 reveals a single nucleotide polymorphism (C/T) in intron 3 (IVS3-18C→T). Polymorphisms, were also identified in exon 9 and exon 12, since they were seen in some tumors and their paired normal as well as tumor cell line and control DNAs. Sequence

analysis was performed only in exon 12 PCR products, from one cell line and one normal control, and revealed three single nucleotide polymorphism: IVS11-14C→G, IVS11-37C→T, and 1351T→G. Polymorphism at position 1351, although causing codon change, ACT→ACG, does not change the coding amino acid (T, threonine). From SSCP screening, no other changes were observed and therefore no mutations were detected which were restricted to these breast cancer samples and cancer cell lines.

In breast cancer, other mutation mechanisms may also be involved and cause the reduction of gene expression or errors in splicing and therefore exon skipping which may lead to the disruption of ORF. To test for these possibilities, RT-PCR was performed using poly-A RNA isolated from 12 breast cancer cell lines. The results were compared with those of two control RNAs, normal mammary gland and fetal brain. No reduction of mRNA expression was detected, in contrast expression of *T3-L* appeared to be up-regulated in more than half of the cell lines tested (*Figure 3.10 A*). The isoform where exon 2 was skipped was identified in three cell lines, but only as a fraction of all mRNA products. This isoform was not seen in normal mammary gland, but was detected in fetal brain as previously described (section 3.3.7, *Figure 3.10 B*), and was therefore considered to be an isoform specifically expressed in fetal brain and tumor tissues.

### **3.4 Discussion**

The transcribed sequences of a gene, designated "transcription unit 3 (*T3*)" have been successfully identified. Initially, two trapped exons, ET6 and ET7, were used as the starting materials for both database searches and cDNA library screening. The cDNA sequence of 808 bp was first obtained, completing with 5' UTR, ORF, and 3' UTR with polyadenylation signal. Based on the results of Northern analysis *T3* appeared to code for two transcription isoforms of 0.9 and 3.7 kb. The first corresponds to a 808-bp transcript

that was assigned as "*T3-S*". Subsequently, with the available genomic sequence encompassing this gene, cosmid 301F3 (*Figure 1.3*), the transcribed sequence of the second isoform of 3,525 bp, designated "*T3-L*", and its complete gene structure were identified.

The *T3* gene structure was initially determined for the short isoform, *T3-S*. Since at the early stage of this project there was no genomic sequence available for this region, this was determined by sequencing of genomic PCR amplification products and by partial sequencing of the genomic clone utilizing transcript sequence specific primers. Following the large-scale genomic sequencing of the region, the genomic organization of the entire gene was characterized. The *T3* gene consists of 12 exons encompassing the genomic sequence of approximately 20 kb. It was shown that ET7 was trapped from the 3' part of the exon 1 which constituted the 5' UTR. Nucleotide sequence analysis subsequently revealed a sequence similar to the consensus of a splicing acceptor at the 5' boundary of ET7. This mimic consensus sequence, in combination with the true exon/intron boundary at the 3' end of exon 1, therefore enabled the trap vector to capture the "exon", ET7.

*T3-S* isoform shares the first 4 exons and part of exon 5 with *T3-L*. The 3' terminal exon for *T3-S* is the exon 5 and the sequence that extends into intron 5 of *T3-L*. This sequence includes part of the extended ORF, 3' UTR, and a polyadenylation signal, ATTA AAA specific for *T3-S* (*Figure 3.2*). The presence of specific polyadenylation signal suggests that transcriptional process of *T3-S* may be independent from that of *T3-L*. More commonly a single primary transcript is generated and different isoforms are generated by alternative splicing of the internal exon or exons. In *T3*, the two isoforms are likely generated during transcription through the selective use of independent polyadenylation signal specific for each isoform. Transcription process by RNA polymerase would either utilize the first ATTA AAA signal (within intron 5) to generate the *T3-S* transcript or run

passing this region until the second signal, also ATTAAA, close to the end of exon 12 was transcribed to complete the *T3-L* transcript (*Figure 3.6*). According to the Northern analysis result, the two principle transcription isoforms of *T3* were ubiquitously expressed with varying degree of expression among the tissue tested. Interestingly, the difference in degree of expression was also observed between the two isoforms in individual tissues. Such variation between these two isoforms may also suggest independent transcriptional regulation.

Pre-mRNA 3' end selection has been reported for the gene that utilize multiple poly(A) signal similar to that of *T3*. The best characterized example is the regulation of immunoglobulin M (IgM) heavy chain synthesis during B-cell differentiation. In this process, switching from the membrane-bound form of IgM heavy chain ( $\mu_m$ ) to the secreted form ( $\mu_s$ ) involves the selective use of a downstream  $\mu_m$ -specific poly(A) site in B-cell and of an upstream  $\mu_s$ -specific poly(A) site in plasma cell respectively. Such selective use of poly(A) site appears to be modulated by the abundance of polyadenylation factors that participate in pre-mRNA 3' end processing complex (Edwards-Gilbert and Milcarek, 1995; Takagaki *et al.*, 1996).

Another mechanism utilizing the splicing factors also involves in the regulation of pre-mRNA processing to determine the 3' terminal exon. As has been shown during alternative processing of pre-mRNA from the human genes for calcitonin/calcitonin gene-related peptide (CT/CGRP) (Lou *et al.*, 1998) where processing choice involves tissue-specific recognition of alternative 3'-terminal exons, exon 4 or 6 (Amara *et al.*, 1982; Rosenfeld *et al.*, 1984). In thyroid C cells, exon 4 is recognized as 3'-terminal exon with the usage of its poly(A) site, and the resulting mRNA molecule, exon 1 to 4, produces calcitonin (CT) peptide. In neuronal cells, exon 4 is excluded and the mRNA molecule containing exons 1, 2, 3, 5, and 6 is generated utilizing exon 6 as the 3' terminal exon. The CGRP is produced

from this mRNA. It has also been shown that exon 4 recognition requires an RNA-processing enhancer sequence located within intron downstream of the exon 4 poly(A) site (Lou *et al.*, 1995; Lou *et al.*, 1996).

Selective expression of the transcripts *T3-S* or *T3-L* may be regulated at least by either of the above-described mechanisms, though requires further investigation, leading to the differential expression of these two isoforms seen between different tissue as well as different isoform within the same tissue. *Symiaf*

*T3* showed a transcribed sequence homology to a partial sequence (MMDEF85) of mouse *def-8* mRNA, a transcript isolated as showing developmentally regulated expression in haematopoietic cells. The complete mouse transcript homologous to the human *T3-L* was also identified by sequencing of the mouse cDNA clones and an overlapping set of RT-PCR products isolated from mouse cDNA libraries. This mouse transcribed sequence provided a resource for interspecies cross-referencing of the *T3* gene. Surprisingly, the result of Northern analysis of this mouse transcript demonstrated only one prominent product of 3.4 kb mRNA corresponding with its transcript sequence. There was no short isoform expressed from this mouse gene. The nucleotide sequence homology of the mouse transcript to *T3-L* is particularly striking by showing 87% identity across the ORF, with 95% identity at its predicted coding amino acid sequence. This mouse transcript was designated "*mt3*" referred to the mouse gene, orthologous to the human *T3* gene. The absence of the short transcript isoform in the mouse orthologue (*mt3*) may indicate the lack of enhancer sequence element downstream of *mt3* exon 5. This may further suggest the recent evolution of the gene sequence in *T3* that may contain enhancer element and facilitate 3' end polyadenylation processing of exon 5. As a result, *T3* gene codes for two protein isoforms of 197 amino acids for *T3-Sp* and of 451 amino acids for *T3-Lp*.



*In silico* functional analysis of the predicted T3 protein of 451 amino acids (T3-Lp) revealed at least two functional motifs, a DAG/PE binding domain in the central region and a PHD zinc finger motif at its C-terminus. In T3-Sp, most of the predicted amino acid sequence comprises part of the N-terminal region of T3-Lp. There is no known functional motif in this region.

The DAG/PE binding domain (PF00130) is a cysteine-rich motif about 50 amino acid residues long and essential for binding of diacylglycerol (DAG) and its analogue, phorbol ester (PE). DAG is an important second messenger in cell signaling, and its analogue (PE) appears to be potent tumor promoters (Castagna, 1987). This domain binds two zinc ions via the six cysteines and two histidines that are conserved and configured as HC<sub>4</sub>HC<sub>2</sub>. It is a conserved functional motif present in several proteins and is most likely involved in cellular signaling pathways. A family of serine/threonine protein kinases, collectively known as protein kinase C (PKC), is a classical kinase that contains one or two copies of this DAG/PE-binding domain, also known as the "C1 domain", at its N-terminal region. The C1 domain has been shown to bind DAG/PE in a phospholipid and zinc-dependent fashion (Ono *et al.*, 1989).

DAG/PE binding domain has also been found in other proteins either possessing or lacking kinase activity. These include diacylglycerol kinase (DGK): a protein that phosphorylates the second-messenger DAG to phosphatidic acid (PA), chimaerin: a protein family possessing GTPase-activating protein activity, raf/mil family of serine/threonine protein kinase, and Vav protein: a 95 kDa protooncogene product with its expression restricted to cells of hemopoietic origin. These proteins are involved in cell signaling pathways, and some are known to promote cell-cycle progression and migration (reviewed in van Blitterswijk and Houssa, 2000; reviewed in Kazanietz, 2000; Bonnefoy-Berard *et al.*, 1996)

The PHD domain, a zinc finger-like motif, is also a cysteine-rich domain binding to two zinc ions. This type of zinc finger has been identified in more than 40 proteins including those of human (Aasland *et al.*, 1995; Saha *et al.*, 1995; Stec *et al.*, 1998; Hasenpusch-Theil *et al.*, 1999; Lu *et al.*, 1999; Bochar *et al.*, 2000). It also consists of eight ligands with seven cysteines and one histidine, arranged in a unique C4HC3 pattern spanning approximately 50-80 residues. The PHD finger is distinct from other class of zinc finger motifs such as the RING finger and LIM domain. The PHD finger containing proteins are thought to belong to a diverse group of transcriptional regulators possibly affecting eukaryotic gene expression by influencing chromatin structure. Recent data provide evidence that PHD finger proteins are associated with chromatin remodelling complexes (Bochar *et al.*, 2000) or contribute to histone acetylation (Loewith *et al.*, 2000). Furthermore, the PHD finger has been shown to activate transcription in yeast, plant, and animal cells (Halbach *et al.*, 2000). The presence of PHD motif in T3-Lp/mt3p implicated its possible function involving in transcriptional regulation, therefore supported the notion that T3-Lp/mt3p is a nuclear protein.

The importance of the PHD finger is evident from the reported mutations in PHD fingers associated with human diseases and provides a good evidence for its functional role in particular proteins. In the ATRX syndrome, an X-linked form of syndromal mental retardation associated with alpha-thalassemia, missense mutations in the non-canonical PHD-like domain of the *ATRX* gene lead to the characteristic phenotype (Gibbons *et al.*, 1997). ATRX has been suggested to play a role in the de-repression or activation of chromatin, and the interaction between its mouse orthologue, Mo ATRX and mouse chromatin protein (HP1) has been demonstrated in a yeast two-hybrid system (Le Douarin *et al.*, 1996). Other examples of PHD finger-containing proteins involved in human disease include the *AIRE* gene product, which is mutated in the autoimmune disease APECED,

autoimmune polyendocrinopathy-candidiasis ectodermal dystrophy (Nagamine *et al.*, 1997; The Finnish-German APECED Consortium, 1997). The AIRE protein harbours a nuclear localization signal (NLS) and two PHD motifs. Most point mutations associated with deleterious effect are within or lead to the truncation of one of the PHD domains of AIRE (Scott *et al.*, 1998; Rinderle *et al.*, 1999). Furthermore, AIRE mutants lacking the PHD motif exhibits altered nuclear localization (Rinderle *et al.*, 1999), suggesting that AIRE mutations in APECED may have their primary effect in the nucleus. ALL1, the human homologue of *Drosophila trithorax*, and other PHD-finger containing genes, are also frequently disrupted by chromosomal rearrangements that occur in acute lymphoid and myeloid leukemias (Gu *et al.*, 1992; Tkachuk *et al.*, 1992; Nakamura *et al.*, 1993; Parry *et al.*, 1993).

The presence of both the DAG/PE binding domain and the PHD finger motif suggests that T3-Lp is involved in a cellular signalling pathway, most likely to promote cell growth. The most likely function is an involvement in transcription regulation, possibly promotion, by influencing chromatin structure. These suggestions are consistent with the computerized predictions of a nuclear localization as well as the observation that gene expression is ubiquitous in embryonic tissues, and all adult tissues. An importance basic function of the T3-L protein in cellular processes is supported by the conservation of the protein during evolution from *C elegans*, *D melanogaster*, to humans.

Study of *mt3* gene expression in the embryo of the mouse (section 3.3.8.2) further demonstrated the importance of this gene during embryonic development. It is likely that this protein is required for embryonic cell proliferation and differentiation.

Mutation screening of *T3* in breast cancer was performed for all 12 coding exons by SSCP analysis. No evidence of mutations restricted to the cancer cells were detected in two panels of breast cancer samples showing 16q24.3 LOH nor in breast cancer cell lines. This

indicated that *T3* was unlikely to be a tumor suppressor gene targeted by LOH at least in sporadic breast cancer tested. This is consistent with the predicted possible function of T3 protein (T3-Lp) for promotion rather than suppression of cell growth. Furthermore, it was observed that expression of *T3-L* was increased in some breast cancer cell lines.

Carcinogenesis requires at least two cellular events, the loss of cell proliferation control pathway and the activation of cell cycle pathway, although changes in other cellular pathways are also required for phenotypic changes into more aggressive forms of cancer. The possible function of *T3* can be categorised into a group that activates or promotes the cell cycle. T3 protein may therefore be a part of onco-protein network and interact with several proteins in the network. With this possibility, study of the associated proteins or the proteins that may interact with T3 is likely to lead to an understanding of the pathway that involves T3. Many biological techniques can be applied to this, for example, the yeast two-hybrid system (Fields and Song, 1989), mass-spectrophotometric analysis of native protein complex purified by affinity capture (Pandey *et al.*, 2000; Pandey and Mann, 2000), and protein array for multi-protein interaction analysis. These procedures, however are beyond the scope of this study.

## Chapter 4

Characterization of Transcription Unit

12 (T12) and

Mutation Study of Spastic Paraplegia

gene, *SPG7*

---

## Table of Contents

	Page
<b>4.1 Introduction</b>	118
<b>4.2 Methods</b>	120
4.2.1 Transcript sequence identification	120
4.2.1.1 Database searches for Expressed Sequence Homology	120
4.2.1.2 Identification of the cDNA sequence by RT-PCR and cDNA library Screening	120
4.2.1.3 Determination of the cDNA end	121
4.2.2 Expression Analysis	122
4.2.2.1 Northern blot Hybridization	122
4.2.2.2 Differential Expression Study	122
4.2.3 Protein Sequence Analysis	123
4.2.4 Identification of Genomic Structure	123
4.2.5 SPG7 Mutation Study	124
<b>4.3 Results</b>	124
4.3.1 Determination of Transcript Sequence	124
4.3.1.1 Database Analysis and cDNA Sequencing	125
4.3.1.2 Transcript Sequence Assembly	126
4.3.2 Transcript Sequence and Expression Analysis	130
4.3.3 Protein Sequence Identity	131
4.3.4 Genomic Organization	132
4.3.5 SPG7 mutation study	133
4.3.5.1 Nonsense Mutation in Exon 5	134
4.3.5.2 Compound Heterozygous Mutation in Exon 11	135
4.3.5.3 Missense Mutations in Exon 9 and Exon 13	135
<b>4.4 Discussion</b>	137

## 4.1 Introduction

In an attempt to search for a new gene that might be targeted by LOH and be a candidate for a breast cancer tumor suppressor gene, six trapped exons isolated from the 16q24.3 physical map (Whitmore *et al.*, 1998a) within close proximity were investigated to see if they were part of the same gene. These trapped products were grouped and from the order determined by the physical map (**Figure 1.3**) were described as transcription unit 12 (*T12*). *T12* was situated approximately 11 kb proximal to the *CMAR* gene. This chapter describes the determination of the transcript sequence based on these trapped exons. *T12* was found to be a transcript of approximately 3.2 kb. The predicted protein sequence was analyzed *in silico*. The likely function was a mitochondrial ATP dependent metalloprotease, based on its high degree of homology to this type of protein in yeast. At completion of this gene characterization, a report appeared (Casari *et al.*, 1998) which described an identical gene designated *SPG7*, which mapped to the region 16q24.3 and was involved in one form of familial neurodegenerative disorder, recessive hereditary spastic paraplegia (**HSP**). Although *SPG7* (*T12*) was not involved in breast cancer the involvement of this gene in HSP was studied in more detail.

Familial, or hereditary spastic paraplegia (HSP), comprises a group of neurodegenerative disorders characterized by slow progressive spasticity and weakness of the lower extremities (Harding 1981; Fink 1997; Reid 1999a). The common neuropathologic finding is axonal degeneration involving the terminal ends of the longest fibers of the corticospinal tracts and dorsal columns. The diseases are clinically and genetically heterogeneous. Two forms of HSP are classified according to the clinical symptoms. In pure (uncomplicated) HSP the spasticity is the prominent feature and occurs in isolation. Most patients present with difficulty walking or gait disturbance, noticed either by themselves or relative. In those with childhood onset, a delay in walking is

common. In the complicated form spastic paraparesis has been associated with many conditions and there are additional neurologic symptoms, such as cerebellar ataxia, mental retardation, dementia, epilepsy, optic atrophy and retinal degeneration (Fink 1997; Reid 1999a). In both clinical forms, autosomal dominant, autosomal recessive and X-linked inheritance have been described. As well as genetic heterogeneity there is often variable penetrance and expressivity. Linkage analysis has identified at least eight genetic loci for the autosomal dominant forms of HSP on chromosomes 14q (*SPG3*) (Hazan *et al.*, 1993), 2p (*SPG4*) (Hazan *et al.*, 1994; Hentati *et al.*, 1994b), 15q (*SPG6*) (Fink *et al.*, 1995), 8q (*SPG8*) (Hedera *et al.*, 1999), 10q (*SPG9*) (Seri *et al.*, 1999), 12q (*SPG10*) (Reid *et al.*, 1999b), 19q (*SPG12*) (Reid *et al.*, 2000), and 2q (*SPG13*) (Fontaine *et al.*, 2000). Four loci associated with autosomal recessive HSPs were mapped to chromosome arms 8p (*SPG5*) (Hentati *et al.*, 1994a), 16q (*SPG7*) (De Michele *et al.*, 1998), 15q (*SPG11*) (Murillo *et al.*, 1999), and 3q (*SPG14*) (Vazza *et al.*, 2000). Two X-linked forms of HSP, both complicated, have been described and are caused by mutation in the L1 cell adhesion molecule (L1CAM) and proteolipid protein (PLP) genes located on regions Xq28 (*SPG1*) and Xq22 (*SPG2*) respectively (Jouet *et al.*, 1994; Saugier-Veber *et al.*, 1994).

Of the 12 autosomal loci, the genes for *SPG4* and *SPG7* have been characterized. *SPG4*, the gene responsible for autosomal dominant HSP, encodes a predicted 67.2 kD nuclear protein known as "spastin" (Hazan *et al.*, 1999). The spastin amino acid sequence shows a high degree of homology with the yeast 26S proteasome subunits, Yta6p and Tbp6p, which belong to a subclass of the AAA protein family (ATPases associated with diverse cellular activities) (Patel and Latterich, 1998). *SPG7*, a gene at 16q24.3 involved in recessive HSP, has also been characterized (Casari *et al.*, 1998). *SPG7* encodes a mitochondrial protein called "paraplegin" which is highly homologous to the yeast mitochondrial ATP-dependent zinc metalloproteases, Afg3p, Rca1p and Yme1p. These



yeast proteins exhibit both proteolytic and chaperone-like functions at the inner mitochondrial membrane and also belong to the AAA protein family (Patel and Latterich, 1998). *SPG7* mutations have been reported in three families, two with pure and one with complicated autosomal recessive HSP (Casari *et al.*, 1998). Mutations in paraplegin cause a mitochondria function defect as a result of impaired oxidative phosphorylation (OXPHOS).

The publication of Casari *et al.* (1998) that reported *SPG7* did not provide the complete gene structure and it was not available in the public sequence database. The genomic structure is necessary as a basis for determining gene mutations associated with a particular genetic syndrome. This chapter will first describe the cloning and characterization of what was called the *transcription unit 12 (T12)*. This will be followed by the determination of the genomic structure. Finally, mutation analysis of *SPG7* will be explored in a large number of spastic paraplegia cases, both pure and complicated forms.

## **4.2 Methods**

### **4.2.1 Transcript Sequence Identification**

#### ***4.2.1.1 Database searches for Expressed Sequence Homology***

Trapped exon sequences were first used to identify the homologous transcribed sequences from both EST and non-redundant databases, using BLASTN program (Altschul *et al.*, 1997) available at NCBI. During the transcript sequence identification was underway, homology searches was also performed on both EST and non-redundant database utilizing the partial transcript sequence available at that time.

#### ***4.2.1.2 Identification of the cDNA equence by RT-PCR and cDNA library Screening***

RT-PCR was initially undertaken to determine the cDNA sequence linking the trapped exons within the close proximity. The procedure described in section 2.2.13 was carried out utilizing a set of primers designed to the trapped exon sequences (**Table 4.1**). The amplified products were subsequently purified by QIAquick columns (Qiagen, section 2.2.14.1) or separated by agarose gel electrophoresis and a single DNA band was purified from the gel (2.2.14.2). The purified cDNA fragments were then sequenced (2.2.16.1) using the same set of primers that were used for RT-PCR. The sequence information was then assembled.

Screening of the cDNA library was performed utilizing the trapped exon insert as probe. Trapped product ET2.5 was excised from the pAMP10 vector by double digestion with *Not I/Sal I*. The insert was  $\alpha P^{32}$ dCTP-labeled and use as probe to hybridize to the  $\lambda$ -fetal brain cDNA library filters (2.2.18). After repeat hybridization for plaque purification the positive clones were eluted from the phage plaques for subsequent analysis. Sequence of the cDNA insert was then determined by PCR amplification using the  $\lambda$ gt11 vector primers, cDNA fragment purification and nucleotide sequencing.

PCR procedure was also use to screen the fetal brain cDNA library for the 5' sequence of cDNA utilizing the vector primer and gene specific primer. This was performed in order to extend the 5' sequence of the existing cDNA obtained from RT-PCR and sequence assembly.

#### **4.2.1.3 Determination of the cDNA end by RACE**

Determination of the 5' end or the sequence toward the 5' end of the transcript was performed utilizing the 5' RACE procedure described in section 2.2.17.1. The nucleotide sequences of the gene specific primers used for 5' RACE are presented in **Table 4.1**

The 3' RACE procedure (2.2.17.2) was used to determine and extend the 3' sequence of the transcript. First strand cDNA was synthesized from poly A<sup>+</sup> mRNA utilizing a specifically designed "anchored oligo-dT" primer. Second strand cDNA was synthesized by a linear PCR using a biotinylated gene specific primer (BioT12-3Ext1, **Table 4.1**). T12 specific biotinylated cDNAs were then captured on streptavidin-coated magnetic beads (DYNAL) for purification. The 3' end was subsequently amplified by nested-PCR using the reverse primers specific for anchored sequence and the forward gene specific primers, T12-3Ext2 and T12-Ext3 (**Table 4.1**).

## 4.2.2 Expression Analysis

### 4.2.2.1 Northern blot Hybridization

Northern blot analysis was used to determine the size of the *T12* transcript as well as its expression pattern. A commercial multiple tissue Northern membrane (Clontech) was hybridized with a group of [ $\alpha^{32}\text{P}$ ]dCTP-labelled cDNA probes generated from RT-PCR products for 1.3-kb cDNA of T12 (section 4.2.1). The hybridization protocol is given in section 2.2.11.2. After washing, the hybridized membrane was exposed to X-ray film and autoradiographic developed.

### 4.2.2.2 Differential Expression Study

To determine the tissue and developmental stage-specific expression of the *SPG7* mRNA, a human RNA Master Blot from 50 different fetal and adult tissues (Clontech) was also hybridized with the same set of cDNA probes used for Northern hybridization. The hybridization protocol and solution was as recommended by the manufacturer. The membrane was washed five times with 2 x SSC, 1% SDS at 65 °C and was subjected to autoradiographic development. The quantity of mRNA spotted for each tissue on the

Master Blot was normalized, as recommended by the manufacturer, by using eight different housekeeping gene transcripts as probes.

#### 4.2.3 Protein Sequence Analysis

Analysis of protein sequence identity and function was done *in silico*. The BLASTP program (Altschul *et al.*, 1997) available at NCBI was used to search for amino acid sequence homology between T12 and protein sequence databases at NCBI. Sub-cellular localization of the T12 protein was performed utilizing a protein sub-cellular localization prediction tool, PSORT (Nakai and Horton, 1999).

#### 4.2.4 Identification of Genomic Structure

The genomic structure was determined from three cosmids, 383H6, 427B11, and 358D12. These cosmids form a contig and encompass the *T12/SPG7* region. A combination of two independent procedures were utilized. The first method involved the use of PCR amplification of genomic DNA with gene specific oligonucleotide primers designed for both strands of cDNA sequence. The amplified genomic products that were larger than those from cDNA were subsequently purified and sequenced from either end. The second method was by direct sequencing of cosmid clones utilizing the same set of gene specific primers. This was used when PCR amplification was not successful because of a large intervening sequence (intron). All primer sequences and their locations in the *T12* cDNA are given in **Table 4.2**. The exon/intron boundaries were identified by the sequence divergence in genomic DNA from that of cDNA. The size of each intron was determined from the PCR product size and some was estimated from the restriction map of the cosmid contig (Whitmore *et al.*, 1998a).

#### 4.2.5 SPG7 Mutation Study

To determine the spectrum of *SPG7* mutation in sporadic cases of HSP, DNA samples isolated from 116 HSP patients from unrelated families were obtained from three centers. These are the Children's National Medical Center, Research Center for Genetic Medicine, Washington, DC (Dr Francisco Murillo), the Department of Neurology, University of Michigan, USA (Dr John K Fink), and the Geriatric Research, Education and Clinical Center, Seattle, USA (Dr Thomas Bird). The available patient data are age of onset of each patient, the clinical form of either pure or complicated, and the family background. The DNA samples isolated from 60 clinically normal individuals of Caucasian background were used as normal control for mutation analysis

Mutation screening was by single-strand conformation polymorphism (SSCP) analysis (Hayashi, 1991; Nataraj *et al.*, 1999). Individual exon sequences were PCR amplified in 20- $\mu$ l reactions containing [ $\alpha^{32}$ P] dCTP with primers designed with respect to its flanking introns (*Table 4.4*). Each completed PCR product was added to an equal volume of formamide dye. The mixture was denatured at 100 °C for 10 min, cooled on ice and loaded onto non-denaturing polyacrylamide gel (containing 6%, 8% or 10% gel, dependent on PCR product size, and 7% glycerol) and MDE gel (FMC bioproducts). Electrophoresis was carried out at a constant 700 volt overnight at room temperature. Gel was blotted onto 3MM filter paper and dried under heat and vacuum. Dried gel was exposed to X-ray film overnight at -70 °C and autoradiograph developed. The samples showing mobility band shift when compared with normal controls were re-amplified and sequenced in parallel with controls.

### 4.3 Results

#### 4.3.1 Determination of Transcript Sequence

A schematic summary of the transcript sequence determination leading to the assembly of a full-length cDNA transcript of T12 is shown in **Figure 4.1**.

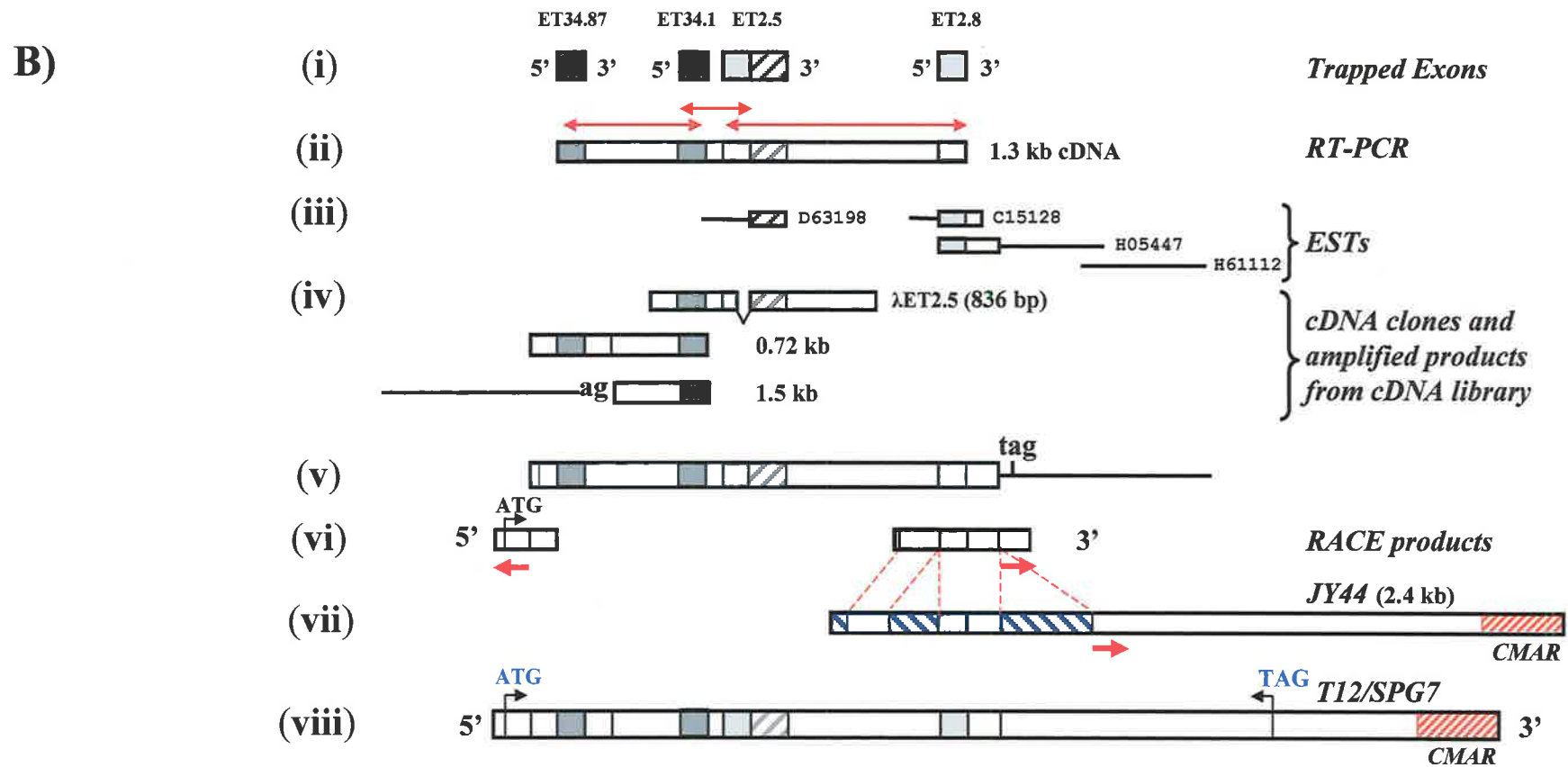
#### **4.3.1.1 Database Analysis and cDNA Sequencing**

A group of trapped exons, arranged in the centromeric to telomeric order in the 16q24.3 physical map as ET23.23, ET24.42, ET34.87, ET34.1, ET2.5, and ET2.8 (Whitmore *et al.*, 1998a) were initially assigned as "transcription unit 12 (T12)" (**Figure 1.3 and 4.1 A**). Database searches at this time revealed only two human ESTs, D63198 and C15128 that showed sequence homology to ET2.5 and ET2.8 respectively. Homology between the sequences of ET2.5 and D63198 was limited to only 126 bp out of the 229 bp of ET2.5. Since the cDNA clone corresponding to D63198 was not available for analysis, the trapped product ET2.5 was excised from the vector and then used as a probe to screen a fetal brain cDNA library (Clontech) from which a cDNA clone designed "λET2.5" was obtained. Sequencing of the λET2.5 cDNA revealed an insert sequence of 836 bp containing the sequence of both exons ET34.1 and ET2.5 (**Figure 4.2**)

Within the λET2.5 cDNA there was a 55 bp deletion of ET2.5 that resulted in the first 48 bp of ET2.5 being contiguous with 126 bp of the sequence that showed homology to the EST sequence D63198 (**Figure 4.2**). Analysis of the sequence at the deletion junction at nucleotide position 48 of ET2.5 revealed a sequence that resembled the consensus for a donor-splice site. Therefore, this suggests that there is alternative splicing of the ET2.5 sequence and that one form was present in λET2.5. This alternative splicing suggested that ET2.5 was likely to be trapped from two adjacent exons present in the genomic sequence cloned into the trap-vector. To verify this, primers were designed from both ends of ET2.5 (2.5F1 and 2.5R2, **Table 4.1**) and used to PCR amplify a fragment of 1.4 kb from genomic

---

**Figure 4.1** : Identification of the *T12/SPG7* transcript sequence. **A)** The physical location of the trapped exons and the restriction map of *EcoRI* site (**E**) within the genomic region encompassing the transcription unit 12 (*T12*). **B)** Schematic representation of the alignment of the trapped exons (**i**), joined sequence of RT-PCR products, “T12-1.3” (**ii**), sequence of homologous ESTs (**iii**), cDNA sequences isolated from cDNA library (**iv**), the primary assembled cDNA sequence (**v**), RACE products (**vi**), and a 2.4 kb *CMAR* cDNA (**JY44**) (**vii**), and the full-length 3083 bp transcript of *T12/SPG7* (**viii**). Sequence homology is indicated by the corresponding “filled”, “opened”, or “hatched” boxes. Reverse-hatched boxes indicate non-homologous regions found in JY44 cDNA. Single lines denote non-coding sequences (intronic) seen in ESTs and one of cDNA clones isolated from cDNA library. Red arrows denote the extended sequence that were obtained from the 5’ (reverse) and 3’ (forward) RACE products. *CMAR* indicates the sequence of original report *CMAR* (Pullman and Bodmer, 1992).





**Table 4.1**

**Oligonucleotide Primers used for T12(SPG7) cDNA sequence assembly**

<b>Primer</b>	<b>Sequences of oligonucleotide primers (5'→3')</b>	<b>Nucleotide position in the cDNA</b>
<b>Primers for trapped exon RT-PCR</b>		
34.87F	tga gac tgc taa ccc cta cc	194-213
34.87R	agt tgc cag agt ctg act gg	281-262
34.1F	atg tac cga atg cag gtt gc	631-650
34.1R	cct tgg cct cga tat tca gc	715-696
2.5F1	tgt act ctg tgg gga tga cg	764-783
2.5R1	agc act gaa tcc acc ttc cc	858-839
2.5F2	ggc tcg ttt cac cat tgt gg	876-895
2.5R2	aca aac tcg cgg act tcc ag	974-955
2.8R	cac tga atc ctg gtg tca gc	1552-1533
<b>Primers for T12/CMAR cDNA sequence assembly</b>		
T12-3Ext1	ctg cat cgt cta cat cga tg	1206-1225
T12-3Ext2	tga cat ttt gga cgg tgc tc	1374-1391
T12-3Ext3	ttg agc agc acc tga aga gc	1466-1485
T11A-R	cgt act cga agt tga gag tg	1642-1623
T12-3Ext4	tgg ttg cgt ttc atg agt cg	1709-1728
T12-3ExtR	gaa gag gtg ctg gtc tct gg	1851-1832
T12-CMAR	aac acc tcc caa cta ctt gg	2400-2381
CMAR1518	gaa cac ttc gag ttc cca ggg tta t	2644-2668
CMAR1230	caa cac tgc atg cgt gac tag acc t	2991-2967

**Note:** Forward primers 3Ext1, 3Ext2, 3Ext3, and 3Ext4 were designed from 3' sequence of T12-1.3 cDNA and 3' RACE product, whereas the reverse primers 3ExtR, T12-CMAR, 1518, and 1230 were designed from the JY44 cDNA.

**Table 4.2**  
**Sequences of Oligonucleotide primers used for Determination of the SPG7 Gene Structure**

<i>Primer Identity</i>	<i>Sequences of oligonucleotide primers (5'→3')</i>	<i>Nucleotide position in the cDNA</i>
T12-E1F	cgt aca tgg cca gca ggc	125-141
T12-E104R	caa caa caa tcc gtt gat cc	240-221
34.87F	tga gac tgc taa ccc cta cc	194-213
T12-EB2	cct tcg act tat cct tct cc	352-333
T12-EB1	cct caa ggt tga agc aga ag	308-327
T12-EB3	atg aca acc gcg atg acc ag	455-436
T12-EB4	aga gcg acg tgg tgg aag tc	563-582
34.1R	cct tgg cct cga tat tca gc	715-696
34.1F	atg tac cga atg cag gtt gc	631-650
2.5R1	agc act gaa tcc acc ttc cc	858-839
2.5F1	tgt act ctg tgg gga tga cg	764-783
2.5R2	aca aac tcg cgg act tcc ag	974-955
2.5F2	ggc tcg ttt cac cat tgt gg	876-895
161R	atg acc tcc acg aac tct gg	1145-1126
161F	aga acg ctt cct cca gct tg	993-1012
T12-EB7	cga tct cat cga tgt aga cg	1231-1212
T12-3Ext1	ctg cat cgt cta cat cga tg	1206-1225
T12-EB8	gca ccg tcc aaa atg tca gc	1391-1372
T12-3Ext2	tga cat ttt gga cgg tgc tc	1374-1391
2.8R	cac tga atc ctg gtg tca gc	1552-1533
T12-3Ext3	ttg agc agc acc tga aga gc	1466-1485
T11A-R	cgt act cga agt tga gag tg	1642-1623
T12-E116R	ccg act cat gaa acg caa cc	1729-1710
T12-3Ext4	tgg ttg cgt ttc atg agt cg	1709-1728
T12-3ExtR	gaa gag gtg ctg gtc tct gg	1851-1832
T12-3Ext5	agc acc tct tca cca agg ag	1841-1860
T12-R1	gtg acc ttc ctc agg tcg tc	1964-1945
T12-EB9	cct act cca tgg tga agc ag	1973-1992
T12-EB10	cct tct cgg tgt gtc tgt ag	2149-2130
T12-3Ext7	aca gac aca ccg aga agg tg	2132-2151
T12-R3	caa tga gag cct caa tgt cc	2242-2223
T12-CMAR-F	acc gaa gag acc cag cag c	2332-2350
T12-CMAR	aac acc tcc caa cta ctt gg	2400-2381

**A** 5' CCAAGGGCGAGGTGCAGCGCGTCCAGGTGGTGCCTGAGAGCGACGTGGTGGAAAGTCTACCTG  
CACCCCTGGAGCCGTGGTGTGTTGGGCGGCCT ET34.1  
ATATTGACAAGTTTGAAGAGAAGCTTCGAGCAGCTGAAGATGAGCTGAATATCGAGGCCAAGGA  
ET34.1 CAGGATCCAGTTTCTACAAGCGAACAGGATTCTTTGGAAAT ET2.5  
ACGGCAGTGGGCCTGGCCATCCTGTG *gtatggtttccgtctggccgggatgactggaagggaag*  
*gtggattcagtgcctttt* AATCAGCTTAAAATGGCTCGTTTCACCATTGTGGATGGGAAGATGGG  
GAAAGGAGTCAGCTTCAAAGACGTGGCAGGAATGCACGAAGCCAAACTGGAAGTCCGCGAGTTT  
ET2.5 GTGGATTATCTGAAGAGCCCAGAACGCTTCCTCCAGCTTGGCGCCAAGGTCCCAAAGGGCGCAC  
TGCTGCTCGGCCCCCCCGGCTGTGGGAAGACGCTGCTGGCCAAGGCGGTGGCCACGGAGGCTCA  
GGTGCCCTTCCTGGCGATGGCCGGCCAGAGTTCGTGGAGGTCATTGGAGGCCTCGGCGCTGCC  
CGTGTGCGGAGCCTCTTTAAGGAAGCCCGAGCCCGGGCCCCCTGCATCGTCTACATCGATGAGA  
TCGACGCGGTGGGCAAGAAGCGCTCCACCACCATGTCCGGCTTCTCCAACACGGAGGAGGAGCA  
GACGCTCAACCAGCTTCTGGTAGAAATGGATGGAATGGGTACCACAGACCATGTCATCGTCCTG  
GCGTCCACGAACCGAGCTGACATTTTGGACGGTGCTCTGATGAGGCCAGGCCGACTGGACC 3'

**B** 2.5F1  
5' TGCCCTGTACTCTGTGGGGATGACGGCAGTGGGCCTGGCCATCCTGT GTATGTTTCCGTC  
TGGCCGGGATGACTGGAAGGGAAGGTGGATTCACTGCTTTT gtaagttctgta-----  
----- 1.2kb -----  
ctttctgcac ag AATCAGCTTAAAATGGCTCGTTTCACCATTGTGGATGGGAAGATGGGGAAAG  
GAGTCAGCTTCAAAGACGTGGCAGGAATGCACGAAGCCAAACTGGAAGTCCGCGAGTTTGTGGA  
TTATCTGAAG 3' 2.5R1

**Figure 4.2** : The nucleotide sequence cDNA λET2.5 **A**) The 836 bp nucleotide sequence of cDNA clone λET2.5 (except the sequence in lower case) that includes both ET34.1 and ET2.5 (underlined with arrows indicate the boundaries), was isolated from fetal brain cDNA library. The sequence in lower case indicates the 55 bp fragment that was deleted from clone λET2.5. **B**) The partial genomic sequence of ET2.5 that was confirmed by PCR as the two exons separated by an intron of approximately 1.2 kb. The cryptic splice site that causes the deletion of 55 bp (underlined) is *boxed*. The normal donor splice site at the exon / intron boundary is underlined (*red*). Also indicated are the sequence positions of the two primers used for PCR

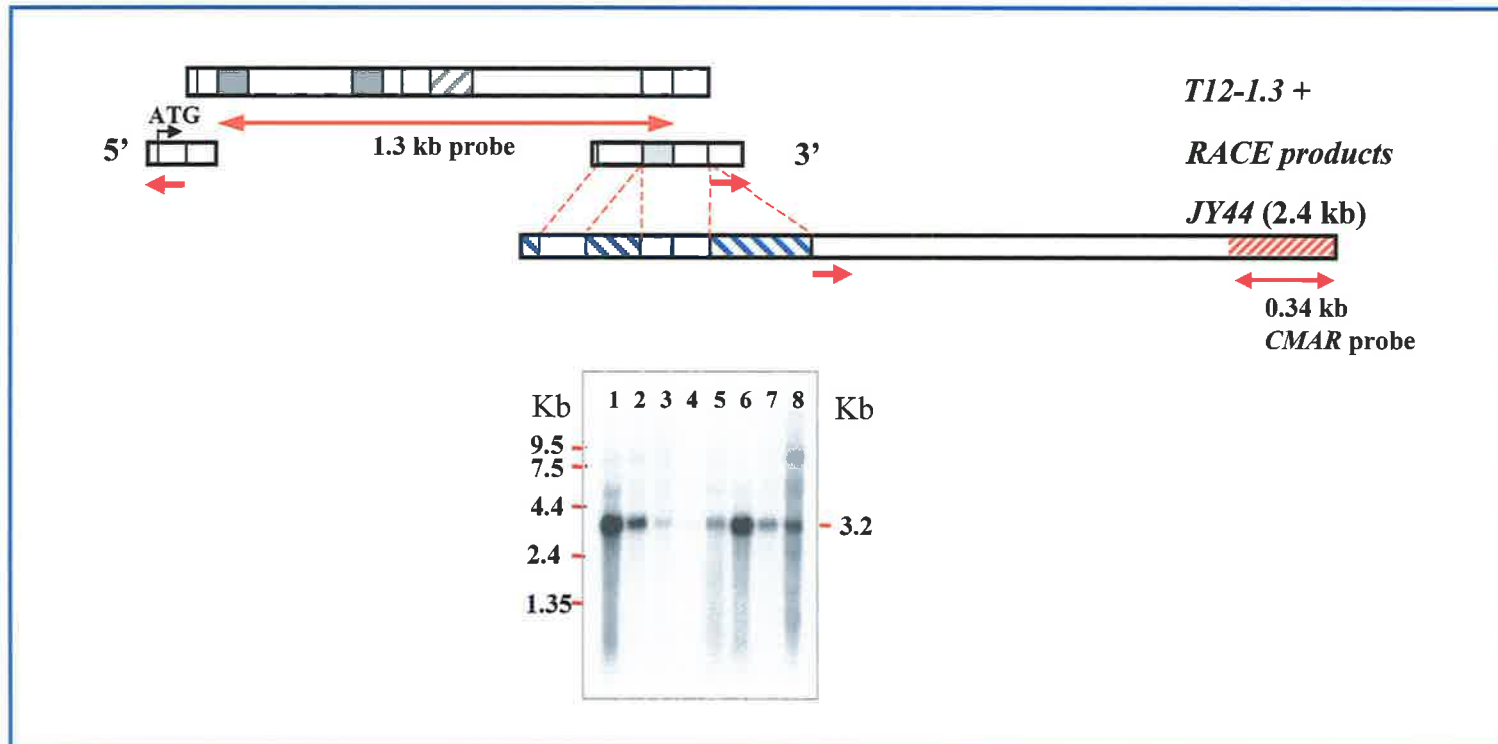
DNA. Sequencing of this 1.4 kb DNA fragment demonstrated that ET2.5 was trapped from two adjacent exons separated by an intron of approximately 1.2 kb.

RT-PCR analysis was used to investigate if the trapped exons ET23.23, ET24.42, ET34.87, and ET2.8 were also part of the same transcript as ET34.1 and ET2.5. Primers were designed from these trapped exon sequences (*Table 4.1*) and used in various combinations in RT-PCR to amplify the transcription products from fetal brain total RNA. Three overlapping RT-PCR products were successfully obtained and sequenced. This was assembled to form a cDNA sequence of 1.3 kb (this will be termed **T12-1.3**) that linked four trapped exons, ET34.87, ET34.1, ET2.5, and ET2.8, with the ET34.87 as the most 5' sequence (*Figure 4.1 B.ii*). This cDNA sequence exhibited an open reading frame (ORF) with no start or stop codons. Therefore, further sequence identification was required to complete a full-length transcript. For both ET23.23 and ET24.42, despite several attempts, RT-PCR was unable to isolate cDNA that joined these two trapped-products nor linked to the T12-1.3 sequence, suggesting they were not part of this transcript.

Hybridization of a multiple tissue Northern membrane (section 4.2.2.1), utilizing the RT-PCR product of T12-1.3 as probe, identified a predominant 3.2 kb mRNA band in all tissue tested, though with varying degree of intensity. With a longer exposure several faint band of various sizes were also observed (*Figure 4.3*).

#### **4.3.1.2 Transcript Sequence Assembly**

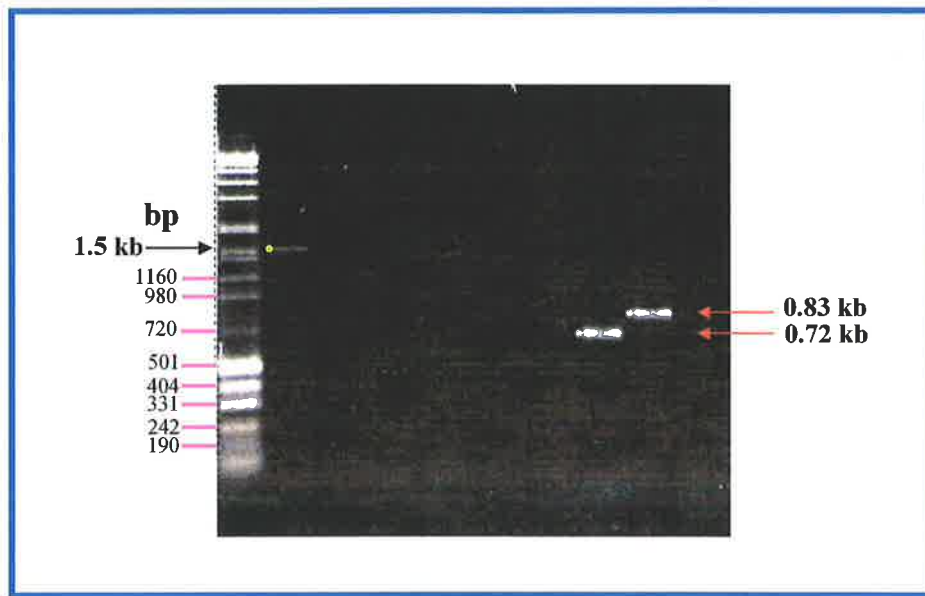
According to the Northern analysis result, the T12-1.3 cDNA sequence was only a part of a full length transcript of 3.2 kb. Extension from both 5' and 3' ends was therefore undertaken. To extend the 5' sequence of the transcript, PCR was first used with 34.87F and 34.1R primers (*Table 4.1*) to screen pools of a fetal brain 5' stretched cDNA library (Clontech, the library had been formatted as described in section 2.2.18.1). The cDNA



**Figure 4.3 :** Northern Blot Hybridization of *T12/CMAR*. The multiple tissue Northern membrane (Clontech) was hybridized with P<sup>32</sup>-labelled cDNA probe derived from either 1.3 kb assembled RT-PCR product (T12-1.3) or 0.34 kb *CMAR* sequence. The same hybridization pattern with the prominent band of 3.2 kb was obtained from either probe. The tissue sources of mRNA are 1 *heart*, 2 *brain*, 3 *placenta*, 4 *lung*, 5 *liver*, 6 *skeletal muscle*, 7 *kidney*, 8 *pancreas*. The RNA size markers are indicated on the left side of the membrane.

pools that gave a strong PCR signal were then selected for subsequent PCR amplification utilizing the reverse primers designed to the 5' sequence of T12-1.3 and a vector primer (**Table 4.1**). As shown in **Figure 4.4**, two PCR products of approximately 0.72 kb and 1.5 kb were obtained. The identity of these products was confirmed by Southern hybridization using 34.87F/34.1R RT-PCR product as probe. Sequencing of the 0.72 kb PCR fragment identified the 5' sequence extended from ET34.87 by a further 98 bp. The 1.5 kb fragment, in contrast, showed only the continuation of the transcript sequence of 242 bp 5' upstream from ET34.1, then the sequence deviated beyond this point and extended a further 1.1 kb (**Figure 4.1 B.iv**). This deviated sequence was later found to be an intron 3, and the 242 bp sequence was exon 4, of the gene. To further extend the transcript, 5' RACE was also carried out using the "SMART" technology developed by Clontech (sections 4.2.1.3 and 2.2.17.2). This allowed the 5' sequence to be extended another 92 bp from the initially obtained 98 bp, mentioned above. Within this 92 bp extended sequence an in-frame ATG start codon was identified. Although this ATG codon was not preceded by any in-frame upstream stop codon, it was situated in a sequence which conformed to the Kozak consensus of translation start site (Kozak, 1991). Therefore, it was considered likely that this represented the 5' start of the transcript, although additional 5'UTR sequence may be present.

To identify the transcript sequence that extended from the 3' end of the T12-1.3 fragment, database searches were first performed using ET2.8 and the EST C15128 as queries. This identified the EST H05447, which overlapped with C15128 and part of ET2.8. Sequencing of the cDNA clone corresponding to H05477 was carried out and identified the sequence that extended a further 580 bp from the 3' end of the T12-1.3 cDNA (**Figure 4.1 B.iii**). Subsequent database searches utilizing the 3' sequence of H05447 identified another EST, H61112, a 5' sequence of a human fetal liver cDNA. This



**Figure 4.4** : Agarose Gel pattern of PCR products amplified from 5' sequence of *T12*. PCR was undertaken on pools of formatted 5'- stretched fetal brain cDNA library (2.2.18.1). The 1.5 and 0.72 kb bands were obtained from pools no. A1 and A8 respectively, using vector primer, *gt11R*, and gene specific primer, *34.1R*. Band of 0.83 kb was also the product amplified from pool no. A8, but using gene specific primer *2.5R1* instead of *34.1R*. The DNA size marker, loaded in the first lane, is indicated in **bp**.

extended another 360 bp from the 3' end of H05447. Further continuation of the 3' sequence by database searching was unsuccessful. Furthermore, the cDNA clone corresponding to H61112 was not available for analysis.

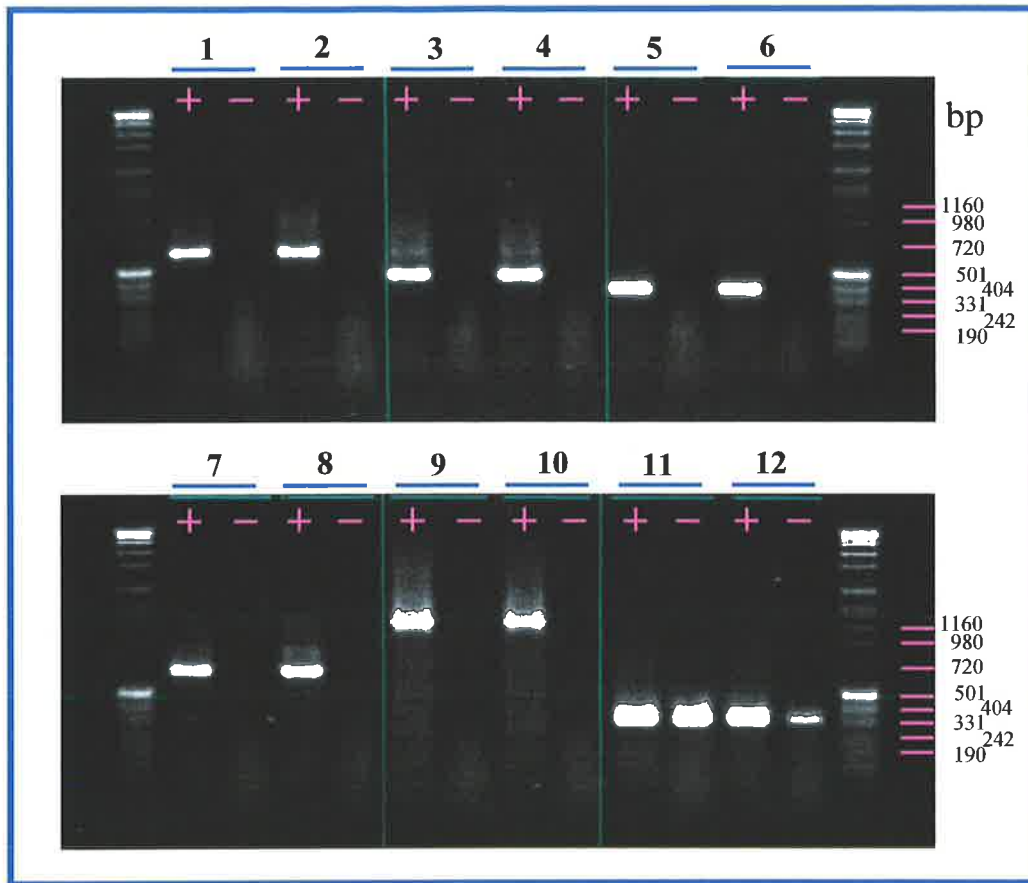
At this stage the sequence from the 5' extension, the T12-1.3 fragment, and the available 3' extended sequence were joined to form a 2.1 kb cDNA sequence. Based on Northern analysis, this assembled sequence represented only 60% of the full-length transcript. An ORF of 1.7 kb was derived from this assembled cDNA sequence. This ORF was terminated by a stop codon, TAG, which is only 157 bp downstream from ET2.8 and situated within the sequence of H05447 (**Figure 4.1 B.v**). Later, upon determination of the T12 gene structure, it was found that the two ESTs, H05447 and H61112, were derived from genomic contaminates in the cDNA library. The 3' extended sequence from T12-1.3 therefore contained part of an intron, which created a stop codon.

3' RACE (section 4.2.1.3) was also undertaken as an alternative attempt to extend the sequence 3' from T12-1.3 cDNA. Three forward gene specific primers were designed to the 3' region of T12-1.3 (**Table 4.1**). Essentially, the first strand cDNA was primed with an anchored oligo(dT) primer and reverse-transcribed from fetal brain poly A<sup>+</sup> RNA. This was then used as a template for a biotinylated gene specific primer to synthesize a second cDNA strand. This cDNA template was purified and used for subsequent nested PCR utilizing gene specific primers and the primers specific for the sequence of the anchored oligo-dT. The final PCR product was cloned and sequenced. This yielded a cDNA sequence that extend a further 160 bp from ET2.8. When compared with the previous assembled sequence, this 3' RACE product revealed an extended sequence identical to the first 111 nucleotides downstream from ET2.8 but then diverged into a continuation of the ORF without the stop signal as seen previously.



BLAST analysis of non-redundant (nr) database at NCBI, using this 3' RACE product subsequently identified a 2.4-kb cDNA sequence of the *CMAR* gene (GenBank accession no. AF034795). This cDNA was isolated from a lymphoblastoid cell line (Durbin *et al.*, 1997). Sequence homology was between the 3' region of T12-1.3 that include the 3' RACE product and the 5' region of a *CMAR* cDNA, clone JY44. As shown in **Figure 4.1 B.vii**, the homology of the JY44 sequence with T12-1.3 cDNA appeared to be interrupted by two regions recognized as *Alu*-like elements in this *CMAR* cDNA sequence (Durbin *et al.*, 1997). The sequence overlapping between the 3' region of T12-1.3 and the 5' region of the *CMAR* transcript provided evidence that *CMAR* is likely to be part of *T12*. To examine this possibility, a set of forward primers designed within the 3' region of T12-1.3 cDNA together with reverse primers from within the *CMAR* sequence (**Figure 4.1 B.vii** and **Table 4.1**) were used for RT-PCR analysis of total RNA isolated from fetal brain and fetal skeletal muscle. Overlapping RT-PCR products (**Figure 4.5**) were obtained and sequenced. The assembled sequence demonstrated the link between the T12-1.3 cDNA and the *CMAR* sequence without any interruption by *Alu*-like elements. Since the cDNA JY44, representing the 2.4-kb *CMAR* transcript sequence, was derived from a human B-lymphoblastoid cell line, insertion of two *Alu* elements may represent a unique mRNA expressed in this cell type. RT-PCR analysis was then undertaken of RNA samples isolated from three different human lymphoblastoid cell lines. Sequence analysis of the RT-PCR products, however, did not reveal any *Alu* element insertions in these sequences.

Finally, the sequences were assembled to form a full-length *T12* transcript sequence of 3,083 bp that included the *CMAR* transcript sequence and extended 3' from the original T12-1.3 cDNA fragment. The 340-bp PCR product amplified from the 3' region of *T12* corresponding with the *CMAR* sequence, was also used to hybridize to a multiple tissue Northern membrane. The same prominent 3.2-kb band was detected similar to that



**Figure 4.5** : RT-PCR analysis of the 3' half of *T12* transcript. Total RNA from fetal brain and fetal skeletal muscle was analysed by RT-PCR utilizing oligonucleotide primers designed to *T12* and *CMAR* cDNA sequence. The products contain in gel are: *lanes 1* fetal brain, *2* fetal sk.muscle, **size** 648 bp, *primers* T12-3Ext1/-3ExtR; *lanes 3* fetal brain, *4* fetal sk.muscle, **size** 480 bp, *primers* T12-3Ext2/-3ExtR; *lanes 5* fetal brain, *6* fetal sk.muscle, **size** 388 bp, *primers* T12-3Ext3/-3ExtR; *lanes 7* fetal brain, *8* fetal sk.muscle, **size** 676 bp, *primers* T12-3Ext4/-CMAR; *lanes 9* fetal brain, *10* fetal sk.muscle, **size** 1282 bp, *primers* T12-3Ext4/CMAR-1230; *lanes 11* fetal brain, *12* fetal sk.muscle, **size** 348 bp, *primers* CMAR-1518/-1230. *Plus (+)* and *minus (-)* indicate the reaction with and without enzyme reverse transcriptase respectively. Bands shown in *lanes 11 (-)* and *12 (-)* indicate the products which were amplified from contaminated genomic DNA. Size markers, in **bp**, are indicated on the right hand side of the gels.

observed when the membrane was probed with the *T12*-1.3 cDNA (**Figure 4.3**). The result of this Northern hybridization confirmed that *CMAR* was indeed part of *T12*. Therefore, the reported 2.4-kb *CMAR* cDNA sequence (JY44) was unlikely to exist as a separate gene. This was shown in at least two fetal tissues, and three lymphoblastoid cell lines as determined by RT-PCR, and eight adult tissues tested by Northern analysis.

Insertion of the *Alu* elements seen in the 2.4 kb version of the *CMAR* transcript, if present in *T12*, would truncate up to 405 bp of the ORF by introducing an in-frame stop codon. Although such *Alu* insertion was not detected in fetal tissues, lymphoblastoid cell lines, and in adult tissues, this event may occur as a mechanism of gene inactivation in tumor tissue. To test if this insertion is involved in breast cancer, RT-PCR of the *T12* transcript was undertaken in RNA isolated from 12 breast cancer cell lines, using the set of primers described in **Table 4.1**. Sequencing of the RT-PCR products obtained from this experiment did not reveal any *Alu* element insertions nor sequence changes indicating single base mutations. Therefore there was no evidence of *T12* inactivation in breast cancer arising from this mechanism.

#### 4.3.2 Transcript Sequence and Expression Analysis

The *T12* transcript contains an ORF of 2385 bp that was predicted to code for a protein of 795 amino acids (**Figure 4.6**). The ORF is followed by a 680-bp 3' UTR, with polyadenylation signal, AATAAA, located 20 bp upstream from the polyadenylation site, as confirmed by alignment of the transcript sequence against the genomic sequence obtained by direct sequencing of cosmid 383H6 encompassing the *CMAR* region. Within the 3' UTR an insertion/deletion polymorphism of CACA was frequently observed among the RT-PCR products and genomic sequence. The CACA insertion/deletion polymorphism

---

***Figure 4.6*** : Nucleotide transcript and deduced amino acid sequence of *T12(SPG7)*. The transcript codes for a protein of 795 amino acid residues. Nucleotide and amino acid positions are indicated by numbers on the right hand column. The adenine (A) nucleotide of the start codon is given as the position "1" of the transcript. The stop codon at positions 2386-2388 is denoted by asterisk underneath. Polyadenylation signal AATAAA is indicated in *blue* at positions 3040-3045. Within the 3' UTR the short region of insertion/deletion polymorphism of CACA is underlined. The transcript is coded for by 17 exons of the gene. The exon boundaries are indicated by the vertical lines with double-head arrows and given numbers of exons. The identity of each trapped product is indicated above its corresponding exon.

TTTCAGGCCAACATGGCCGTGCTGCTGCTGCTGCTCCGTGCCCTCCGCCGGGTCCAGGCCGGGTCTCGGCCGCTGTG 68  
M A V L L L L L R A L G R G P G P G P R P L W 23

GGGCCAGGCCGGCCTGGAGTCCAGGGTCCCCGCCAGGCCGGGAGGGGGCGGCCGTACATGGCCAGCAGGCCCTCCGG 148  
G P G P A W S P G F P A R P G R G R P Y M A S R P P 49

GGGACCTCGCCGAGGCTGGAGGCCGAGCTCTGCAGAGCTTACAATTGAGACTGCTAACCCCTACCTTTGAAGGGATCAAC 228  
G D L A E A G G R A L Q S L Q L R L L T P T F E G I N 76

GGATTGTTGTTGAAACAACATTTAGTTCAGAATCCAGTCAGACTCTGGCAACTTTAGGTGGTACTTTCTATTTTAACAC 308  
G L L L K Q H L V Q N P V R L W Q L L G G T F Y F N T 103

CTCAAGGTTGAAGCAGAAGAATAAGGAGAAGGATAAGTCTGAAGGGGAAGGCCCTGAAGAGGACGAAGAGGAGGAGAC 388  
S R L K Q K N K E K D K S K G K A P E E D E E E R R 129

GCCGTGAGCGGGACGACCAGATGTACCGAGAGCGGCTGCGCACCTTGCTGGTCATCGCGGTTGTCTGAGCCTCCTGAAT 468  
R R E R D D Q M Y R E R L R T L L V I A V V M S L L N 156

GCTCTCAGCACCAGCGGAGGCGAGCATTTTCTGGAACGACTTTGTCCACGAGATGCTGGCCAAGGGCGAGGTGCAGCGCT 548  
A L S T S G G S I S W N D F V H E M L A K G E V Q R V 183

CCAGGTGGTGCCTGAGAGCGAGCTGGTGGAACTTACCTGCACCCTGGAGCCGTTGGTGGTGGGGCGGCTCGGCTAGCCT 628  
Q V V P E S D V V E V Y L H P G A V V F G R P R L A 210

**ET 34.1**  
TGATGTACCGAATGCAGGTTGCAAATATTGACAAGTTTGAAGAGAAGCTTCGAGCAGCTGAAGATGAGCTGAATATCGAG 708  
L M Y R M Q V A N I D K F E E K L R A A E D E L N I E 236

GCCAAGGACAGGATCCCAGTTTCTTACAAGCGAACAGGATTTTGGAAATGCCCTGTACTCTGTGGGGATGACGGCAGT 788  
A K D R I P V S Y K R T G F F G N A L Y S V G M T A V 263

GGGCTGGCCATCCTGTGGTATGTTTTCCGTCTGGCCGGGATGACTGGAAGGGAAGGTGGATTTCAGTCTTTAATCAGC 868  
G L A I L W Y V F R L A G M T G R E G G F S A F N Q 289

**ET2.5/2**  
TTAAAATGGCTCGTTTTCCACATTGTGGATGGGAAGATGGGAAAGGAGTCCAGCTTCAAAGACGTGGCAGGAATGCACGAA 948  
L K M A R F T I V D G K M G K G V S F K D V A G M H E 316

GCCAACTGGAAGTCCGCGAGTTTGTGGATTATCTGAAGAGCCGAGAAGCTTCTCCAGCTTGGCGCCAAGGTCCCAAA 1028  
A K L E V R E F V D Y L K S P E R F L Q L G A K V P K 343

GGGCGCACTGCTGCTCGGCCCCCCGGCTGTGGGAAGACGCTGCTGGCCAAGGCGGTGGCCACGGAGGCTCAGGTGCCCT 1108  
G A L L L G P P G C G K T L L A K A V A T E A Q V P 369

TCCTGGCGATGGCCGGCCAGAGTTCGTGGAGGTCATTGGAGGCCCTCGGCGCTGCCGCTGCGGAGCCTCTTTAAGGAA 1188  
F L A M A G P E F V E V I G G L G A A R V R S L F K E 396

GCCCGAGCCCGGCCCCCTGCATCGTCTACATCGATGAGATCGACGCGGTGGGCAAGAAGCGCTCCACCACCATGTCCGG 1268  
A R A R A P C I V Y I D E I D A V G K K R S T T M S G 423

CTTCTCCAACACGGAGGAGGAGCAGACGCTCAACCAGCTTCTGGTAGAAATGGATGGAATGGGTACCACAGACCATGTCA 1348  
F S N T E E E Q T L N Q L L V E M D G M G T T D H V 449

TCGTCTGGCGTCCACGAACCGAGCTGACATTTTGGACGGTGTCTGATGAGGCCAGGCCGACTGGACCGGCACGTCTTC 1428  
I V L A S T N R A D I L D G A L M R P G R L D R H V F 476

ATTGATCTCCCCACGCTGCAGGAGAGGGCGGAGATTTTTGAGCAGCACCTGAAGAGCCTGAAGCTGACCCAGTCCAGCAC 1508  
I D L P T L Q E R R E I F E Q H L K S L K L T Q S S T 503

CTTTTACTCCCAGCGTCTGGCAGAGCTGACACCAGGATTCAGTGGGCTGACATCGCCAACATCTGCAATGAGGCTGCGC 1588  
F Y S Q R L A E L T P G F S G A D I A N I C N E A A 529

TGCACGGCGCGGGAGGGACACACTTCCGTGCACACTCTCAACTTCGAGTACGCCGTGGAGCGCGTCTCGCAGGACT 1668  
L H A A R E G H T S V H T L N F E Y A V E R V L A G T 556

GCCAAAAGAGCAAGATCCTGTCCAAGGAAGAACAGAAAGTGGTTGCGTTTCATGAGTCCGGCCACGCTTGGTGGGCTG 1748  
A K K S K I L S K E E Q K V V A F H E S G H A L V G W 583

ET 34.87

ET2.5/1

ET 2.8

13      14  
 GATGCTGGAGCACACGGAGGCCGTGATGAAGTCTCCATAACCCCTCGGACAAAACGCCGCCCTGGGCTTTGCTCAGATGC    1828  
 M L E H T E A V M K V S I T P R T N A A L G F A Q M    609

TCCCCAGAGACCAGCACCTCTTACCAAGGAGCAGCTGTTTGGAGCGGATGTGCATGGCCTTGGGAGGACGGGCCTCGGAA    1908  
 L P R D Q H L F T K E Q L F E R M C M A L G G R A S E    636

14      15  
 GCACCTGTCTTCAACGAGGTCACTTCTGGGGCACAGGACGACCTGAGGAAGTCCACCCGCATCGCCTACTCCATGGTGAA    1988  
 A L S F N E V T S G A Q D D L R K V T R I A Y S M V K    663

GCAGTTTGGGATGGCACCTGGCATCGGGCCCATCTCCTTCCCTGAGGCGCAGGAGGGCCTCATGGGCATCGGGCGGGCC    2068  
 Q F G M A P G I G P I S F P E A Q E G L M G I G R R    689

15      16  
 CCTTCAGCCAAGGCCTGCAGCAGATGATGGACCATGAAGCAAGACTGCTGGTGGCCAAGGCCTACAGACACACCGAGAAG    2148  
 P F S Q G L Q Q M M D H E A R L L V A K A Y R H T E K    716

16      17  
 GTGCTGCAGGACAACCTGGACAAGTTGCAGGCGCTGGCAAACGCCCTTCTGGAAAAGGAAGTGATAAACTATGAGGACAT    2228  
 V L Q D N L D K L Q A L A N A L L E K E V I N Y E D I    743

TGAGGCTCTCATTGGCCCCGCCCCCATGGGCCGAAGAAAATGATCGCACCCGAGAGGTGGATCGACGCCAGAGGGAGA    2308  
 E A L I G P P P H G P K K M I A P Q R W I D A Q R E    769

AACAGGACTTGGGCGAGGAGGAGACCGAAGAGACCCAGCAGCCTCCACTTGGAGGCGAAGAGCCGACTTGGCCCAAGTAG    2388  
 K Q D L G E E E T E E T Q Q P P L G G E E P T W P K \*    795

TTGGGAGGTGTTGGCTGCACGTGCGGGTGGTCCGGGAAGTGAGGGCTCACTCAGCCACCCTGAGTTGCTTTTCAGCTGAG    2468  
 GTTTGCACTTCCTCTCGCGGCCCTCAGTAGTCCCTGCACAGTGACTTCTGAGATCTGTTGATTGATGACCCTTTTCATGA    2548  
 TTTTAAGTTTCTCTGCAGAACTACTGACGGAGTCTGTGTTTGTGAGTCGTTTCCCTATGGGGAAGGTTATCAGTGCT    2628  
 TCCCGAGTGAGCATGGAACACTTCGAGTTCACAGGGTTATAGACAGTCGTTCCAGTGTGGCTGAGGCCACCCAGAGGCA    2708  
 GCAGAGCATTGAGACTCCAAACAGACCCCTGTTTCATGCCGACGCTTGCACGACCCGCCAGTTCCTGTGGCTCCCTCGGA    2788  
 ATGCTAAGGGGATCGGACATGAAAGGACCCCTGTGAGCCGATGTCCTATCTCCAGCGGCCCTGTCATCCAGCTCACTCAT    2868  
 CAATGGGGCCACACAGTCAGGCCAGGCCTGGGCTCCGGAGGACTCACCCTGCCCCCTGCTGCCATGTGGACTGGTGC    2948  
 AAGTTGAGGACTTCTTGTGCTGTTAGTCACGCATGCAGTGTGGGGATGCCTTGGTTTTTACTGCTCTGAGAATTGTTGA    3028  
 GATACTTTACTAATAAACTGTGTAGTTGGAAAATCTAA    3065

<sup>has</sup> have also been described in the previous reports of *CMAR* (Koyama *et al.*, 1993; Durbin *et al.*, 1994).

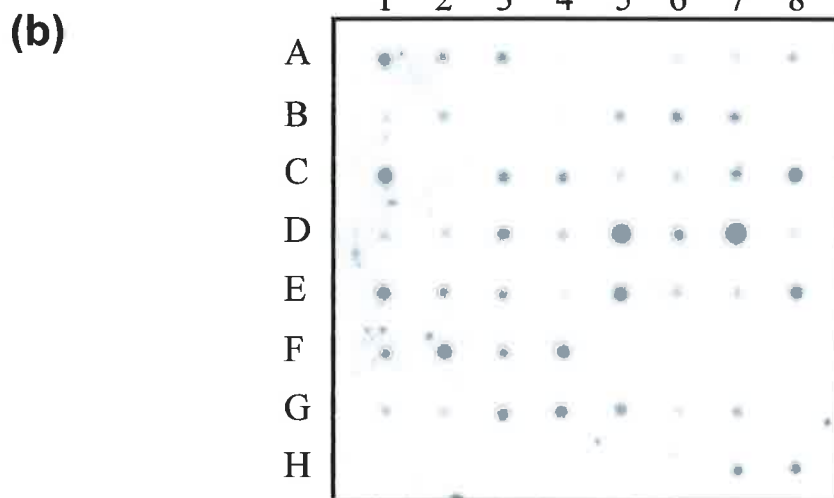
In addition to the expression study in multiple tissue Northern, tissue distribution of *T12* mRNA expression was also undertaken. A cDNA probe derived from 3' UTR of *T12* was used to hybridize to a human mRNA Master Blot from 50 different fetal and adult tissues (Clontech). As shown in **Figure 4.7**, expression was high in a series of tissues from the central nervous system: in the amygdala, caudate nucleus, thalamus, subthalamic nucleus, and spinal cord. Intermediate expression was observed in the frontal lobe of the cerebrum, the hippocampus, <sup>e</sup>mdulla oblongata, and putamen, with the lowest expression in the substantia nigra. Compared with adult brain, heart, and thymus, the corresponding fetal tissues showed a lower expression. In contrast, the level of expression was approximately the same in fetal kidney and liver when compared with the corresponding adult tissues. A high level of mRNA was present in adrenal gland and salivary gland, with a very low expression in mammary gland and testis

### 4.3.3 Protein Sequence Identity

With the complete *T12* amino acid sequence of 795 residues, homology searches of protein database using the BLASTP algorithm identified homology to a diverse group of proteins which were member of the AAA protein family (ATPases associated with diverse cellular activities). The most striking similarity was observed with the yeast mitochondrial proteins, Afg3p and Rca1p (both at 55% amino acid identity), and with a lesser similarity to Yme1p (52% identity). Afg3p, Rca1p and Yme1p are metalloproteases, situated in the inner membrane of mitochondria and exhibit both proteolytic and chaperone-like functions (Rep and Grivell, 1996; Patel and Latterich, 1998). Within the *T12* protein, which is similar to those of yeast, an ATP-binding motif was observed between residues 349 and

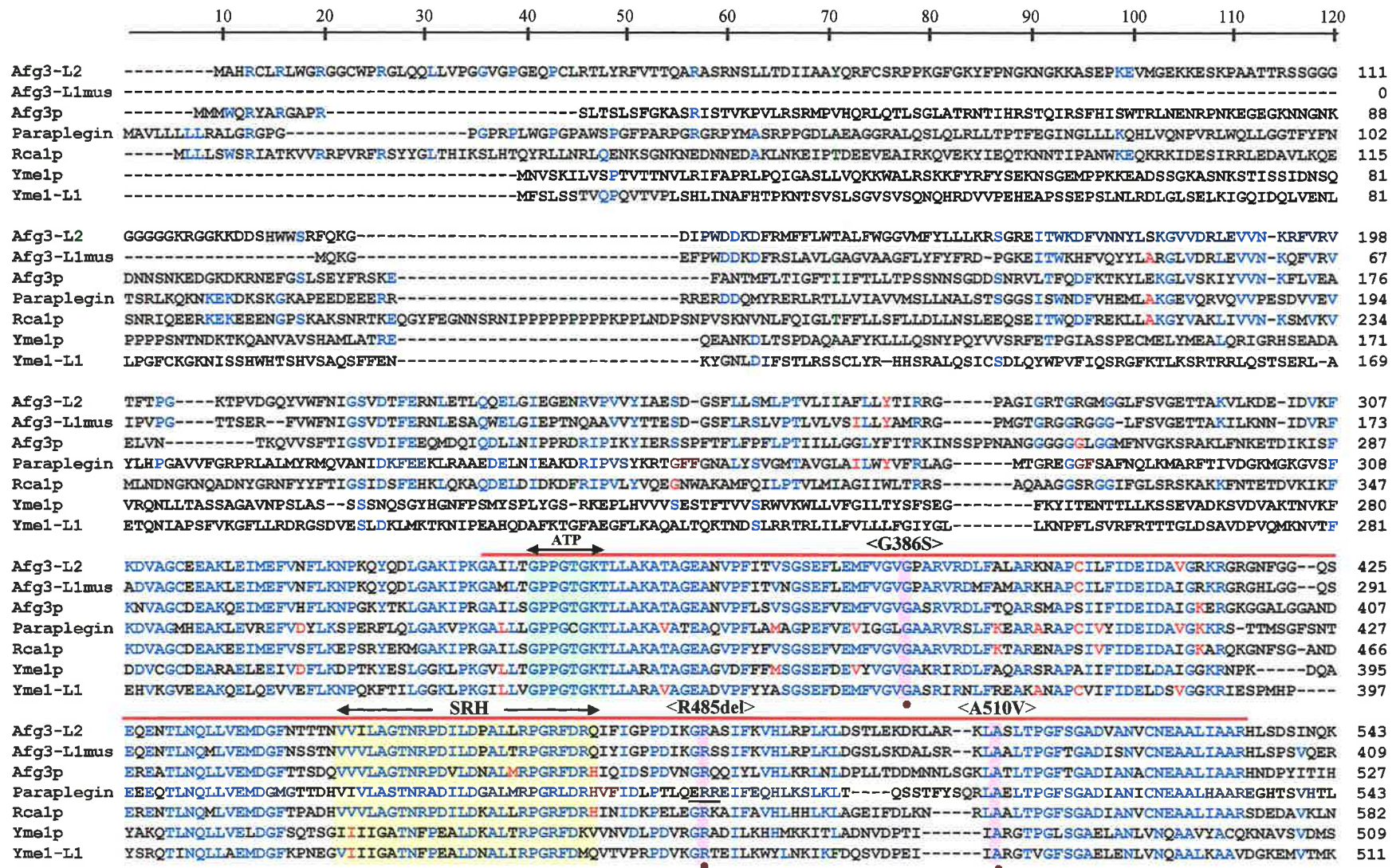
**(a)**

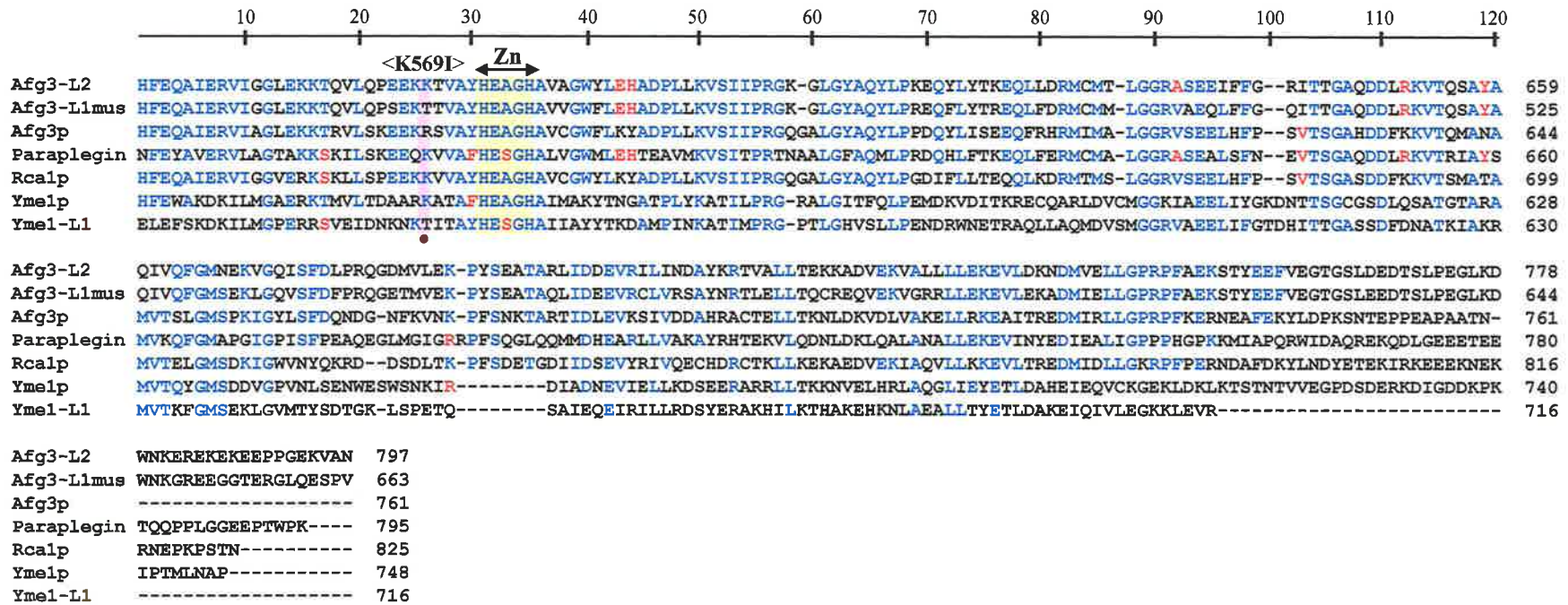
	1	2	3	4	5	6	7	8
A	whole brain	amygdala	caudate nucleus	cerebellum	cerebral cortex	frontal lobe	hippocampus	medulla oblongata
B	occipital lobe	putamen	substantia nigra	temporal lobe	thalamus	sub-thalamic nucleus	spinal cord	
C	heart	aorta	skeletal muscle	colon	bladder	uterus	prostate	stomach
D	testis	ovary	pancreas	pituitary gland	adrenal gland	thyroid gland	salivary gland	mammary gland
E	kidney	liver	small intestine	spleen	thymus	peripheral leukocyte	lymph node	bone marrow
F	appendix	lung	trachea	placenta				
G	fetal brain	fetal heart	fetal kidney	fetal liver	fetal spleen	fetal thymus	fetal lung	
H	yeast total RNA	yeast tRNA	E.coli rRNA	E.coli DNA	poly r(A)	Human Cot1 DNA	human DNA 100 ng	human DNA 500 ng



**Figure 4.7** : Differential Expression of *T12/SPG7*. **(a)** Diagram of the tissue origin and position of human polyA<sup>+</sup> mRNA dot-blotted on a nylon membrane. **(b)** Hybridization with the [<sup>32</sup>P]dCTP-labelled *T12* cDNA probe derived from the region between nt 1206 and 2991. The amount of polyA<sup>+</sup> mRNA in each dot was normalized by manufacturer using eight different housekeeping gene cDNA probes (Clontech)







**Figure 4.8** : Comparison of the T12/SPG7 protein, “paraplegin”, sequence with the sequence of closely related proteins. Protein sequences were aligned by CLUSTAL-W method. Numbers on the right-hand column denote the amino acid positions. Scale bars above the aligned sequence are given to assist in locating the amino acid position. Homologous sequence with identical amino acids are highlighted in blue. Gaps are introduced as broken lines to optimize alignment. The AAA family consensus sequence is indicated by thick red line above the aligned sequence and, with respect to paraplegin sequence, is located between residues 344 and 534 with the ATP-binding motif (<ATP>) indicated between residues 349 and 356. The AAA minimal consensus or the second region of homology (SRH) is also indicated. The zinc-dependent binding domain (HESGHA) is indicated (<Zn>) between amino acids 574 and 579 of paraplegin. Four conserved amino acid changes which were observed in paraplegin in the patients with HSP are indicated as G386S, R485(del), A510V, and K569I.

356, and the AAA family consensus sequence located between residues 344 and 534. The zinc-dependent binding domain (HESGHA) was found between amino acids 574 and 579. The protein localization prediction by the program PSORT suggested a mitochondrial localization. Two transmembrane segments of T12 were also predicted between amino acids 144 and 162, and between amino acids 249 and 272, which were situated within the less conserved N-terminal region. Sequence alignment of T12 protein and its yeast homologue as well as human paralogue is presented in **Figure 4.8**.

At the completion of this work Casari *et al* (1998) reported an identical gene called “**SPG7**”, that was involved in an autosomal recessive type of hereditary spastic paraplegia (HSP), a neurodegenerative disorder. The protein product of *SPG7* was designated “**paraplegin**”.

#### 4.3.4 Genomic Organization

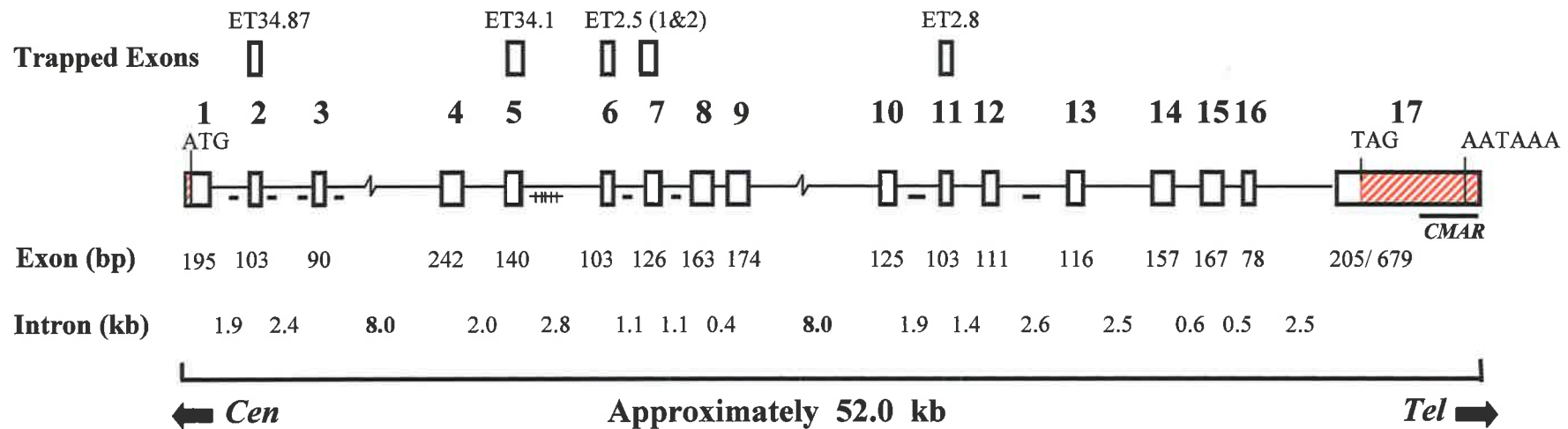
Since the report of Casari *et al* (1998) did not provide data of gene structure of *SPG7* and it was not present in the public database, characterization of the gene structure was, therefore, undertaken.

Because the complete genomic sequence encompassing this region was not available at this time, the gene structure was characterized by PCR amplification using exon sequences of the genomic clones and direct sequencing (section 4.2.4). Alignment of the transcript sequence with this resulting genomic sequence determined that the *SPG7* gene consists of 17 exons ranging in size between 78 and 242 bp. This gene encompasses genomic DNA of approximately 52 kb (**Figure 4.9**) which is contained within three overlapping cosmids: 383H6, 427B11, and 358D12. These three cosmids are part of a cosmid contig of the physical map established at chromosome region 16q24.3 (Whitmore *et al.*, 1998a). As shown in the **Table 4.3**, all splice junction boundaries conform to the

**Table 4.3**

**Nucleotide Sequences of the *SPG7* exon and intron boundaries**

<b>Exon</b>	<b>3' Splice site (intron/exon)</b>	<b>Nucleotide positions</b>	<b>5' Splice site (exon/intron)</b>
1	tttcaggccaac <b>ATGGCC</b>	(-12)-183	CTGCAG <b>gt</b> aaatccccgc
2	tttttttttc <b>agAGCTTA</b>	184-286	TTTTAG <b>gt</b> atgtatctgt
3	tatttctcat <b>agGTGGTA</b>	287-376	ACGAAG <b>gt</b> atattcatct
4	ctgccggctc <b>agAGGAGA</b>	377-618	CGGCCT <b>gt</b> gagtgagggc
5	tctgcccccc <b>agCGGCTA</b>	619-758	TGGAAA <b>gt</b> atggttgatg
6	ttcttccggc <b>agTGCCCT</b>	759-861	GCTTTT <b>gt</b> aagttctgta
7	ctttctgcac <b>agAATCAG</b>	862-987	CTGAAG <b>gt</b> gagagcagca
8	gctgccgtcc <b>agAGCCCA</b>	988-1150	TTGGAG <b>gt</b> aggtgctgtg
9	gtgtcccctc <b>agGCCTCG</b>	1151-1324	TGGATG <b>gt</b> cagtgctcgt
10	ctgtcccctc <b>agGAATGG</b>	1325-1449	CTGCAG <b>gt</b> cagagccagg
11	tggttctggc <b>agGAGAGG</b>	1450-1552	TCAGTG <b>gt</b> acttcctcaa
12	ttgtgttgac <b>agGGGCTG</b>	1553-1663	TCGCAG <b>gt</b> acagggggcg
13	gcattctttc <b>agGGACTG</b>	1664-1779	ATGAAG <b>gt</b> gggtcttggc
14	accttggtgcc <b>agGTCTCC</b>	1780-1936	CTTCTG <b>gt</b> gaggagcagc
15	cccgccctcc <b>agGGGCAC</b>	1937-2103	GACCAT <b>gt</b> gagtcggctc
16	ctccactcac <b>agGAAGCA</b>	2104-2181	CAGGCG <b>gt</b> gaggccctgg
17	tgttctttct <b>agCTGGCA</b>	2182-3065	CAAGTAG <b>gt</b> ttggg-3'UTR
<b>Splicing consensus</b>	<b>ttctttncagG cctccc</b>		<b>AGgtaagt g</b>



**Figure 4.9** : Genomic structure of *SPG7*. *Open boxes* indicate translated portions of the gene. Both 5' and 3' untranslated regions (UTR) are denoted by the *hatched boxes*. Intron size is estimated from the PCR product and is partially sequenced from both ends. Three introns (introns 8,14, and 15) have been completely sequenced. The approximate sizes of introns 3 and 9 have been estimated from the restriction map of the cosmid contig. The *thick line* lying under introns indicate the *Alu*-repeat elements and +++++ under intron 5 represents repeat units of 5'GTCTCAGCCCCCGACGT 3'. The corresponding trapped exons used in identifying the transcript are also indicated: ET34.1 (accession no. AF039807) was aberrantly trapped from exon 5 missing nucleotides 727-758, whereas ET2.5 (accession no. 039804) was trapped from two spliced exons, viz. Exons 6 and 7. *CMAR* indicates the position of the previously reported *CMAR* sequence (Pullman and Bodmer, 1992) within the *SPG7* 3'UTR

published consensus sequences for donor and acceptor splicing sites (Shapiro and Senapathy, 1987). The sequence data of each exon and at least 120 bp of its flanking introns have been deposited with the GenBank Database under accession nos. AF080511 to AF080525. The 5' UTR and a translation initiation codon are contained in exon 1. The predicted transmembrane regions of the protein, amino acid residues 144-162 and residues 249-272, are encoded by exon 4 and exon 6, respectively. The AAA family consensus sequence is coded for by the sequence from exon 8 to exon 12, with the ATP-binding domain (GPPGCGKT) being encoded by exon 8. Exon 13 codes for the protein region containing a zinc-binding motif (HESGHA). The last exon, exon 17, contains part of the coding sequence, stop codon (TAG), and the 680-bp 3' UTR. Other notable sequence features within T12/SPG7 are *Alu* repeat elements within introns 1, 2, 3, 6, and 7 (and possibly introns 10 and 12), and a series of tandem repeat units of 17 nucleotides (GTCTCAGCCCCCGACGT) in intron 5. From the available genomic structure information, the trapped exon ET34.87 was equivalent to exon 2 whereas ET34.1 appeared to be derived from part of exon 5, but missing nucleotides 727-758. This was likely due to a cryptic splicing site at nucleotide position 727. ET2.5 was trapped from two exons, 6 and 7. The trapped product ET2.8 was from exon 11. The *Alu* sequence insertions seen in the 2.4 kb-CMAR cDNA were derived from part of the introns between the corresponding exons 10 and 11, and between exons 12 and 13 of T12/SPG7. These exon/intron boundary sequences provided the basis for mutation study of SPG7 in patients with hereditary spastic paraplegia.

#### 4.3.5 SPG7 mutation study

The frequency of *SPG7* mutations in sporadic spastic paraplegia was not known. Therefore a total of 116 unrelated patients with the diagnosis of spastic paraplegia of both

**Table 4.4**  
**Oligonucleotide primers used for SSCP analysis of *SPG7***

<i>Exon</i>	<i>Sequences of oligonucleotide primers</i> (5'→3')		<i>Product size</i> (bp)
1	<i>F</i>	cac gca ggc gcg gct ttc	271
	<i>R</i>	gct ggg cct tac aga gca g	
2	<i>F</i>	tgg tgt gac ctc cag tat tg	190
	<i>R</i>	gcg ata agt gtg cag tca tg	
3	<i>F</i>	agg agt aca ctg ttg tcc tg	211
	<i>R</i>	atc cag agg cag cta ctg ac	
4	<i>F</i>	ccg tgt ctg ttg ctc atg tg	366
	<i>R</i>	agc ctc act ctc aca ggc tg	
5	<i>F</i>	act gta ggg ttg ctc gtc tg	267
	<i>R</i>	ctg ttt ctc aga tta caa agc c	
6	<i>F</i>	cat tgc cag cag tgg ttc tg	247
	<i>R</i>	gta ttc agc aaa cac aaa cca g	
7	<i>F</i>	ggc atc gtg ctg ctg att tc	282
	<i>R</i>	att cca ggt ccc gca gga g	
8	<i>F</i>	tga ccc aga gag acc tta cc	300
	<i>R</i>	aaa gga gcc agg tag tca gc	
9	<i>F</i>	ggg tac agg aag agg ctt tg	292
	<i>R</i>	cgg aaa acc tcc tct gat gg	
10	<i>F</i>	agg gga aat ctg ttg tgt cag	254
	<i>R</i>	cca gac cac tca gag cga g	
11	<i>F</i>	gca cct gtg gca gta act ag	236
	<i>R</i>	tga tgc tgt ttg cgc agg tg	
12	<i>F</i>	agc cct gat agc aga aac cc	245
	<i>R</i>	cac ctc tca ata cct gcc tg	
13	<i>F</i>	att aca ggc gtg agc cac c	207
	<i>R</i>	ttt ggc caa ggc ccg tgg	
14	<i>F</i>	tga cct ggg tca tct tga cc	228
	<i>R</i>	aag cgc tct gaa acc tcc ag	
15	<i>F</i>	gga tgc ctc tgt ctc gac c	270
	<i>R</i>	cct tgt gtg gta gac cca g	
16	<i>F</i>	ctg aca cag ttc cct cca c	193
	<i>R</i>	ttg gaa ata ctt cac gag agg	
17	<i>F</i>	tgg gga ctc aca cac tgc ta	265
	<i>R</i>	acg tgc agc caa cac ctc c	

**Table 4.5**  
**Summary of SPG7 mutation study in HSP**

Exon	Base/Codon change	Amino acid change	Consequence	Frequency	
				Patients N=116	Control N=60
1	120G→A (GGG→GGA)	No change	None	1	0
4	IVS3-8C→T	None	None	1	0
	IVS4+12C→T	None	None	1	0
5	679C→T (CGA→TGA)	<b>R227STOP</b>	<b>Truncation</b>	1	0
6	762C→T (GCC→GCT)	No change	None	1	ND
7	IVS6-34G→T	None	None	2	1
	IVS7+5A→G	None	None	12	7
8	1014C→T (GGC→GGT)	No change	None	1	0
	1032C→T (GGC→GGT)	No change	None	3	1
9	1156G→A (GGC→AGC)	<b>G386S</b>	<b>Missense</b>	1	0
	IVS9+10C→T	None	None	15	7
10	IVS10+19A→G	None	None	10	6
11	1457G→A (CGG→CAG)	<b>R486Q</b>	<b>Missense</b>	1	0
	1507A→G (ACC→GCC)	<b>T503A</b>	polymorphism	24	13
	1450-1458del9	<b>ERR(484-486)del</b>	<b>3 aa deletion</b>	1	0
	1529C→T (GCA→GTA)	<b>A510V</b>	<b>Missense</b>	1	0
12	IVS12+13C→T	None	None	11	6
13	1706A→T (AAA→ATA)	<b>K569I</b>	<b>Missense</b>	1	0
	1770C→T (GCC→GCT)	No change	None	1	0
	IVS12-15C→A	None	None	1	0
15	2063G→A (CGG→CAG)	<b>R688Q</b>	polymorphism	21	12
17	2292C→T (ATC→ATT)	No change	None	1	ND
	2295C→T (GAC→GAT)	No change	None	1	ND

*ND: not determine*



pure and complicated forms were analyzed. Mutation screening was by PCR-SSCP analysis of genomic DNA isolated from peripheral blood leukocytes. Briefly, radioactive-labeling PCR product of each *SPG7* exon was applied to SSCP gel and followed by electrophoresis (section 4.2.5). The primer pairs used to amplify each *SPG7* exon were generated from the flanking intron sequences, shown in **Table 4.4**. Any PCR product showing a mobility band shift as compared with most of the samples and normal controls were sequenced to determine the basis of nucleotide changes. A number of mutations and sequence polymorphisms were detected in *SPG7* and these are summarised in **Table 4.5**.

#### 4.3.5.1 Nonsense Mutation in Exon 5

The 679C→T mutation in exon 5 was detected in the patient **S26-3** and was heterozygous with the wild type allele (**Figure 4.10 A**). This mutation (679C→T) converts the codon 227 (CGA) for arginine (R) into a potential **stop codon** (TGA) and thus truncates up to 72% of the open reading frame. This would remove all of the functional AAA domain and proteolytic region from the paraplegin molecule. Furthermore, the truncated protein product is probably conformational unstable and likely to be degraded by the cellular proteolytic machinery. Sequencing of other exons failed to detect any additional mutations as a compound heterozygote in this patient. Polymorphism was also detected in this patient, S26-3, in the intron 4 (IVS4+12C/T), intron 6 (IVS6-34T/G) and intron 7 (IVS7+5G/A). To investigate whether such mutation <sup>is a disease?</sup> associated with the disease phenotype, DNA samples from both parents and from all sibs of patient S26-3 were also analysed for sequence changes in *SPG7* exon 5 as well as other exons. As shown in **Figure 4.11**, the pedigree of family S26 showed the segregation of the exon 5 nonsense mutation from the unaffected father to two out of three affected sons and one out of four unaffected sons. Similar to patient S26-3, no additional mutations were observed in other members of

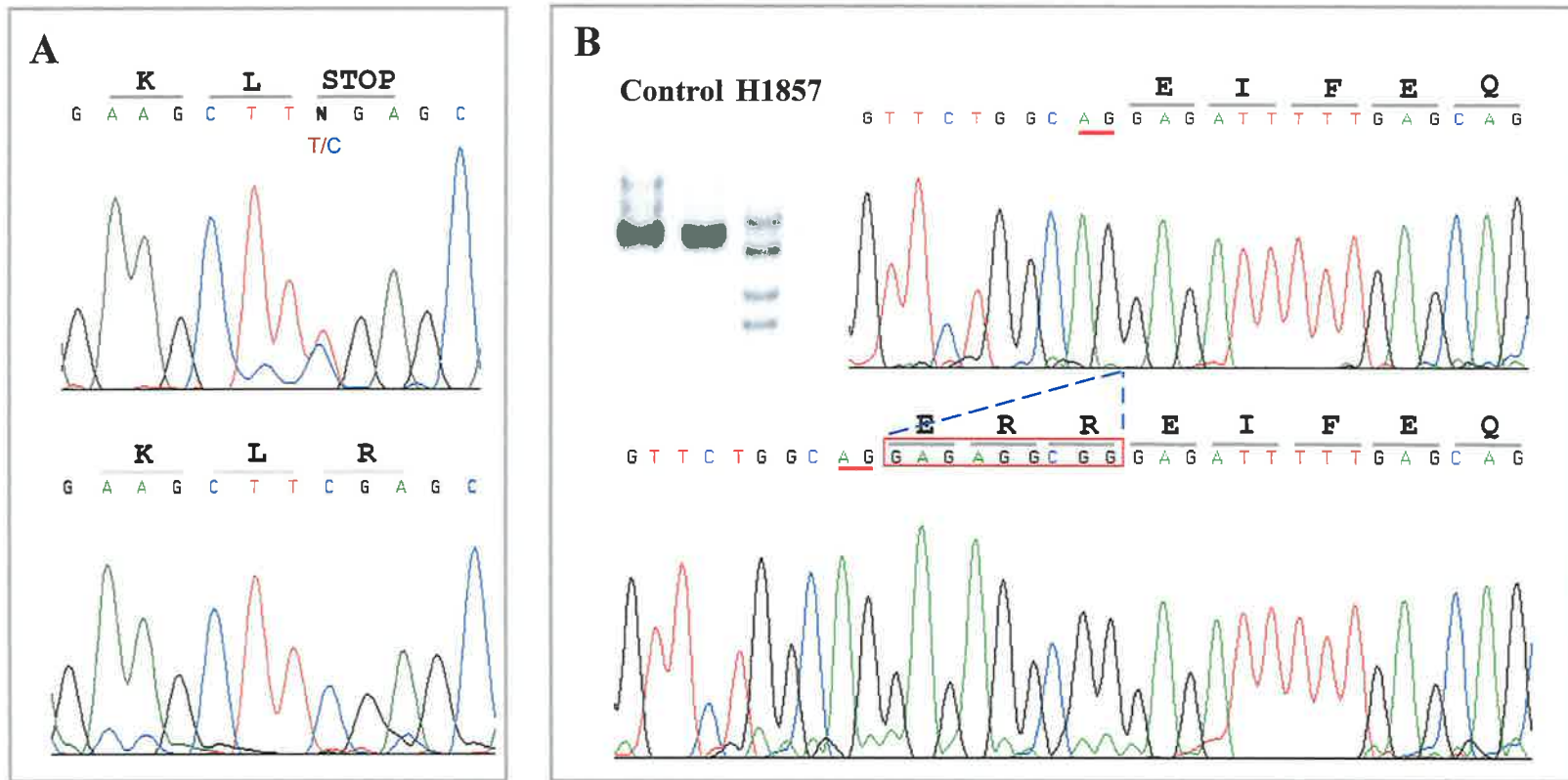
this family. Since one of the affected sons carried no mutation of exon 5 or other exons, mutation of *SPG7* may not be the major cause of disorder in this family. However, by its function in the integrity of mitochondria biogenesis, *SPG7* mutation may play indirect role in, or increase the risk of spastic paraplegia in this family.

#### 4.3.5.2 Compound Heterozygous Mutation in Exon 11

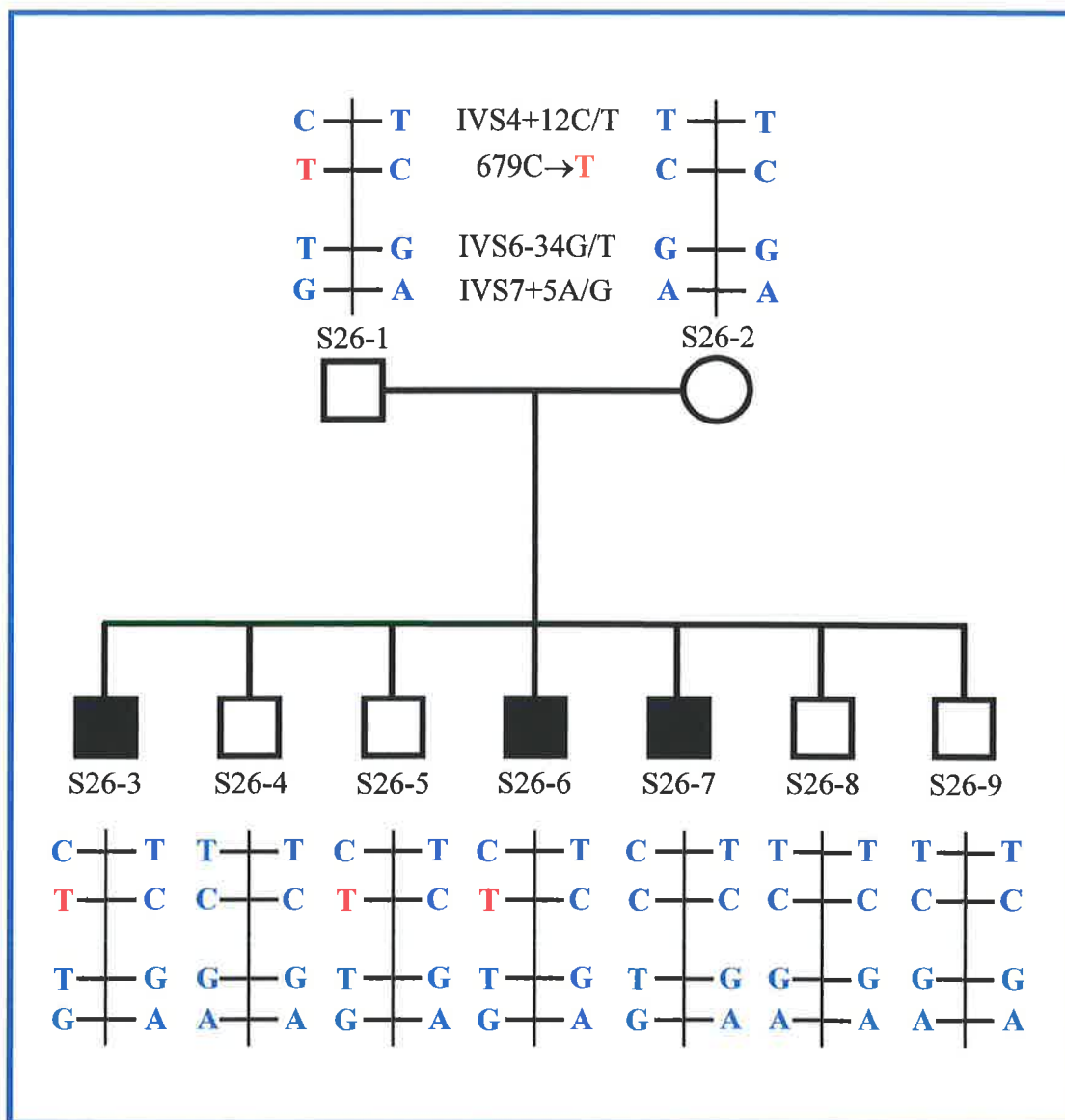
Patient **H1857** with complicated spastic paraplegia was found to be a compound heterozygote, 1450del9 and 1529C→T, both changes in exon 11. To confirm that the two mutations were heterozygous, the exon 11 PCR products were amplified from the patient's genomic DNA, cloned into a plasmid vector, and six clones sequenced. Only products containing either of the single mutation were found. The 9 bp deletion from nucleotides 1450-1458 in *SPG7* cDNA gives rise to the removal of three amino acid residues, glutamine (E), arginine (R), and arginine (R) (aa 484-486) from the paraplegin protein (**Figure 4.10 B**). In addition, the nucleotide transition 1529C→T causes an amino acid change at residue 510 from alanine (A) to valine (V). Both mutations are located within the AAA consensus domain of paraplegin. Alignment identity of R485 and A510 with yeast Afg3p, Rca1p and Yme1p and bacterial Ftsh1p suggest that these residues are functionally conserved throughout evolution. Both changes were not detected in the other 115 SPG samples or in 50 normal controls. Therefore, such compound changes, the removal of R485 by 9 bp deletion and the amino acid substitution, A510V, could contribute to the clinical phenotype.

#### 4.3.5.3 Missense Mutations in Exon 9 and Exon 13

Two missense mutations, 1156G→A in exon 9 and 1706A→T in exon 13, were separately found in two patients and both were heterozygous with wild type allele.



**Figure 4.10** : Mutation analysis of *SPG7*. **A)** Partial sequencing profiles of *SPG7* **exon 5** from patient S26-3 (top) and normal control (bottom). A single nucleotide mutation, C → T, is indicated in patient sequence (heterozygous T / C) and converts the codon 227 for arginine to stop codon. **B)** SSCP gel electrophoresis and partial sequencing profiles of *SPG7* **exon 11** PCR products from patient H1857 and normal control. The gel (upper left) shows mobility band-shift of PCR product from patient as compared to those of normal. Nine base pair deletion of *SPG7* at the 5' boundary of exon 11 is seen in the patient (top) but not in normal control (bottom). AG (underlined) indicate the splice acceptor site at the intron boundary. Deletion of nine base pairs removes three amino acid residues, glutamine (E), arginine (R), and arginine (R), at positions 484-486 from paraplegin.



**Figure 4.11** Pedigree of family S26 with the Genotype of *SPG7*. The symbols indicated are *opened square*, unaffected male; *filled square*, affected male; and *opened circle*, unaffected female. Sequence changes are indicated in intron 4 (IVS4+12C/T), exon 5 (679C→T), intron 6 (IVS6-34G/T), and intron 7 (IVS7+5A/G).

The first, 1156G→A transition replaces the glycine (G) residue 386 with serine (S), a larger uncharged polar amino acid. Glycine at this position is within the AAA domain and is strongly conserved between paraplegin and its orthologous yeast protein (**Figure 4.8**). This change was detected only in the patient **PC12** but not in other patients or in normal controls. On the basis of the nature and location of this amino acid changes, it is possible that it is of functional significance. Study of the family members of patient **PC12** showed that this change was inherited from his unaffected mother and was also present in an unaffected sister. Therefore, this change is likely to represent a rare and unique variant for this family and is unlikely to be a disease causing mutation, at least when heterozygous.

The second heterozygous change, 1706A→T, in exon 13 was detected only in the patient **PC55** and results in amino acid substitution at residue 569, lysine (K) to isoleucine (I) (K569I). Lysine is a basic amino acid while isoleucine is a non- polar amino acid. The lysine at this position is conserved between paraplegin and the yeast proteins Rca1p and Yme1p, while in Afg3p this residue is arginine (R). Arginine is also a basic amino acid and therefore shares a similar physico-chemical property with lysine. Since amino acid residue 569 is in the close proximity to the zinc- binding motif critical for the protease activity (**Figure 4.8**), the change from basic to non-polar amino acid would therefore be expected to affect the function of this domain. The family members of patient **PC55** were not available for studies and, hence, it is not known whether this change is functionally significant when heterozygous.

A number of single base substitutions which cause no amino acid change were also found in exons 1, 6, 8, 13, and 17 (**Table 4.5**) and were likely to be low frequency polymorphism though not detected in normal controls.

#### **4.4 Discussion**

The present study describes the isolation and characterization of a transcript sequence of 3,083 bp, which was later found to be identical to the subsequently reported *SPG7*, a gene involved in a recessive type of hereditary spastic paraplegia (HSP) (Casari *et al.*, 1998). Initially, from the physical map of chromosome region 16q24.3 (Whitmore *et al.*, 1998) six trapped exons were grouped together as part of a tentative gene, designated "transcription unit 12 (T12)". Upon analyses, however, only four: ET34.87, ET34.1, ET2.5, and ET2.8 were successfully joined and formed part of a transcript sequence. Northern analysis indicated the full-length transcript to be approximately 3.2 kb. The sequence was extended by utilizing both PCR from a 5' stretched cDNA library and 5' RACE. This extended cDNA sequence contained an in-frame translation initiation codon, ATG, perfectly embedded in a Kozak consensus sequence, and therefore suggested a true start site. In contrast, the 3' sequence extension was initially complicated by the finding that overlapping EST sequences were derived from genomic DNA that contaminated the cDNA library. 3' RACE was then utilized but this only allowed a short sequence extension. However, it was then determined that this 3' sequence of T12 overlapped with the 5' part of a newly reported *CMAR* cDNA sequence of 2.4 kb (Durbin *et al.*, 1997). RT-PCR experiments were undertaken using T12 primers and primers designed from *CMAR* sequence to confirm that they were part of the same transcript. These results finally allowed the assembly of a 3,083-bp transcript sequence of T12 and also provided the evidence that *CMAR* was in fact a part of the 3' UTR of this transcript.

The *CMAR* gene has originally been reported as an intronless gene with its 459-bp cDNA that has been predicted to code for a protein of 142 amino acids (Pullman and Bodmer, 1992). This gene has been demonstrated by transfection experiment to enhance binding of the cell to extra-cellular matrix component, collagen type1 (Pullman and

Bodmer, 1992). This *CMAR* sequence is situated close to the 3' end of the 3,083-bp T12 transcript. From the initial report, the CMAR protein structure has been described to be integrin  $\beta$ 1 dependent, with probable tyrosine phosphorylation activation, suggesting a role in signal transduction (Juliano and Haskill, 1993; Romer *et al.*, 1992; Novelli *et al.*, 1995). Furthermore, since this gene has been mapped to the 16q24.3 LOH region observed in several cancers, it has been proposed that CMAR might function as a tumor suppressor. A subsequent publication by Durbin *et al.* (1997) presents a larger *CMAR* transcript of 2.4 kb, but the ORF was the same as of the original published sequence. This 2.4-kb transcribed sequence is interrupted with the insertion of two *Alu*-like elements within its 5' half. The original reported coding sequence (Pullman and Bodmer, 1992) is situated within the 3' region of the 2.4-kb transcript. Within this coding region the CACA insertion polymorphism has been reported and resulted in the change in the putative ORF (Koyama *et al.*, 1993; Durbin *et al.*, 1994). However, none of these reports of *CMAR* transcript have presented evidence of a protein product or a decent Northern result.

The 3,083-bp transcript of T12 was predicted to code for a protein of 795 amino acids. This protein was predicted by sequence homology to have functions similar to the yeast mitochondrial proteins, Afg3p, Rca1p, and Yme1p. In addition, the protein localization prediction by PSORT program also suggested a mitochondrial protein. Shortly following completion of this result, Casari *et al* (1998) reported the gene, *SPG7*, which was identical to T12 and coded for the same 795 aa-protein designated "**paraplegin**". The result of *in silico* prediction by PSORT also agreed with that of *in vitro* protein localization study of the *SPG7* protein product presented in this report (Casari *et al.*, 1998). The orthologous yeast proteins are ATP-dependent metalloproteases belong to a subclass of the AAA protein family, the AAA proteases. They possess both chaperone-like and proteolytic activities (Rep and Grivell, 1996; Patel and Latterich, 1998). There are three conserved

functional domains characteristic of this subgroup of AAA protein: the ATP-binding motif (GPPGCGKT), AAA minimal consensus sequence, and the zinc-dependent binding motif (consensus HEXXHA). The AAA consensus domain is crucial for the chaperone-like activity in sensing and binding with the substrate (Langer, 2000; Juhola *et al.*, 2000). This domain is followed by a region responsible for proteolytic activity with the consensus metal-binding motif HEXGHA as the proteolytic center (Langer, 2000; Juhola *et al.*, 2000). All of these functional domains were also found in paraplegin. The function of such yeast mitochondrial metalloproteases involves in the quality control of mitochondrial protein biogenesis, including protein folding and the assembly of respiratory chain complexes (Rep and Grivell, 1996; Patel and Latterich, 1998). Defects of these mitochondrial metalloproteases are likely to be lead to the deterioration of mitochondria function.

In the yeast mitochondrial inner membrane, Afg3p and Rca1p form subunits of a hetero-oligomeric proteolytic complex at equimolar amounts (Arlt *et al.*, 1996) with its active site present at the matrix side. Both Afg3p and Rca1p are anchored to the membrane by two transmembrane domains in their N-terminal regions. In contrast, Yme1p is a homo-oligomeric complex (Leonhard *et al.*, 1996) with its catalytic sites facing the intermembrane space. The Yme1p is anchored by only one transmembrane segment. In analogy to the Afg3p and Rca1p, paraplegin is likely to be anchored to the membrane by two transmembrane domains in its N-terminal region, which were identified between amino acid residues 144-162 and between residues 249-272 (section 4.3.3). This similarity suggests that paraplegin function requires other protein partners to form a functional complex. The protein partners of paraplegin have not yet been reported. Two recently identified human genes, *AFG3L2* (Afg3-like 2) and *YME1L1* (Yme1-like 1) on chromosomes 18 and 10 respectively, have also been reported to encode members of the



mitochondrial AAA metalloprotease (Banfi *et al.*, 1999; Coppola *et al.*, 2000). The AFG3L2 protein sequence is highly homologous to both Afg3p and Rca1p and similar to that of paraplegin (Banfi *et al.*, 1999). AFG3L2 has also been reported to localize to mitochondria. It was likely that AFG3L2 may have function by forming a complex with paraplegin, however this has yet to be determined from molecular interaction analysis. *YME1L*, on the other hand, appears to be an orthologue of yeast *Yme1* and its expression in yeast has been able to complement a *yme1* null mutant (Shah *et al.*, 2000). Therefore this is unlikely to be a protein partner of paraplegin.

The genomic structure of *SPG7* was determined based on the *T12* transcript sequence and three genomic clones spanning the *T12/SPG7* region: cosmids 383H6, 427B11, and 358D12. The gene consists of 17 exons encompassing the genomic sequence of approximately 52 kb. Based on the exon/intron boundary information, a cryptic splice site was identified in exon 6. During the course of transcript sequence identification, a cDNA clone  $\lambda$ ET2.5 containing a sequence between exon 4 and exon 9 of *SPG7* was isolated from a fetal brain cDNA library. Sequence analysis revealed a 55-bp deletion of the sequence corresponding to the 3' part of exon 6. Analysis of the complete exon 6 sequence identified a cryptic donor splice site (TGgtatgt: consensus score 81) situated 55 bp upstream from the 3' end of the exon (**Figure 4.1**). Therefore, the 55 bp deletion from exon 6 in this cDNA clone is probably attributable to the use of this cryptic donor splice sequence alternative to the normal site (TTgtaagt: score 79). This deletion however truncates the ORF of the transcript by creating an immediate in-frame stop codon. This aberrant form of transcript is not detected by RT-PCR and is considered to be a rare transcription isoform with no functional significance.

Mutation studies of *SPG7* were undertaken in 116 unrelated cases of spastic paraplegia, both pure and complicated forms. Two cases were found that have mutations

and were thought to be related to the clinical phenotype. A nonsense mutation was detected in the patient S26-3, an index case, diagnosed with a pure form of HSP. The family of this index case consists of normal parents, three affected sons, and four unaffected sons. The familial pattern of disease transmission was referred to as autosomal recessive based on the presence of unaffected parents and absence of family history. The *SPG7* mutation identified in individual S26-3 was a single base substitution, 679C→T, resulting in the change of codon 227 from CGA for arginine to a potential termination codon, TGA. This change abolishes up to 72% of the coding sequence and therefore removes all functional motifs of the protein. The truncated protein product, if exist, is probably conformational unstable and likely to be degraded by the cellular proteolytic machinery. The most likely mechanism is the degradation of mRNA that contains premature termination of ORF by the pathway referred to as nonsense-mediated mRNA decay (NMD) (Culbertson, 1999; Frischmeyer and Dietz, 1999). This mutation is therefore pathogenic. The mutation, however, is heterozygous with no other mutations detected to be a compound heterozygote. Sequencing of this gene in the parents and all sons indicated that the 679C→T mutation was inherited from the unaffected father to two out of three affected sons and one out of four unaffected sons. When heterozygous this mutation would lead to the reduction of paraplegin synthesis. However, it was unknown whether this contributed to the clinical phenotype in this family. Both the father and one of his son (S26-5), were heterozygous for this mutation but unaffected, could be caused by a dominant gene with low penetrance. However, the presence of an affected son with no *SPG7* mutations indicated that it is likely that a defect of other gene may be the cause of the clinical phenotype in this family. Since the paraplegin protein is involved in the integrity of mitochondrial function, it is possible that *SPG7* mutations even when heterozygous may indirectly contribute to the

pathogenesis of disease or increase the risk of disease in an individual who carries such mutation.

A compound heterozygous mutations of exon 11, 1450del9 and 1525C→T, was observed in case H1857. Both mutations are in the conserved AAA domain that is important for paraplegin function. The nine base pair deletion, nt 1450-1458, results in the loss of three amino acids, glutamine, arginine, and arginine (residues 484-486) from the highly conserved region of AAA domain (**Figure 4.8**). The removal of these amino acids is predicted to affect the protein conformation of this domain and therefore affect the paraplegin function. The AAA domain is important in sensing and binding of paraplegin to its substrate. This is a chaperone-like activity. In addition, the 1529C→T mutation replaces alanine at amino acid residue 510 with valine (A510V) and also occurs within the conserved sequence of AAA domain. This change would also affect the AAA motif function. The mutations as a compound heterozygote are likely to compromise the function of the paraplegin protein and therefore be consistent with a sporadic case of recessive spastic paraplegia. The patient H1857 was clinically diagnosed as a complicated spastic paraplegia and developed the clinical disease at the age of 25. Both of his parents were reported to be clinical normal but there were no clinical data of other family members and they were unavailable for study. The same compound heterozygous changes, 1450del9 and 1525C→T, were also presented in one of 30 index cases, studied by McDermott *et al* (2001). This case was also complicated spastic paraplegia. In this family, the father who was heterozygous for the 1450del9 also presented with a milder form of complicated spastic paraplegia, suggesting a dominant behavior of this mutation.

Two heterozygous missense mutations, 1156G→A in exon 9 and 1706A→T in exon 13 were found in patients PC12 and PC55 respectively but were not presented in at least 50 normal controls. Both changes cause amino acid substitutions affecting a conserved amino

acid. The 1156G→A results in the replacement of glycine at residue 386 with serine (G386S). The 1706A→T mutation on the other hand causes amino acid substitution, lysine by isoleucine (K569I). The amino acid substitutions, G386S and K569I, affect the conserved amino acid within the AAA domain, and the conserved residue close to a zinc-binding motif respectively. However, both variants are heterozygous with no other mutations detected in this gene. When analysis was performed in the family members of either case, such mutations were also detected in the clinical normal individuals. Therefore, both were unlikely to be related to a clinical phenotype at the heterozygous stage. It is possible, however, that carriers are at increased risk of developing diseases.

Casari *et al* (1998) reported three different *SPG7* mutations in three families of which all affected cases were homozygous. The first mutation was a 9.5 kb genomic deletion covering the last five exons of *SPG7*, including its 3'UTR. This deletion is likely to interrupt the transcription process of the gene due to the absence of polyadenylation signal within 3'UTR of the gene and the down stream sequence element necessary for the transcription machinery (Lou *et al.*, 1995; Lou *et al.*, 1996). The second was homozygous 2-bp deletion at nucleotide 784-785 (in exon 6) which causes a frame shift that changes the amino acid sequence down stream and creates a "TAA" stop at the first codon of exon 7. This abolishes about 60% of the ORF. The resulting mRNA is likely to be degraded by the mechanism called "nonsense-mediated mRNA decay (NMD)" (Culbertson, 1999; Frischmeyer and Dietz, 1999). The last family was homozygous for an "A" insertion at position 2228 within exon 17 of *SPG7* cDNA. This mutation causes a frame shift and creates a premature termination of the ORF which is predicted to code for a truncated form of paraplegin that is missing 57 amino acids at the C terminus.

The three mutations reported by Casari *et al* (1998) were from a screen of 14 unrelated recessive families. This suggested that mutations in this gene may be a relatively

common cause of autosomal recessive HSP. The present study, however, demonstrates that *SPG7* mutations are rare in sporadic cases and small families with HSP. This is likely the result of the mode of ascertainment. In the study by Casari *et al* (1998), each case was from a single ethnic group in a relatively close small community. All three mutation-families were consanguineous. In the present analysis, in contrast, all subjects are from a population with a diverse ethnic mix. In recessive families, there is a low frequency of linkage to *SPG 7* that has been reported in other studies (Coutinho *et al.*, 1999). This also suggests that *SPG7* mutations are a rare cause of spastic paraplegia.

Finally, the expression of *SPG7* was found ubiquitously in all the tissues tested, consistent with its mitochondrial function. Mutation of this gene in other syndromes, which have a causal basis of mitochondrial impairment may also exist. Spastic paraplegia has also been reported in association with other neurologic disorders, such as Alzheimer's disease (Orth and Schapira, 2001; Turner and Schapira, 2001) and some types of epilepsy (Gigli *et al.*, 1993; Yih *et al.*, 1993; Webb *et al.*, 1997; Nobile *et al.*, 2002). It is possible that *SPG7* mutations and the resulting defects in paraplegin may contribute to these diseases, though this requires additional study.

## **Chapter 5**

# Exon Amplification Study and Characterization of Transcription Unit

13 (T13)

---

## **Table of Contents**

	<b>Page</b>
<b>5.1 Introduction</b>	145
<b>5.2 Methods</b>	146
5.2.1 Exon Trapping	146
5.2.1.1 <i>Digestion of BAC DNA</i>	147
5.2.1.2 <i>Purification of digested DNA by phenol/Chloroform Extraction</i>	147
5.2.1.3 <i>Preparation of the Exon trapping Vector</i>	148
5.2.1.4 <i>Cloning of BAC digested products into the pSPL-3B Vector</i>	148
5.2.1.5 <i>Preparation of COS-7 Cells</i>	149
5.2.1.6 <i>Transfection of COS-7 Cells</i>	149
5.2.1.7 <i>Isolation of Cytoplasmic RNA from COS-7 Cells</i>	150
5.2.1.8 <i>Reverse transcription of Isolated RNA</i>	150
5.2.1.9 <i>Primary PCR Amplification</i>	151
5.2.1.10 <i>Secondary PCR Amplification</i>	151
5.2.1.11 <i>Uracil DNA Glycosylase Cloning of Secondary PCR Products</i>	152
5.2.2 Exon Confirmation	152
5.2.2.1 <i>Colony PCR</i>	153
5.2.2.2 <i>DNA Isolation and Sequencing</i>	153
5.2.2.3 <i>Physical Mapping of Trapped Products</i>	153
5.2.2.4 <i>Trapped Exon Sequence Analysis</i>	154
5.2.3 Identification of Transcript Sequence	154
5.2.4 <i>In silico</i> Protein Function Analysis	154
<b>5.3 Results</b>	155
5.3.1 Exon Trapping	155
5.3.2 Trapped Exon Sequence Analysis	155
5.3.3 Physical Mapping of Trapped Products	158
5.3.4 Verification of the Transcript Sequence	159
5.3.4.1 <i>Transcript Sequence generated from 561-4 and 561-13</i>	159
5.3.4.2 <i>Transcript Sequence of 561-72 and 561-114</i>	162
5.3.5 Assembly of the Transcription Unit 13 (T13) Sequence	163
5.3.6 Analysis of T13 Transcript Sequence	166
5.3.7 T13 Protein Sequence Analysis	166
5.3.8 T13 Gene Structure	168

5.3.9 T13 Alternative Splicing and Overlapping Transcripts	169
5.3.10 Mutation Study of T13 in Sporadic Breast Cancer	171
<b>5.4 Discussion</b>	172
5.4.1 Exon trapping	172
5.4.2 Characterization of Transcription unit 13 (T13)	175



## **5.1 Introduction**

The construction of the physical and transcription map at 16q24.3 LOH region (Whitmore *et al.*, 1998a) has allowed the isolation and mapping of many new genes and potential transcription units. In addition, the map location of known genes in the region were also established. These included *FAA*, a gene involved in Fanconi anemia complementary group A (The Fanconi anaemia/Breast cancer consortium, 1996); *GAS11*, a human homologue of murine growth arrested specific gene (Whitmore *et al.*, 1998b); transcription unit 3 (*T3*) (chapter 3); transcription unit 6 (*T6*) (Whitmore, 1999 unpublished data); *SPG7*, the gene for recessive spastic paraplegia (Casari *et al.*, 1998; chapter 4), and a recently identified gene, Copine VII (Savino *et al.*, 1999). The map location of the *PISSLRE* gene (Grana *et al.*, 1994) was also established in this region through its sequence homology to trapped exons isolated from the cosmid contig at 16q24.3.

Among these genes, *FAA* and *PISSLRE* were initially considered to be potential candidates as a tumor suppressor targeted by LOH in breast cancer. Patients with Fanconi anemia possess a predisposition to a variety of malignancies such as acute myeloid leukemia and squamous-cell carcinoma (Butturini *et al.*, 1994). Cells from FA patients show increased sensitivity to bi-functional DNA crosslinking agents such as diepoxybutane and mitomycin C, with characteristic chromosome breakage (Auerbach, 1993), a feature of genomic instability found in most cancer. The *PISSLRE* gene on the other hand belongs to a family of cycline dependent protein serine/threonine kinases that are critical for cell cycle progression. Dominant-negative constructs of the gene have been shown to halt cell cycle progression in G2-M phase and, when overexpressed in U2OS cells, suppress growth (Li *et al.*, 1995). Despite the potential involvement in genomic instability and cell growth, both have been excluded by mutation analysis in sporadic breast cancer as tumor suppressors

targeted by LOH (Cleton-Jansen *et al.*, 1999; Crawford *et al.*, 1999).

Other identified genes in this region including *GAS11*, *SPG7*, Copine VII, T6, and T3 have also been excluded by mutation studies as candidate breast cancer related tumor suppressors (Whitmore *et al.*, 1998b; Settasatian *et al.*, 1999; Savino *et al.*, 1999; Whitmore, 1999 unpublished data; Chapter 3).


The proximal region of the physical map established at the 16q24.3 breast cancer LOH region (Whitmore *et al.*, 1998a) consists of a group of overlapping cosmids with five clones representing the minimal tiling path. This group of cosmids was linked to a longer cosmid contig by the genomic clone, PAC 168P16 (**Figure 1.3**). Of five overlapping cosmids, the cosmid 383F1 represents the proximal boundary of this group. From this boundary a BAC clone **561E17** was subsequently identified and extended the contig proximally a further 140 kb (Whitmore, 1999 unpublished data). Within this region a cDNA sequence of clones 118499 and IMAGE:1098126 were initially mapped and were assigned as transcription units, *T17* and *T18* respectively. In addition, the Cadherin 15 gene (*CDH15*) has also been mapped to this region (Kremmidiotis *et al.*, 1998). To analyse this extended region for the identification of new genes, the large genomic clone, BAC 561E17, was chosen for further study.

The present studies will describe the use of exon trapping for exploring genes within this region. Subsequent identification of the genes or potential transcripts from the trapped products will also be described. Characterization of the gene designated transcription unit 13 (*T13*) will be described in detail as the major transcript.

## **5.2 Methods**

### **5.2.1 Exon Trapping**

The principle of the exon trapping procedure (Buckler *et al.*, 1991; Church *et al.*,

1994) is summarised in **Figure 5.1**. The detailed procedures are described in the subsequent sections. The sequences of all primers used for exon trapping procedure are given in Table 

#### **5.2.1.1 Digestion of BAC DNA**

In a reaction volume of 50  $\mu$ l, 5  $\mu$ g of BAC DNA was digested at 37 °C, overnight, with the restriction enzymes *Bam*HI and *Bg*III, or with *Pst*I alone, in the specific buffer containing 100  $\mu$ g/ml of BSA as recommended by the manufacturer. The amounts of restriction enzymes used were 15 units of *Bam*HI, 30 units of *Bg*III, and 15 units of *Pst*I. Since double enzyme digestion was performed in the *Bam*HI specific buffer, which was not optimal for 100% digestion by *Bg*III, the amount of *Bg*III used was doubled.

#### **5.2.1.2 Purification of digested DNA by Phenol/Chloroform Extraction**

Following restriction enzyme digestion, an aliquot of 250 ng of each digested DNA was taken for subsequent analysis. The remaining DNA was adjusted to a final volume of 200  $\mu$ l with sterile water and mixed well. An equal volume of phenol was then added to each reaction tube, mixed and followed by centrifugation at 13,000 rpm for 7 minutes. The aqueous layer was transferred to a new 1.5 ml tube and an equal volume of chloroform/iso-amyl alcohol (24:1) was added and mixed. Each mixed sample was again centrifuged for 7 minutes at 13,000 rpm and the aqueous layer transferred to a clean eppendorf tube. The DNA was precipitated by the addition of one-tenth the volume of 3M NaOAc (pH 5.2) and 2 volumes of absolute ethanol, mixed, and kept at - 80 °C for 1 hour. The samples were then spun at 13,000 rpm for 30 minutes. DNA pellets were washed with 1 ml of 70% ethanol, air-dried for 15 minutes, and the DNA was dissolved in 10  $\mu$ l of sterile water. An aliquot of 2  $\mu$ l of each sample was analysed by electrophoresis on a 0.8% agarose gel. The

250 ng aliquot of the original digested product was also loaded on the gel to determine the efficiency of the purification procedure.

### **5.2.1.3 Preparation of the Exon Trapping Vector**

Ten micrograms of the exon trapping vector, pSPL3B, was digested overnight at 37 °C with 15 units of either *Pst*I or *Bam*HI in a 50 µl reaction buffer. Then, 50 µl of water was added and the digested products purified with QIAquick columns according to the manufacturer's instructions (2.2.14.1). The purified digested vectors were then dephosphorylated with 2.5 units of calf intestinal alkaline phosphatase (CIAP) in a final volume of 100 µl using the specific buffer supplied with the kit (Boehringer Mannheim). The reactions were carried out at 37 °C for 1 hour and followed by a further purification step using QIAquick columns. The DNA was quantitated by spectrophotometry (2.2.4).

### **5.2.1.4 Cloning of BAC digested products into the pSPL3B Vector**

The digested BAC DNA was ligated to a linear and dephosphorylated pSPL3B DNA according to the procedure described in 2.2.15.1. The reaction was performed in a final volume of 10 µl containing 50 ng of vector and 200 ng of digested BAC DNA. The ligation reaction containing 50 ng of digested vector alone was also carried out as a control.

The transformation reaction was undertaken by the procedure described in 2.2.15.4, except for the plating of transformed cells. Instead of plating out 200 µl of the transformed cells, the cells were spun down and the pellet was resuspended in 100 µl of LB-broth. The entire 100 µl was then plated onto a LB-agar/Ampicillin plate, and incubated overnight at 37 °C ( X-gal and IPTG were not spread with the cells). Subcloning of the BAC DNA into trapping vector was considered successful if the number of colonies on the experimental

plate exceeded by 10 folds the number of colonies on the control, vector-only ligation plate.

After overnight incubation, 3 ml of LB-broth plus ampicillin (100 µg/ml) was added onto each successful ligation/transformation plate. The cells were scraped off into the media solution with a sterile glass spreader. The cell suspension was then transferred to a sterile 10 ml tube. DNA isolation was performed by "plamid mini-prep" using Qiagen Tip-20 (2.2.1.3).

#### **5.2.1.5 Preparation of COS-7 cells**

The COS-7 cells were grown in supplemented DMEM culture medium. The cells were passaged one day prior to transfection by placing  $8 \times 10^5$  to  $16 \times 10^5$  cells into 10 ml of supplemented DMEM in 75 cm<sup>2</sup> tissue culture flasks. This resulted in 40 to 60% confluence the next day.

#### **5.2.1.6 Transfection of COS-7 Cells**

The culture flasks were inspected for the appropriate cell confluence and pSPL3B-BAC DNA was transfected into the cells using the LipofectACE reagent (Gibco-BRL). Two subcloned DNA samples, *Pst*I generated and *Bam*HI/*Bgl*II generated, were added to one flask of cells per transfection. The transfection was performed essentially as follows. In a sterile 10 ml tube, 500 µl of Opti-MEM I (OMI) medium was added to 3 µg of each subcloned DNA to be transfected. In a sterile 1.5 ml eppendorf tube, 500 µl of OMI was added to 30 µl of LipofectACE and left at room temperature for 5 minutes. This OMI-LipofectACE mix was then added to the 10 ml tube containing OMI-subcloned DNA and incubated for another 10 minutes at room temperature. Whilst waiting, the supplemented DMEM was removed from the COS-7 cell culture and 10 ml of OMI was added to each

culture flask. The cells were then incubated for 5 minutes at 37 °C in a 5% CO<sub>2</sub> incubator. After 10 minute incubation of the above 10 ml tube was complete, 4 ml of OMI was then added to each tube containing the subcloned DNA/LipofectACE/OMI mix. The 10 ml of OMI medium from each flask of cells was removed after 5 minute incubation and the combined DNA/LipofectACE/OMI above was added to the flask of cells. Following incubation overnight at 37 °C in a 5% CO<sub>2</sub> incubator, the lipid-DNA complex was removed from the COS-7 cells and 10 ml of supplemented DMEM was added and incubated for a further 48 hours at 37 °C in 5% CO<sub>2</sub> environment.

#### ***5.2.1.7 Isolation of Cytoplasmic RNA from COS-7 Cells***

All RNA isolation procedures were done under RNase free conditions as described in section 2.2.2. After 48 hour incubation, the supplemented DMEM was removed from each flask and 10 ml of ice-cold PBS was added and the flask was left on ice. The PBS was then removed from each flask and the wash step was repeated two more times. The washed cells were removed from culture flask by adding 10 ml of PBS and cells were scraped off using a sterile plastic scraper (Costar). The cell suspension was transferred into a sterile 10 ml tube, which was placed on ice, then centrifuged at 1,200 rpm for 5 minutes. After the supernatant was removed, the cell pellet was resuspended in Trizol reagent (Gibco-BRL) and RNA was isolated as described in 2.2.2.2.

#### ***5.2.1.8 Reverse Transcription of Isolated RNA***

The total RNA isolated from COS-7 cells was reverse-transcribed as per the procedure described in 2.2.13.1, but using the pSPL3B specific primer, SA2. The sequence of this primer is presented in *Table 5.1*. In each reaction, 1 µl of 20 µM stock of SA2 was used. After a 30 minute incubation at 42 °C with the Superscript enzyme, the mixture was

heated to 55 °C and maintained for 10 minutes. Two units of RNaseH was then added and the reaction tube was incubated for a further 10 minutes at 55 °C. The reverse-transcribed products were stored at -20 °C while not in use.

#### **5.2.1.9 Primary PCR Amplification**

In the reaction, 8 µl of the reverse-transcribed product was mixed with 4 µl of 10X PCR buffer, 0.8 µl of 10 mM dNTPs, 1.2 µl of 50 mM MgCl<sub>2</sub>, and 2 µl each of a 20 µM stock of oligonucleotides SA2 and SD6 (**Table 5.1**). The mixture was made up to a volume of 39.5 µl with sterile water, and overlaid with a drop of paraffin oil. The tube was heated at 94 °C for 5 minutes, followed by holding at 80 °C while 0.5 µl (2.5 units) of *Taq* DNA polymerase was added. Each reaction tube was then incubated at 94 °C for 1 minute, 60 °C for 1 minute, 72 °C for 5 minutes for 6 cycles, followed by a 10 minute incubation at 72 °C. To prevent the amplification of false-positive (vector-only) products in the subsequent PCR reaction, 2.5 µl (25 units) of *Bst*XI was added to the PCR reaction and incubated overnight at 55 °C. This was followed by incubation with a further 0.5 µl ( 5 units) of *Bst*XI for a further 2 hours at 55 °C.

#### **5.2.1.10 Secondary PCR Amplification**

Each reaction was carried out by combining 5 µl of the first PCR reaction with 4.5 µl of 10X PCR buffer, 1 µl of 10 mM dNTPs, 1.5 µl of 50 mM MgCl<sub>2</sub>, 1 µl each of a 20 µM stock of oligonucleotides dUSD2 and dUSA4 (**Table 5.1**), and sterile water to a final volume of 49.5 µl. The contents were mixed and overlaid with a drop of paraffin oil. After heating the tube at 94 °C for 5 minutes, the tube was maintained at 80 °C while 0.5 µl (2.5 units) of *Taq* DNA polymerase was added. The tube was then subjected to a 30 cycles of

94 °C for 1 minute, 60 °C for 1 minute, and 72 °C for 3 minutes. This was followed by a 10 minute incubation at 72 °C. Five microlitres were analysed on a 2.5% agarose gel electrophoresis to examine the potential number of trapped exons amplified.

In this step the oligonucleotides SD2 and SA4 were modified to include deoxy-UMP residues at their 5' ends (the oligonucleotide sequences are presented in *Table 5.1*). This was to allow cloning of PCR products utilizing the Clone AMP pAMP10 kit (Life Technologies).

#### ***5.2.1.11 Uracil DNA Glycosylase Cloning of Secondary PCR Products***

The PCR products from secondary amplification possess the dUMP-containing sequence at their 5' termini. Treatment with uracil DNA glycosylase renders dUMP residues abasic, and unable to base-pair, resulting in 3' protruding termini which can then be ligated to complementary pAMP10 ~~vector~~ ends.

The amount of 2 µl of the secondary PCR reaction was mixed with 2 µl of pAMP10 cloning vector, 15 µl of 1X pAMP10 annealing buffer, and 1µl (1unit) of uracil DNA glycosylase. After mixing, the contents were incubated at 37 °C for 30 minutes. Five microlitres of the reaction was then transformed into 100 µl of competent *E. coli* XL-1 Blue cells (2.2.15.3). One-tenth of the transformation reaction was plated on LB-Ampicillin plates and incubated overnight at 37 °C.

#### **5.2.2 Exon Confirmation**

To confirm that exon products were trapped, the inserts were amplified, colony PCR using pAMP10 specific primers flanking the cloning site was first carried out to determine the product size. Clones were grouped according to the product size and a representative clone from each group was chosen for further analysis.



### **5.2.2.1 Colony PCR**

The recombinant colonies were randomly picked from overnight cultures with a sterile pipette tip, streaked on a master LB-agar plate for later utilized, and the tip placed in a PCR tube containing a drop of paraffin oil. The colony PCR procedure was followed as described in section 2.2.12.2, the only variation was the primers used in the PCR. The primers specific for pAMP10 used for the colony PCR were pUC-F and pUC-R. Clones with an insert size of 409 bp, as determined by colony PCR, were omitted from further analysis. The 409 bp product represented those being generated from vector-only clones containing no trapped product. Colony PCR was also performed using SA4 and SD2 primers, with the PCR product size of 153 bp was from the vector-only clones and such clones were omitted from further analysis. The sequences of all primers are included in *Table 5.1*.

### **5.2.2.2 DNA Isolation and Sequencing**

Clones selected for further analysis were grown in 20 ml of LB-broth plus ampicillin (100 µg/ml) and DNA was isolated using Qiagen Tip-20 columns (2.2.1.3). The trapped products were sequenced using Dye terminator Sequencing kits (2.2.16.1) with either SD2 or SD4 primers (*Table 5.1*). As the exons were trapped directionally in the trapped vector via the use of splicing consensus sequence, the DNA sequence obtained for each trapped product was then aligned in the sense orientation by its relation with the trapped vector sequence.

### **5.2.2.3 Physical Mapping of Trapped Products**

The DNA inserts of trapped products were removed from the vector with a *NotI/SaII*

double digestion. The digested products were separated on a 2.5% agarose gel electrophoresis and inserts were excised from the agarose gel, directly labelled as described (2.2.10.2) with  $\alpha^{32}\text{P}$ -dCTP, and used as probes to hybridise to Southern blots containing restriction digested BAC DNA from which the product was originally trapped.

#### **5.2.2.4 Trapped Exon Sequence Analysis**

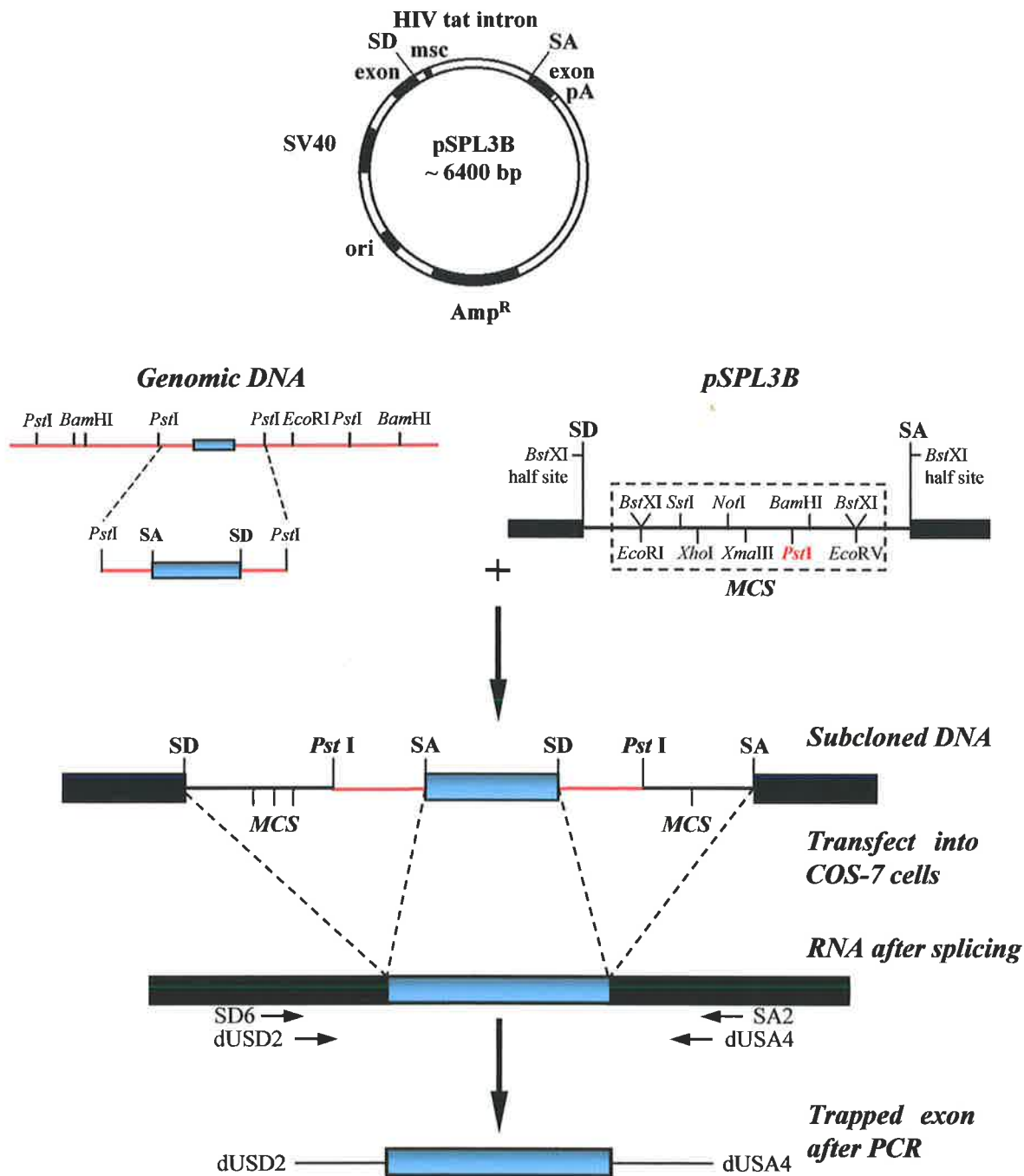
The BLASTN program (Altschne *et al.*, 1997) was used to identify the nucleotide sequence homology between the trapped products and sequences in the GenBank non-redundant and EST database (<http://www.ncbi.nlm.gov/index.html>). In the majority of cases a p value of less than  $10^{-5}$  ( $1.0e^{-5}$ ) was taken to indicate significant homology. The open reading frames within the sequences of trapped products was revealed using the Applied Biosystems SeqEd (version 1.0.3) software.

#### **5.2.3 Identification of Transcript Sequence**

Identification of a transcript sequence was based on trapped exon sequences and the use of bioinformatics for homology sequence searches. Subsequently, RT-PCR experiments and direct sequencing of homologous cDNA clones were utilized for sequence confirmation and assembly. The sequence information of oligonucleotide primers used for these RT-PCR experiments and cDNA sequencing are presented in **Table 5.3**. Finally, 5' and 3' RACE procedures were also used to complete the full-length transcript.

#### **5.2.4 In silico Protein Function Analysis**

The function of the characterized gene was analysed *in silico* utilizing the predicted protein sequence as template to compare, by BLASTP program, with either known functional or novel protein sequences in protein database. Analysis of the protein function



**Figure 5.1** : Schematic diagram illustrating the exon trapping procedure. The pSPL3B vector (Burn *et al.*, 1995) contains a multiple cloning site within an intron of the HIV-1 *tat* gene. This intron is flanked by exons of the rabbit  $\beta$ -globin gene which provide a splice donor (SD) and splice acceptor (SA) site. Genomic DNA (eg BAC) is digested at specific restriction sites (eg *Pst*I) such that a single exon may be contained within a digested fragment. The exon is flanked by SA and SD site. After subcloning BAC DNA into pSPL3B, recombinant plasmid vector is isolated and transfected into COS-7 cells. Cytoplasmic RNA is isolated for RT-PCR analysis. First strand cDNA is generated using a vector specific primer and followed by a first round PCR utilising primers SD6 and SA2. A second round PCR is then carried out using the nested primers dUSD2 and dUSA4. These primers allow uracil DNA glycosylase cloning of the PCR products.

**Table 5.1**  
**Sequences of Primers used for Exon trapping**

<b><i>Primer</i></b>	<b><i>Sequence (5'→3')</i></b>
SA5	cta gaa cta gtg gat ctc cag g
SD5	ccc tcg agg tcg acc cag c
SA2	atc tca gtg gta ttt gtg agc
SD6	tct gag tca cct gga caa cc
dUSA4	<u>cua cua cua cua</u> cac ctg agg agt gaa ttg gtc g
dUSD2	<u>cua cua cua cua</u> gtg aac tgc act gtg aca agc tgc
pUC-F	cac gac gtt gta aaa cga cgg cca gt
pUC-R	tgt gag cgg ata aca att tca cac agg a

was also carried out using the protein domain prediction program as described in previous chapters.

## **5.3 Results**

### **5.3.1 Exon Trapping**

Exon trapping was performed on a genomic clone BAC 561E17. This BAC extends approximately 140 kb from the proximal boundary of the cosmid contig of the 16q24.3 physical map constructed at this time. BAC DNA was double digested with *Bam*HI and *Bg*III or with *Pst*I alone and subcloned into the pSPL3B exon trapping vector (*Figure 5.1*). The subsequent protocol then involved transfection into COS-7 cells, RT-PCR amplification of the trapped fragments, which were then cloned into the pAMP10 vector. From this transformation, a total of 137 colonies were randomly selected from the primary plates. These were transferred to a master plate for further analysis. Colony PCR was carried out to identify the recombinants containing trapped sequences. Colonies that gave products corresponding with "vector-only" inserts (5.2.2.1) were omitted from further analysis. The remaining colonies were grouped according to the insert size and one to three colonies representing each group of the same insert size were then selected and grown for DNA isolation and sequencing. A total of 36 colonies were finally selected and subjected to DNA isolation and sequencing.

### **5.3.2 Trapped Exon Sequence Analysis**

Sequence analysis of the selected clones for the presence of open reading frame and sequence homology to known genes or partially sequenced ESTs in the public sequence database are summarised in *Table 5.2*. Initial analysis indicated that the trapped exons could potentially code for 5 different transcripts.

***Table 5.2***  
**Trapped Exons from BAC 561E17**

Grouped as a Transcript	Clones Selected and Sequenced	Size (bp); ORF; Frame	BLAST search results
A	561-9 561-38 561-71 561-89	159; +; 1, 3	Cadherin 15 ( <i>CDH15</i> ) cDNA, nt 120-278
	561-88	236; +/-	The sequence from nt 71-236 is identical to nt 120-278 of <i>CDH15</i> cDNA
	561-3 561-10 561-18 561-22 561-101 561-115	156; +; 1, 2	<i>CDH15</i> cDNA, nt 279-434
	561-17	289; +; 1	The sequence between nt 1-156 is identical to <i>CDH15</i> cDNA nt 279-434
B	561-4 561-45 561-58 561-69 561-95 561-107 561-111	93; +; 1	AA543575 (mouse mammary gland), AA733571 (mouse skin), AA794849 (mouse 2-cell embryo)

**Table 5.2 (continue)**

<b>Grouped as a Transcript</b>	<b>Clones Selected and Sequenced</b>	<b>Size (bp); ORF; Frame</b>	<b>BLAST search results</b>
<b>B</b>	<b>561-13</b>	<b>169; No</b>	<b>The reverse sequence between nt 132-169 shows homology to mouse ESTs: AA733571 (mouse skin), AA794849 (mouse 2-cell embryo) AA065769 (mouse melanoma)</b>
<b>C</b>	<b>561-72</b>	<b>148; +; 1, 3</b>	<b>R41095 (human heart), AA895310 (mouse lung), AA553169 (mouse mammary gland) AA153265 (mouse melanoma)</b>
	<b>561-114</b>	<b>268; +; 1</b>	<b>AA312702 (Jurkat T cell) AA345035 (human gall bladder) AI029504 (rat cDNA)</b>
<b>D</b>	<b>561-36 561-109</b>	<b>84; +; 2, 3</b>	<b>No EST homology but the deduced amino acid sequence shows homology to mouse zinc finger proteins: Zfp70p (44% identity), Zfp61p (62% identity), and Zfp67p (40% identity)</b>
<b>E</b>	<b>561-117</b>	<b>108; +; 1</b>	<b>AA285067 (human B cell)</b>

Trapped exons **561-9**, **561-38**, **561-71**, and **561-89** were from the same exon and had two possible ORFs either in frame 1 or frame 3. Homology sequence analysis showed that they were trapped from the gene sequence corresponding with nucleotides 120-278 of Cadherin 15 cDNA (GenBank accession no. NM-004933). Human cadherin 15 (*CDH15*: M-cadherin) has previously been mapped to this region (Kaupmann *et al.*, 1992; Kremmidiotis *et al.*, 1998). The trapped product **561-88**, though of a larger size (236 bp) when compared with that of **561-9** (159 bp), was also trapped from the same part of *CDH15* (cDNA nt 120-278). Present, however, was an extra 70 bp of 5' sequence. Analysis of the sequence at this junction of *CDH15* exon and the extra 70 bp indicated the presence of a nucleotide sequence resembling a splice acceptor consensus. This indicates that part of the adjacent "intron" was likely to be included in this trapped product (*Figure 5.2*).

The region between nucleotides 279 and 434 of the *CDH15* cDNA were also repeatedly trapped as shown by sequencing of the clones **561-3**, **561-10**, **561-18**, **561-22**, **561-101**, and **561-115**. The trapped product **561-17** was also trapped from this part of the *CDH15* sequence but also included a 133 bp sequence likely to be part of the adjacent 3' intron (*Figure 5.2*).

A group of 7 trapped exons represented by clone **561-4** were homologous and were likely trapped from an exon of an uncharacterized human gene. There was no homology to human ESTs though homology was found to three overlapping mouse EST sequences. These were vj82g04.r1, vu73b05.r1, and vr48d04.s1 (GenBank accession no. AA543575, AA733571, and AA794849 respectively). These three overlapping ESTs form a cDNA sequence of approximately 1.1 kb. The sequence homology between **561-4** and mouse cDNA was high, with 89% identity.



### CDH15 cDNA (5' sequence)

5' ACTTGCCTGTCACTCAGCCTGGACGCGCTTCTTCGGGTTCGCGGGTGCACCTCCGGCCCGGCTCCCGCC  
TCGGCCCCGATGACGCGCGGTTCTCTCTCGTCTCGGGCTGTTGGCCAGAGCCTCTGCCTGTCTTTGG  
GGGTTCTGGATGGAGGAGGCCACCACCCTGTACCCCTGGCGCCGGGCGCCTGCCCTGAGCCGCGTGCG  
561-9 nt 119/120 nt 278/279  
GAGGGCCTGGGTATCCCCCGATCAGCGTATCCGAGAACCACAAGCGTCTCCCTACCCCTGGTTTCAG  
ATCAAGTCGGACAAGCAGCAGCTGGGCAGCGTCATCTACAGCATCCAGGGACCCGGCGTGGATGAGGAGC  
561-3 nt 434/435  
CCCGGGGCGTCTTCTCTATCGACAAGTTCACAGGAAGGTCTTCTCAATGCCATGCTGGACCGCGAGAA  
GACTGATCGCTTCAGGCTAAGAGCGTTTGCCTGGACCTGGGAGGATCCACCCTGGAGGACCCACGGAC  
CTGGAGATTGTAGTTGTGGATCAGAATGACAACCGCCAGCCTTCTGCAGGAGGCGTTCACTGGCCGCG  
TGCTGGAGGGTGCAGTCCCAGGCACCTATGTGACCAGGGCAGAGGCCACAGATGCCGACGACCCCGAGAC  
GGACAACGCAGCGCTGCGGTTCTCCATCCTGCAGCAGGGCAGCCCCGAGCTCTTACAGCATCGACGAGCTC  
ACAGGAGAGATCCGCACAGTGCAAGTGGGGCTGGACCGGAGGTGGTTCGCGGTGTACAATCTGACCCTGC  
AGGTGGCGGACATGTCTGGAGACGGCCTCACA -----3'

### 561-88

5' -

ctcccagctgcaagctgccgcccagggctccaggagacgggtactgtgggacgcatctgtctttgttgag  
AGCCTCTGCCTGTCTTTGGGGTTCTGGATGGAGGAGGCCACCACCCTGTACCCCTGGCGCCGGGGCGC  
CTGCCCTGAGCCGCTGCGGAGGGCCTGGGTATCCCCCGATCAGCGTATCCGAGAACCACAAGCGTCT  
CCCTACCCCTGGTTCAG-3'

### 561-17

5' -

ATCAAGTCGGACAAGCAGCAGCTGGGCAGCGTCATCTACAGCATCCAGGGACCCGGCGTGGATGAGGAGC  
CCCGGGGCGTCTTCTCTATCGACAAGTTCACAGGAAGGTCTTCTCAATGCCATGCTGGACCGCGAGAA  
GACTGATCGCTTCAGgtgaggagctggtggtgacactgtgcccctcaagcaggcctggtggaaagcca  
gtgcccctcctcccAccaggettctctcccgcctcggtgggcttcaggccagactgcaagatccaggc  
accottaa-3'

**Figure 5.2:** The 5' sequence of *CDH15* cDNA and the corresponding trapped exon isolated from BAC 561E17. The ATG start codon is indicated in *blue* and *underlined*. Clone 561-9 was trapped from exon 2 of *CDH15* gene corresponding with the nucleotide positions 120-278 of its transcript. The trapped exon 561-88 was also from *CDH15* exon 2 but included part of its 5' up-stream intron sequence (lower case). Trapped product 561-3 was from exon 3 of *CDH15*, corresponding with base positions 279- 434 of transcript. Trapped exon 561-17 was also from exon 3 but included part of its down-stream intron sequence. The consensus sequence of splice acceptor (*ag*) and donor (*gt*) are indicated.

A trapped product **561-13**, 169 bp in size, had a 94% sequence homology to two of the three previously identified overlapping mouse EST, vr48d04.s1 (AA794849) and vu73b05.r1 (AA733571), and an additional mouse EST, mm44g06.r1 (AA065769). This homology, however, was by the reverse complementary sequence and only between nucleotides 132 to 169 of 561-13. Analysis of the reverse sequence of 561-13 at the boundary between homologous and non-homologous sequence revealed a sequence that resembled a consensus donor-splicing site (*Figure 5.4A*). This product might be trapped from part of an exon and its 3' intron, using potential splice sites in the reverse orientation.

By their sequence homology to the same set of overlapping ESTs both trapped exons **561-4** and **561-13** could therefore belong to the same gene

Two trapped products **561-72** (148 bp) and **561-114** (268 bp) appeared to be part of the same gene. Sequence homology searches using 561-72 as a query initially identified a human EST sequence R41095 from cDNA clone K146-f, and three mouse ESTs, AA895310, AA153265, and AA553169. These EST sequences were then used to search for more extended sequences in the public sequence database. A number of overlapping cDNA sequences were finally combined to form a sequence of approximately 1.4 kb. On the other hand, clone 561-114 showed homology to at least eight EST sequences from human, mouse and rat, from which a 0.92 kb cDNA sequence was assembled. One of these ESTs, AA312702 (EST 183373), from a human Jurkat T cell cDNA library was found to link these overlapping ESTs from 561-114 to those detected by 561-72 (*Figure 5.5*). Together, these formed a cDNA sequence of approximately 1.95 kb.

Trapped exons **561-36** and 561-109 were identical and therefore from the same exon. BLAST analysis of the nucleotide sequence from the EST sequence database (dbEST)

initially detected no sequence homology (as of October 2000). However, by using all six frames of translation of the nucleotide sequence as queries using BLASTx program, significant amino acid sequence homology was detected. The deduced amino acid sequence of 27 residues from frame 3 of the 561-36 ORF showed sequence homology to a KRAB-B motif of mouse gene products Zfp61p, Zfg67p, and Zfg70p. These are members of the Krüppel-associated box (KRAB)-containing group of zinc finger proteins (KRAB-ZFP) (Bellefroid *et al.*, 1995 ).

The trapped product **561-117**, of 108 bp contained an ORF at frame 1 and showed sequence homology to the 3' part of EST AA285067.

At this stage 5 potential transcripts represented by trapped exons were assigned (**Table 5.2**) according to sequence homology to the expressed sequence database at NCBI as **transcript A** (*CDH15*: 561-9, 561-88, 561-3, and 561-17), **transcript B** (561-4 and 561-13), **transcript C** (561-72 and 561-114), **transcript D** (561-36), and **transcript E** (561-117).

### 5.3.3 Physical Mapping of Trapped products

Representative trapped exons from each putative transcript were successfully mapped by Southern hybridization to the original genomic clone, BAC 561E17, and to other overlapping BACs and cosmid clones that formed a regional physical map (**Figure 5.3**). Also mapped were other trapped products, which did not show homology, at this time, to any expressed sequence in database at NCBI. These were 561-33, 561-55, 561-59, 561-7, and 561-40.

Six clones were clustered within a region of approximately 30 kb at the telomeric end of the BAC 561E17. By order in the BAC these were 561-33, transcript C (561-114 and

---

***Figure 5.3*** : Physical and Transcription Map of 16q24.3 LOH sub-region. The diagram shows a physical map of the genomic region extending from the previous published physical and transcript map at 16q24.3 LOH region (Whitemore *et al.*, 1998a). The restriction map of two restriction enzymes is indicated: **E**, *EcoRI*, and **Ea**, *EagI*. The size of restriction fragments is indicated in kilobase pair (kb). The trapped exons that were mapped to this physical region are denoted by *filled circles* with vertical dash lines indicate their map locations. The trapped products beginning with "561-" were trapped from BAC 561E17 whereas those beginning with "ET-" were trapped from cosmids in the contig. Also indicated, the assigned transcription units T13, T14, T14A, T15, T16, T17, and T18, based on the trapped exons and/or ESTs mapped to this region.

T18

T17 T16

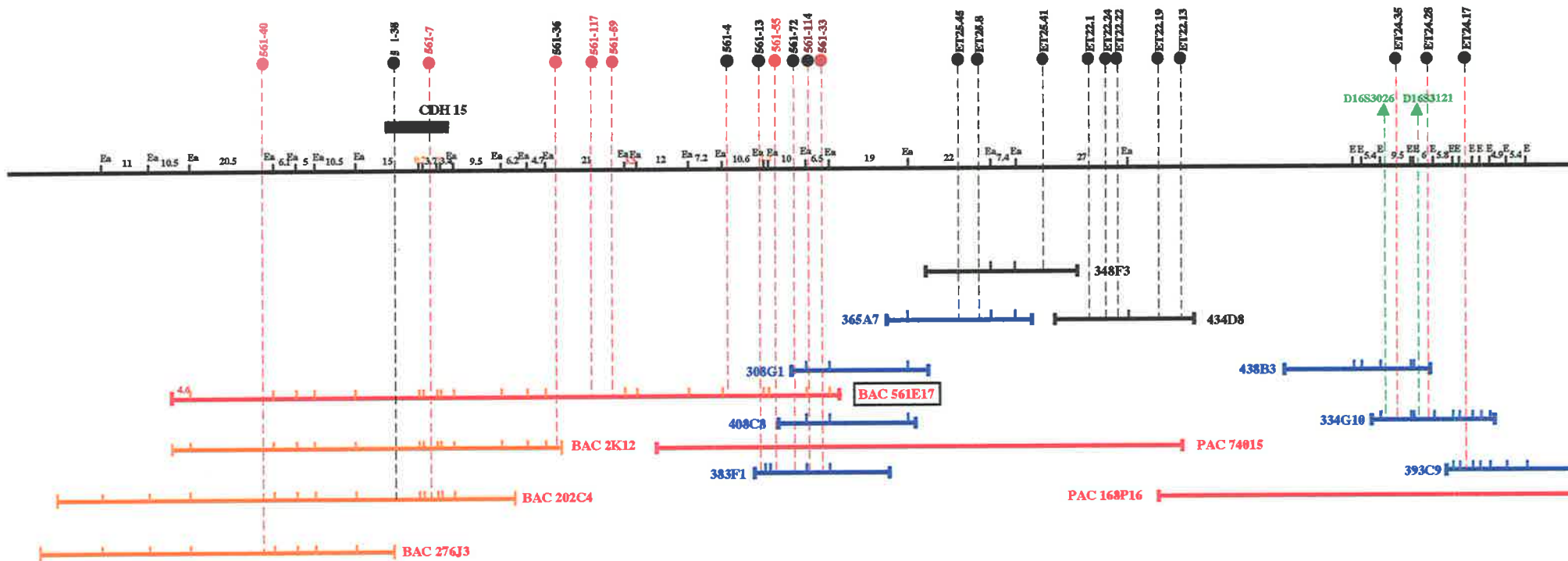
T15 T14A T14

T13

SGC33145 WI-13072

C113

AA027336



561-72), 561-55, and transcript B (561-13 and 561-4). At this region of the physical map, a cDNA 118499 (initially designed STS SGC33145) belonging to a group of cDNA clones collectively referred to a Unigene Cluster Hs.10238 was also mapped initially by PCR (Whitmore, S.A. and Powell, J., personal communication), and was assigned as transcription unit 17 (*T17*). It was then confirmed by Southern hybridization at this time to this region of the BAC 561E17.

Three trapped products, 561-59, 561-117 (transcript E), and 561-36 (transcript D) were localized near the middle region of the BAC 561E17.

The trapped clones 561-38 (*CDH15* nt 120-178) was co-localized with 561-7. In the present study, PCR products amplified from 5' and 3' region of *CDH15* cDNA were also used as probe for Southern hybridization and mapped specifically to this region of BAC 561E17.

Trapped exon 561-40 was mapped close to the proximal end of the BAC and distal to the region that cDNA clone IMAGE:1098126 (*T18*) was initially mapped by Southern hybridization (Whitmore, S.A., personal communication).

### **5.3.4 Verification of the Transcript Sequence**

Among the trapped products isolated from the BAC 561E17, there were two group of trapped exons initially assigned as *transcript B* (561-4 and 561-13) and *transcript C* (561-72 and 561-114) and the trapped products 561-33 and 561-55 clustered at the distal region of this BAC (**Figure 5.3**). Due to their physical proximity it was considered likely that they were part of the same transcript and so were chosen for further study.

#### **5.3.4.1 Transcript Sequence generated from 561-4 and 561-13**

Two trapped exons, 561-4 and 561-13 were homologous to the mouse cDNA

sequence formed by at least three overlapping mouse ESTs, AA543575, AA733571, and AA794849. This group was initially assigned as *transcript B* (section 5.3.2). The order of these two trapped exons was arranged with respect to the mouse EST sequences, and showed that 561-13<sup>was/is</sup> situated approximately 350 bp 5' upstream from the exon 561-4 (Figure 5.4 B). RT-PCR using a forward primer within the 561-13 sequence and a reverse primer from the 561-4 sequence (Table 5.3) successfully amplified a product of 357 bp from human fetal heart total RNA. The total human cDNA sequence of 378 bp between 561-13 and 561-4 was assembled and was homologous to the overlapping mouse cDNA sequences (Figure 5.4 B).

To identify additional human cDNA sequence extending the 378-bp product, an assembled mouse cDNA sequence of 1,022 bp generated from the three overlapping mouse ESTs, mentioned above, was used to screen for homologous human DNA sequences from both EST and non-redundant databases. The search results identified a genomic sequence which was part of the PAC 203P18 (GenBank accession no. Z97180). This PAC was derived from the human X-chromosome region Xq27.1-q27.3. Homology between the PAC 203P18 sequence and that of the mouse cDNA was 88%. This homology was only from the first 835 bp of the assembled 1,022-bp mouse cDNA sequence. The remaining 187 bp showed no significant homology (Figure 5.4 B). The 378-bp human cDNA between exons 561-13 and 561-4 was 100% homologous to this corresponding sequence of the X chromosome.

The finding that part of genomic DNA from the X chromosome showed sequence homology to the cDNA sequence of at least 835 bp suggested that this part of the X chromosome sequence is a processed pseudogene. Further analysis of the PAC 203P18 DNA sequence, which flanked this region of homology, showed that there were adjacent L1 elements. A 1.39-kb L1 element was situated 338 bp 5' to this 835 bp homologous

region and a 0.85-kb L1 element was situated 854 bp 3' to this region. Flanking by the L1 elements at both ends suggested the likelihood that this DNA segment had been transposed as a mobile element. A processed pseudogene generally is a DNA fragment being copied from a processed RNA transcript and transposed to the genomic region by the transposable elements such as L1 repeat (Esnault *et al.*, 2000). The above evidence again supported that this genomic sequence of the X chromosome, flanking by L1 elements, was a processed pseudogene of the transcript B. The processed pseudogene did not contain a promoter and therefore would not be expressed.

The additional sequence of the pseudogene which flanked the 835-bp homologous sequence was then utilized to detect additional sequence homologies from <sup>the</sup> database. Homology searches had identified at least one human- and three mouse-ESTs, accession no. AI203683, AI461784, AI197222, and AI121535 respectively. These aligned with the sequence 3' to the region homologous to 561-13/561-4 product (**Figure 5.4B**). To further identify if these EST sequences belonged to, and formed part of, the same transcript with 561-13/561-4, the human cDNA clone qf54a03 (EST AI203683) was sequenced. The result showed that the sequence of this clone overlapped with the sequence generated from the trapped exons 561-4 and 561-13. The combined cDNA sequence of 1,263 bp was again identical to the X chromosome processed pseudogene.

Further analysis of pseudogene sequence was undertaken to provide additional information related to its original transcript. The total 2,026-bp sequence of this pseudogene revealed an ORF of 1167 bp which extended from the first nucleotide (**Figure 5.4B**). There was no start codon. A stop codon, TGA, was identified between the nucleotide positions 1168 and 1170 which was then followed by the non-coding sequence likely to be a 3' UTR. The stop codon was also detected in the corresponding mouse cDNA at the homologous and non-homologous junction (**Figure 5.4B**). The lack of significant

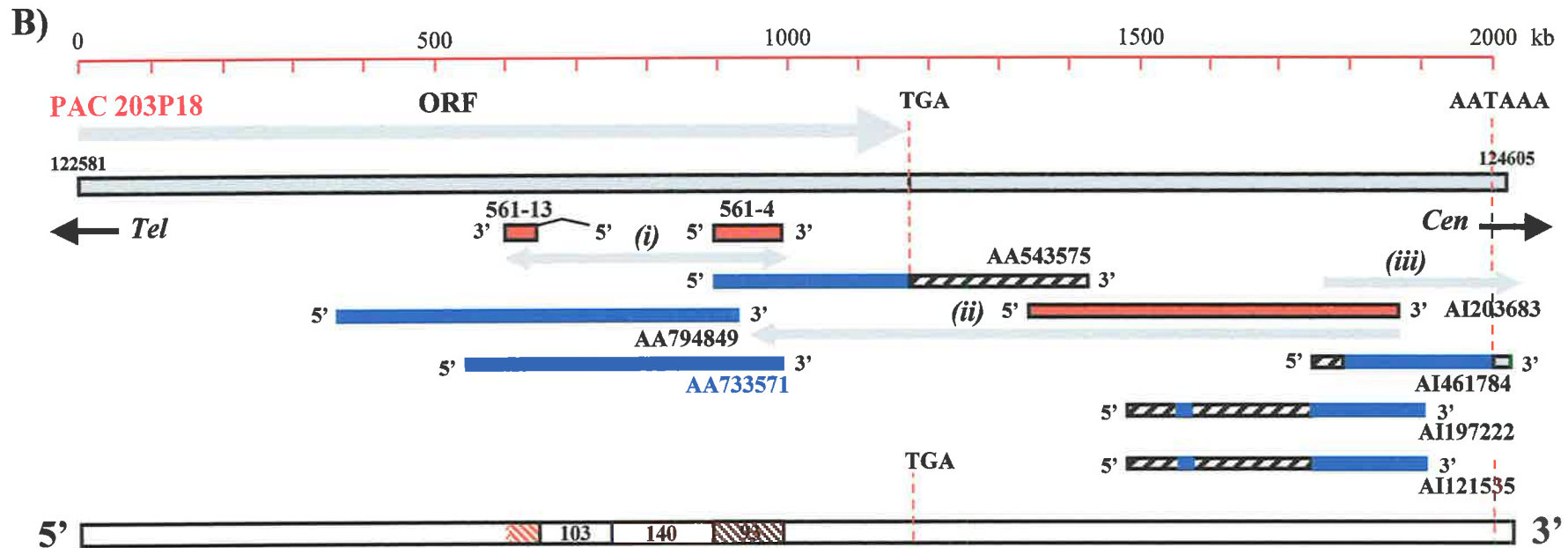


---

**Figure 5.4** : Alignment of the trapped exons 561-13 and 561-4 with the overlapping ESTs, and the pseudogene on PAC 203P18 sequence (nucleotide position 122581- 124605) from X chromosome region Xq27.1-27.3. **A)** Nucleotide sequence (5' to 3') of the trapped product 561-13 (sequence # 1) which was trapped from part of an exon and its flanking intron (sequence # 2, reverse complement sequence of 561-13) in the reverse orientation. **B)** Scheme representing the alignment of the trapped products, ESTs, and genomic sequence from X chromosome, 5' to 3' from telomeric (*Tel*) to centromeric (*Cen*) direction. The sequence of 561-13 that is homologous to the ESTs is only the 42 bp exon sequence (*uppercase* in #2). The tissue sources of ESTs from human (red) and mouse (blue) are AI203683, human testis; AA543575, AI461784, AI197222, and AI121535, mouse mammary gland; AA794849, mouse 2-cell embryo; and AA733571, mouse skin. The human cDNA sequence were also isolated by RT-PCR (product *i*), sequencing of the cDNA clone of AI203683 (product *ii*), and 3' RACE (product *iii*). *Hatched boxes* indicate part of the mouse sequences that are not homologous to the human sequence. The open reading frame (ORF) was derived from the sequence of processed pseudogene on PAC 203P18. The assembled sequence, *bottom bar*, was later found to be the 3' part of transcription unit 13 (*T13*).

**A)** *561-13*

1. gtggcagaag gtcgagagag gtcaagggcc atgagtggga caagacgggc tggagaaagc cgtgactagg ggccccagac gcatcccaga  
gagagaaggc agtggctctc ccgggccccg cactcacCAC GGGGATGTGG AGCTTGCTGA GCGGCTTGCC GTCCAGGAG
2. CTCCTGGACG GCAAGCCGCT CAGCAAGCTC CACATCCCCG TGgtgagtgc gggggccggg agagccactg ctttctctct ctgggatgcg  
tctggggccc ctagtcaagg ctttctccag ccgctcttgt ccactcatg gcccttgacc tctctcgacc ttctgccac



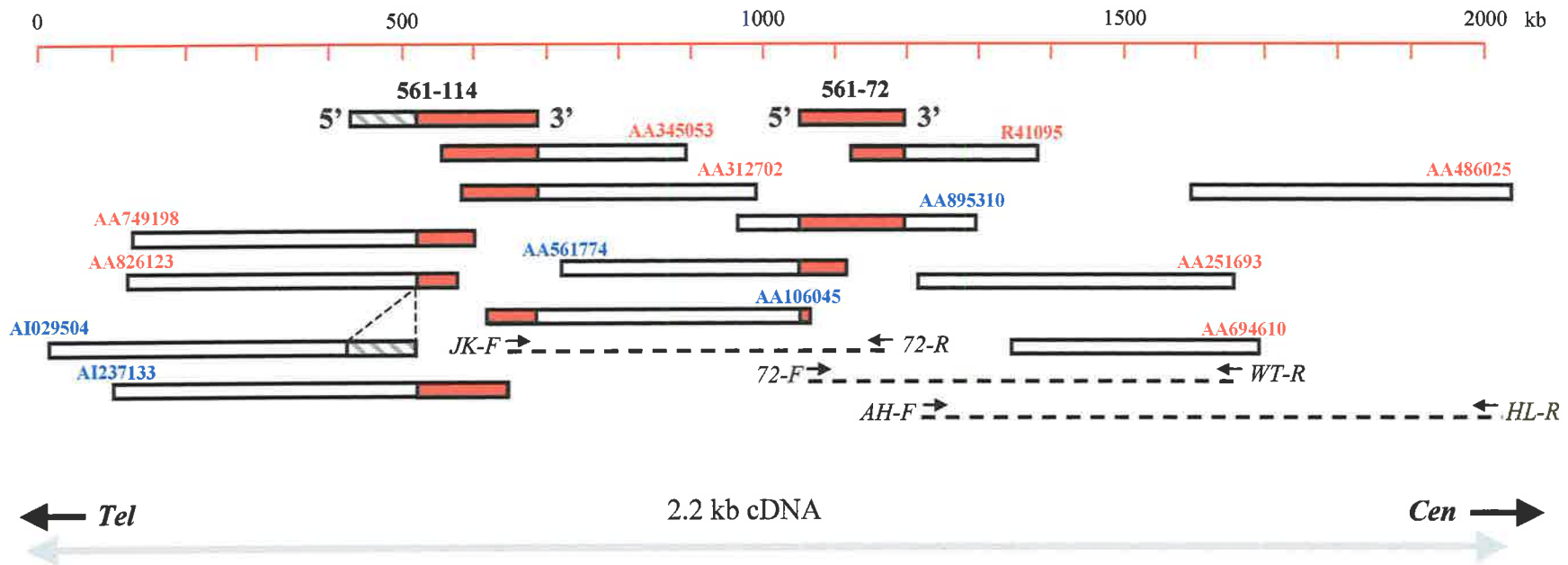
homology of the mouse sequence to that of the human 3' down-stream from the stop codon was also observed and supported the likelihood that this sequence was the 3' UTR. Furthermore, the nucleotide sequence characteristic of the polyadenylation signal, AATAAA, was also present at the 3' end of this pseudogene. With respect to its process pseudogene, the identifying transcript sequence (initially given as transcript B) was likely to be the 3' part of the mRNA transcript of a gene.

To confirm the 3' end of the transcript, 3' RACE using a set of primers (*Table 5.3*) designed from the sense strand of the cDNA qf54a03 was performed (the procedure described in section 2.2.17). The result revealed the same sequence as that present in the pseudogene but with an extra seven nucleotides prior to the polyadenylation site. These extra nucleotides were later confirmed by the available genomic sequence of PAC clone 74015 that overlapped with BAC 561E17 (the sequence data has been deposited with GenBank as part of NT-010542.13).

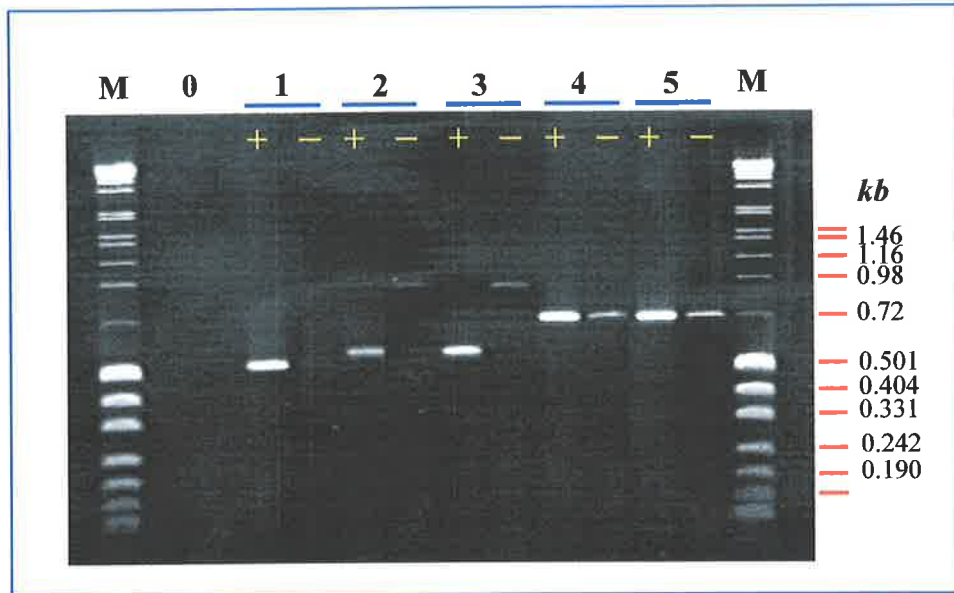
#### **5.3.4.2 Transcript Sequence of 561-72 and 561-114**

Trapped exons 561-114 and 561-72 were shown to be linked to each other by a set of overlapping human and mouse EST sequences as mentioned in section 5.3.2 and shown in *Figure 5.5*. To confirm this, primers were designed for RT-PCR analysis of these overlapping ESTs and trapped products (*Table 5.3*). RT-PCR products were successfully obtained from total RNA isolated from human fetal heart and fetal kidney (*Figure 5.6*). These products were purified and sequenced. Three human cDNA clones, ab40b03 (AA486025), ah20h11 (AA694610), and od53h03 (AA826123), that formed part of this overlapping sequence, were also chosen for sequencing. The sequences obtained from the RT-PCR and cDNA clones were then assembled to form a 2.2-kb cDNA sequence.

Analysis of clone 561-114, 268 bp in size, showed that it was likely to have been



**Figure 5.5** : Alignment of the trapped products 561-114 and 561-72 with a set of overlapping ESTs. Homology between trapped exon sequences and ESTs is denoted by the corresponding *filled* or *hatched* bars. The overlapping ESTs were from various tissue sources: R41095, human heart; AA251693, human tonsilar B-cells; AA694610, Wilms tumor; AA486025, Hela cell; AA345053, human gal bladder; AA312702, Jurkat T cells; AA749198, human cDNA; AA826123, human cDNA; AA895310, mouse lung; 561774, mouse blastocyst; AA106045, mouse testis; AI029504, rat cDNA; and AI237133, rat ovary. The *forward* and *reverse* arrows with primer identities indicate the primers utilized for RT-PCR amplification of the transcribed sequence fragments, *dash lines* (products are presented in **Figure 5.6**). The hatched bar at the 5' half of 561-114 represents the sequence from nt 1-97. The cDNA sequence of 2.2 kb was assembled from the RT-PCR product and clones of ESTs



**Figure 5.6** : RT-PCR analysis of the trapped exon 561-72 and overlapping EST sequences. RT-PCR was performed on total RNA isolated from fetal heart (*lanes 1, 2, 4,*) and fetal kidney (*lanes 3, 5*). The primer pairs used for cDNA amplification are *JK-F/72-R* (*lanes 0, 1*), *72-F/WT-R* (*lanes 2, 3*), and *AH-F/HL-R* (*lanes 4, 5*). All primer sequences are presented in *Table 5.2a*. The *plus (+)* and *minus (-)* indicate the RT-PCR reaction *with* and *without* reverse transcriptase respectively (section 2.2.14). *Lane M*, DNA size marker in kb, *lane 0*, control reaction without RNA.

***Table 5.2a***

**Sequences of the primers used for RT-PCR amplification and Sequencing of the initial transcript sequence.**

<b><i>Primer</i></b>	<b><i>Sequence (5' → 3')</i></b>	<b><i>Primer location in T13 and PCR product size</i></b>	
<i>JK-F:</i>	cccttctcatgcagatgacg	782-801	528 bp
<i>72-R:</i>	cctcgtggaagtgaagtg	1309-1290	
<i>72-F:</i>	gaagctgctgctgcggtac	1179-1197	573 bp
<i>WT-R:</i>	lctcgtgtcttactaccagg	1751-1732	
<i>AH-F:</i>	aagaggaagacgcacatcc	1331-1350	745 bp
<i>HL-R:</i>	tctgtccgccagtgcttgg	2075-2057	

trapped simultaneously from two adjacent exons. From the results of database searches, the sequence of 561-114 between nucleotide positions 1 and 97 was only found in a limited number of cDNA clones. The majority of the identified cDNA clones showed sequence homology to the region of 561-114 from nucleotide positions 98 to 268 (*Figure 5.5*). It was suggested that the exon corresponding to nucleotides 1 to 97 of 561-114 could have been alternatively included in some transcription products of some tissues, at least those included in the EST sequence database.

The transcripts B and C were later found to be part of the gene assigned as transcription unit 13 (*T13*) described in the next section.

### **5.3.5 Assembly of the Transcription Unit 13 (*T13*) Sequence.**

Transcription unit 13 was originally assigned to the physical map of 16q24.3 LOH region from three trapped exons, ET24.35, ET24.28, and ET24.17 which clustered within two overlapping cosmids, 438B3 and 334G10 (*Figure 5.3*). Ms Karen Lower performed these studies, first by RT-PCR experiments. These, however, failed to provide any evidence that these exons were part of the same gene. Homology searches of the cDNA sequence database at NCBI were then performed using the sequence of these exons. Only ET24.35 identified an overlapping sequence of ESTs. Further homology searches, cDNA sequencing, and RT-PCR experiments allowed the determination of a cDNA sequence of 2,290 bp. To identify the transcription product size and expression pattern, a multiple tissue Northern filter was hybridized with a cDNA insert from the clone zs04a09 which was part of the assembled 2,290-bp cDNA. The result showed a common 9.5 kb mRNA band with varying degree of expression between the different tissues (Lower, K.M., personal communication). Bands of 3.2, 3.0, and 1.0 kb with varying intensity in different

tissue were also detected in this Northern.

During the identification of the combined cDNA sequence of 2.2 kb generated from the trapped exons 561-114 and 561-72 (transcript C, described in 5.3.4.2) it appeared that this sequence was identical to the cDNA sequence of *T13*, described above. From the 16q24.3 physical map, these two sequences were separated by approximately 120 kb (**Figure 5.3**). It is likely that these two cDNA fragments are derived either from the same gene or from duplicated genes. This was resolved by Southern hybridization to the cloned DNA of the physical map using RT-PCR products from the *T13* 2,290-bp cDNA. The results showed that most of this sequence mapped to BAC 561E17 and three cosmids, 308G1, 408C8, and 383F1, which overlapped with the distal end of this BAC (**Figure 5.3**). However, the 5' end of the transcript including ET24.35, was located in cosmids 438B3 and 334G10 which situated approximately 120 kb from BAC 561E17. This result indicated that the *T13* transcript encompassed more than 120 kb of genomic DNA and confirmed that the two independently isolated sequences were in fact the same transcript.

Northern analysis was also performed using a probe generated from RT-PCR product of 561-72 and the overlapping ESTs (**Figure 5.5**). This also identified a 9.5 kb band (**Figure 5.7**). Subsequent cDNA database analysis and cDNA clone sequencing extended the *T13* cDNA in a 3' direction resulting in a *T13* cDNA of 4,570 bp. Further extension of the 3'cDNA sequence based on homology searches to partially sequenced cDNAs in database (dbEST) was not possible as no additional cDNAs were found which extended the sequence. Transcript B was also considered to be part of the same gene with transcript C, and therefore with *T13*, however, several attempts by RT-PCR failed to link transcript B with the partial sequence of 4.5 kb of *T13*.

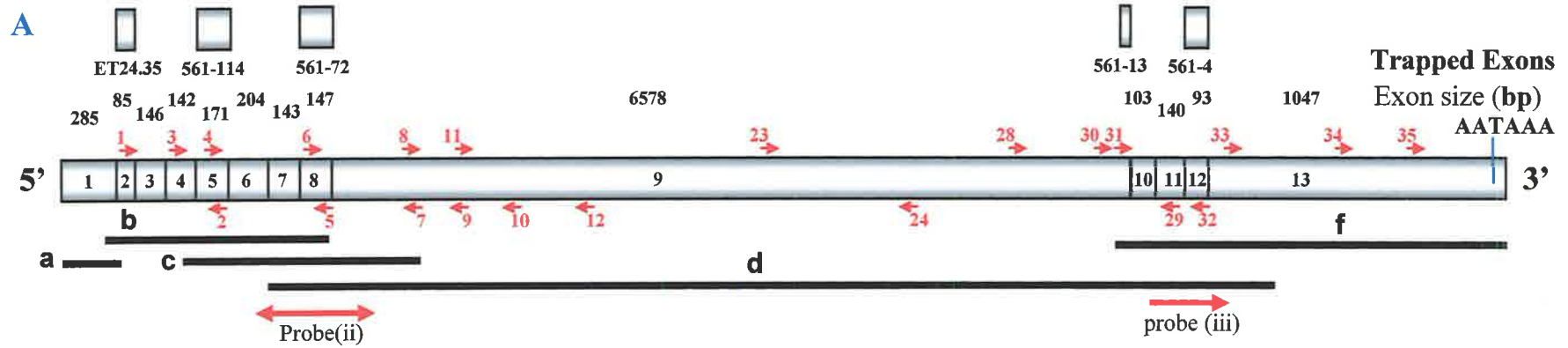
During the time that this work was in progress, a large scale sequencing project of



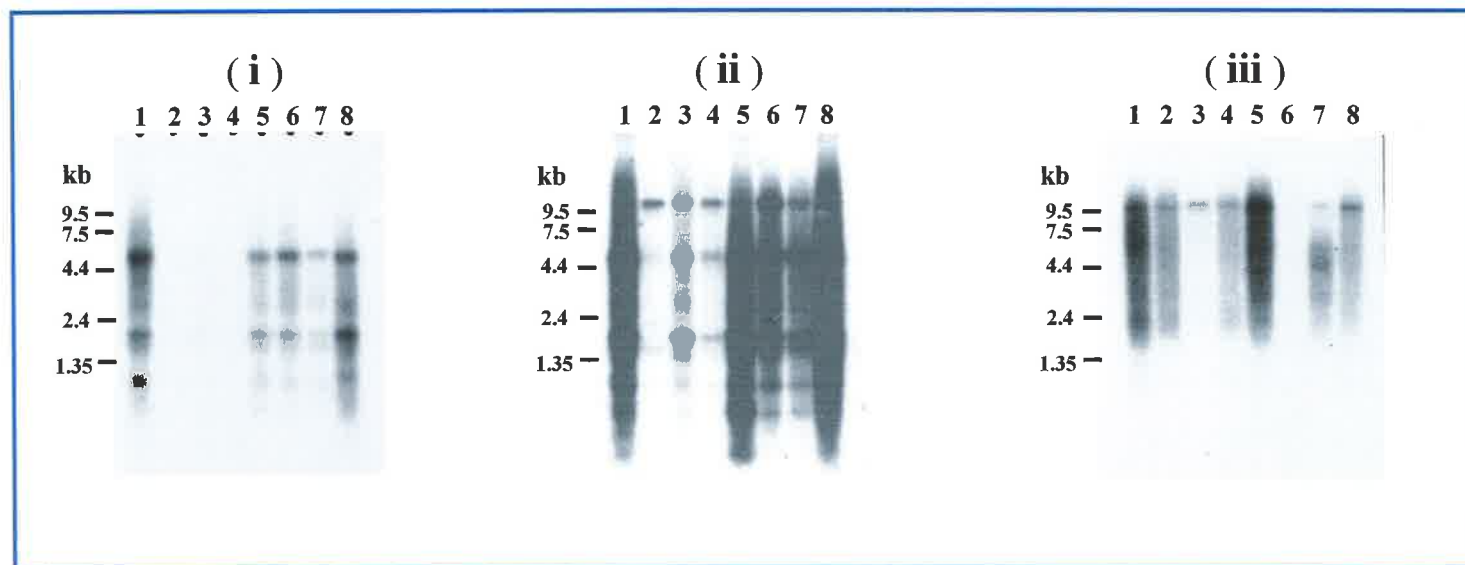
chromosome 16<sup>20-23</sup> established and generated parts of the genomic sequence within the 16q24.3 region (Kremmidiotis, G. and Gardner, A., personal communication, GenBank accession no. NT-010542.13). One of the genomic sequences was a fragment encompassing the distal region of 561E17. Alignment of the *T13* cDNA with this genomic sequence revealed the partial genomic structure of the *T13* gene. This gene featured a large intron of approximately 90 kb between exons 2 and 3. The *T13* exon 2 had been trapped as ET24.35 whereas 561-114 and 561-72 were trapped from exon 5 and exon 8 respectively. With the availability of genomic sequence a gene prediction program, GeneFINDER, was utilized to identify the possible cDNA sequence in this region. This program successfully predicted and extended the cDNA sequence from the 3' end of the 4,570-bp cDNA. Interestingly, the extended cDNA sequence <sup>was co-linear with</sup> co-linear with the genomic sequence and joined with the transcript B which was identified from the two trapped exons 561-13 and 561-4 situated approximately 10 kb distal from 561-72. Transcript B was identified as the 3' part of cDNA (5.3.4.2) and therefore completed the 3' end of *T13*. The 5' RACE was also utilized to identify the 5' end of *T13*. This, however, failed to obtain the extended sequence. The 5' sequence of *T13* was, later, extended by the overlapping sequence of human EST bb15c08.y1 (AW732837). Finally the 9,282-bp full-length transcript sequence of *T13* was assembled (**Figure 5.7**), corresponding approximately with its 9.5 kb mRNA band identified by Northern analysis. The *T13* cDNA predicted from genomic sequence by GeneFINDER, with no intron interruption, was also confirmed by RT-PCR analysis using a number of primer pairs designed from this cDNA sequence (**Table 5.3**) and was carried out on mammary gland poly A<sup>+</sup> RNA to avoid genomic contamination. In addition, RT-PCR products from this and other part of *T13* transcript sequence all detected the same 9.5 kb mRNA by Northern hybridisation (**Figure 5.7**).

---

**Figure 5.7:** *T13* sequence assembly and Northern blot analysis. **A)** Diagrammatic representation of the *T13* sequence assembled from ESTs (**a**), RT-PCR products (**b**, **c**, **f**), and the result of gene prediction from genomic sequence (**d**). Product “**b**” represents an assembled cDNA of the originally described *T13*. Products “**c**” and “**f**” are the results of assembled sequences described in Fig 5.5 and Fig 5.4 respectively. Also indicated are the approximate location of the primers, *red arrows* with primer numbers, used in RT-PCR analysis of the transcript (*Table 5.3*), the exon boundaries, and the size of each exon in **bp**. The corresponding trapped exons are indicated above the transcript. **B)** Northern blot hybridization results utilizing three different cDNA probes: (i) *T17* probe, (ii) *T13* probe derived from cDNA sequence between nt 782-2075, (iii) mouse Northern membrane probed with cDNA of aa794849, homologous to the 3’ part of *T13* (see *Figure 5.4*). The 9.5 kb band is seen only when using *T13* probes. The tissue sources of human mRNA are: **1** heart, **2** brain, **3** placenta, **4** lung, **5** liver, **6** skeletal muscle, **7** kidney, **8** pancreas. The mouse tissue sources of mRNA are: **1** heart, **2** brain, **3** spleen, **4** lung, **5** liver, **6** skeletal muscle, **7** kidney, **8** testis.



**B**



**Table 5.3****Primer pairs for RT-PCR analysis of *T13*: sequence, location, and PCR product size**

<b>No.</b>	<b>Primers</b>	<b>Primer sequences (5' → 3')</b>	<b>Exon and nucleotide location in cDNA</b>	<b>Approximate size of PCR product</b>	
1	24.35F	tatatacagccctgctctgg	Exon 2	288-307	0.540 kb
2	T13-8	gggctgttggcagactcc	Exon 5	824-807	
3	251293T3	gcggaagctgcccttcacc	Exon 4	600-618	0.713 kb
5	561-72R	cctcgctggaagtgttaagtg	Exon 8	1309-1290	
4	561-JKF	cccttctcatgcagatgacg	Exon 5	782-801	0.528 kb
5	561-72R	cctcgctggaagtgttaagtg	Exon 8	1309-1290	
6	561-72F	gaagctgctgctgcggtac	Exon 8	1179-1197	0.712 kb
7	ST13-20	ctctgaggactcgctctcc	Exon 9	1890-1872	
6	561-72F	gaagctgctgctgcggtac	Exon 8	1179-1197	0.949 kb
9	ST13-22	ctgataaagaactgacctctg	Exon 9	2128-2108	
8	ST13-21	cggaaagcggagcgacaag	Exon 9	1836-1854	0.533 kb
10	ST13-24	aacctgtcccttctccttg	Exon 9	2368-2349	
11	T13-6	atcgctgaagccagtgagg	Exon 9	2189-2208	0.561 kb
12	581362F	tctaattttgaggcccggtc	Exon 9	2759-2740	
13	T13-11	aaagcagcatcgcgaccac	Exon 9	2490-2508	1.296 kb
14	aa173-3	gaagatgtctgcatgtacc	Exon 9	3786-3767	
15	581362R	gctttccctgggatcatctc	Exon 9	3697-3716	0.501 kb
16	aa173-5	gtgtttgcttttagccttgtc	Exon 9	4197-4177	
17	aa173-2	agtacaaggacaggaaggac	Exon 9	4076-4095	0.703 kb
18	aa1739F	gatccataaggcttttagttcc	Exon 9	4778-4758	
19	aa1739R	ctttcacggagccacctgg	Exon 9	4400-4418	0.521 kb
20	T13-RT1	tctgctcgctccctgtgatg	Exon 9	4930-4912	
21	T13-CF1	tcacagggacgagctcctg	Exon 9	4890-4908	0.853 kb
22	T13-CR3	aaagcccacttgaagccacg	Exon 9	5743-5724	
23	T13-CF2	gagtcgactcctattccacc	Exon 9	5395-5414	0.922 kb
24	T13-CR2	atcaggctagaggcaagcg	Exon 9	6317-6299	
23	T13-CF2	gagtcgactcctattccacc	Exon 9	5395-5414	1.347 kb
26	T13-RT3	cgccaggaagctatcttcc	Exon 9	6742-6722	
25	T13-RT2	cctgctgaagtctccacagag	Exon 9	6366-6386	0.909 kb
27	T13-CR1	agcttgtgcccacagtgttcg	Exon 9	7275-7256	
28	13EX3FA	tgcagagctcgagcctgag	Exon 9	7029-7047	0.995 kb
29	13EX3R1	gctcacaggatacgatcagc	Exon 11	8023-8004	
30	T13-RT4	tcgaataacctgcagatcagg	Exon 9	7745-7764	0.488 kb
32	561-4R	catgcggtcatacttgtcatc	Exon 12	8232-8212	
31	561-13F	ctcagcaagctccacatcc	Exon 9	7876-7894	0.375 kb
32	561-4R	catgcggtcatacttgtcatc	Exon 12	8232-8212	
33	3'Ext-1	tggtcgacgtcaacgacgac	Exon 13	8381-8400	-----
34	3'Ext-2	gagttggagaacggagtacg	Exon 13	8768-8787	-----
35	3'Ext-3	taacttgaaacactacagagc	Exon 13	9070-9090	-----

**Note.** The primers 3'Ext-1, 3'Ext-2, and 3'Ext-3 was used for 3'RACE

### 5.3.6 Analysis of T13 Transcript Sequence

A full length 9,282 bp transcript sequence of *T13* contains an ORF of 7,989 bp (**Figure 5.8**). Two possible methionine start codons, ATG, were observed. The first start codon was localized at the nucleotide positions 430-432 of the transcript sequence, and was preceded by a sequence of three in-frame stop codons, TGA TGA TGA, 43 bp upstream from the start site. The second start site was located 58 bp down stream from the first site. Comparison of these start sites and their surrounding nucleotide sequence to the Kozak consensus (**GCCA/GCCATGG**) (Kozak, 1989, 1991) showed that the second site (**AGCGACATGG**) conformed to the Kozak's more closely than the first site (**AGGACGATGC**). The ORF with the stop codon at positions 8419-8421 was followed by the 3' UTR of 863 bp with the polyadenylation signal, AATAAA, situated at nucleotide positions 9254-9259 of the transcript and was 13 bp before the polyadenylation site. The *T13* ORF was predicted to code for a high molecular weight protein of 2663 amino acids (if translated from the first start site, **Figure 5.8**). The *T13* transcript sequence was characterized by the adenine (A)-rich region between the nucleotide positions 1,500 and 5,000. Within this region, long sequences of A-tracts were found dispersed throughout the region.

### 5.3.7 T13 Protein Sequence Analysis

T13-protein sequence homology searches were performed using the BLASTP program to analyse the protein sequence database at NCBI. Overall, no significant homology to any known functional proteins was detected. However, partial homologies to a sequence of a protein with unknown function, as well as part of some known functional proteins were observed. The amino acid sequence of 343 residues extending from the N-terminus of T13 was highly homologous, 91% identical, to an unpublished sequence (366

amino acids) of a nasopharyngeal carcinoma susceptibility protein, LZ16 (Protein database accession no. NP-037407). Sequence homology was also identified to a 2,062-amino acid protein of unknown function, KIAA0874. The KIAA0874 protein sequence has been deduced from a sequence of cDNA isolated from human brain cDNA library (Nagase *et al.*, 1998). Two regions of high homology were observed between T13 protein and KIAA0874. The first was a 467 amino acids N-terminal region of T13, which was homologous to the N-terminal region, residues 1 to 484, of KIAA0874 (45% identical). The second region of homology was at the C-terminal sequences of both proteins (68% identical), between amino acid residues 2,322-2,664 of T13 and amino acids 1,730-2,062 of KIAA0874. In addition, a low degree of similarity was also observed between the remaining region of both proteins.

A sequence of 99 amino acids between residues 167 and 265 of T13 showed sequence homology to ankyrin repeats, found in a large number of proteins. The highest degree of homology was detected to the ankyrin domain of BRCA1-associated ring domain protein (BARD-1, accession no. XP-002364) as well as that of the protein product of *NFKBIL2* (nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor-like 2) gene, also known as *IkappaBR* gene (accession nos. AAH08782 and CAB63467). The sequence homology was 40% identical between T13, aa 151-307, and BARD1, aa 410-569, and 43% identical between amino acids 161-273 of T13 and 362-477 of *IkappaBR* protein. This ankyrin domain region was also detected by three protein domain prediction programs: SMART, Pfam, and Profile Scan. Within this domain of T13, three tandem ankyrin-like repeats were located between amino acids 167-199, 200-232, and 233-265. These ankyrin repeat sequences were also conserved in the homologous region of BARD1, *IkappaBR*, and KIAA0874. Significant homology (72% identical) was also observed between the residues 977 and 1056 of T13 protein and the mouse protein AK019393, a partial protein

---

**Figure 5.8** : Nucleotide transcript and deduced amino acid sequence of *T13*. The transcript sequence of 9282 bp is expanded up to 6 pages. The nucleotide and amino acid positions are indicated by numbers in the left hand column. The translation start site and the termination codon are indicated at positions 430-432 and 8419-8421 respectively. Three tandem repeats of in-frame stop codon (TGA) preceding the translation start site are located at positions 379-387 (*underline with asterisk underneath*). A polyadenylation signal beginning at nucleotide 9254 is underlined. The exon boundaries are indicated by the vertical lines with double-head arrows and the given numbers of exons. The corresponding trapped exons are also indicated: ET24.35, exon 2; 561-114, exon 5; 561-72, exon 8; 561-13, exon 9 (partial 3' sequence, nt 7858-7899 ); and 561-4, exon 12. Three tandem *ankyrin-like repeats* are located and underlined (**red**) between amino acids 167-199, 200-232, and 233-265. The *nuclear localization signals* (NLS) at residues 456-475, 906-923, 1184-1201, 1208-1236, 1358-1377, 1427-1444, 1460-1479, and 1640-1657 are also underlined (**black**)

1 GAGGGCC<sup>1</sup>TTTGCCGGGGAGGCGCCGCCGCCGAGCCCGCGGCCAGACGCCCCATCAGTAGCGTCCGCACCGGGAGC  
 76 CGCGGCTCTCGCCCGAGCCGTGGGCGCGCCCGAGGGGGCGGGCTCGCCT<sup>2</sup>CCCGCCGTCCCTCGCAGCTCTGCCGGG  
 151 CCCGAGCCCGCGCCGCCGCCGCCGCCCTT<sup>3</sup>GCCGCTCGGGCCGCGCGGCCCGGAAACCGGCGCGGGGTGCA  
 226 TGGGCAGCGCCCGCGCCCGCCGCTGAGCCGT<sup>4</sup>CGCGGAGCCGCGCAGCCCTCGGAGCAC<sup>5</sup>GAATATACAGCCCT  
 301 GCTCTGGGACACACCTCCATTGGATTTAAAGACAGTCTCTCGTCAGCACTGACTTTCAGCTATGGAATCGCAGAC  
 376 GGT<sup>6</sup>TGATGATGAA<sup>7</sup>AGCGCCGGCCGTGTA<sup>8</sup>AATGAAGATCGGGT<sup>9</sup>GAGGAGCAGGACGATGCCCAAGGGTGGGTGCCCT  
 1 \* \* \* M P K G G C P  
 451 AAAGCACCACAGCAGGAAGAGCTTCCCCTCAGCAGCGACATGGTGGAGAAGCAGACTGGGAAAAAGGATAAAGAT  
 8 K A P Q Q E E L P L S S D M V E K Q T G K K D K D  
 526 AAAGTTTCTCTAACCAAGACCCCAAAACTGGAGCGTGGCGATGGCGGGAAGGAGGTGAGGGAGCGAGCCAGCAAG  
 33 K V S L T K T P K L E R G D G G K E V R E R A S K  
 601 CGGAAGCTGCCCTTACC<sup>10</sup>CGCGGGCGCCAATGGGGAGCAGAAGGACTCGGACACAGAGAAGCAGGGCCCTGAGCGG  
 58 R K L P F T A G A N G E Q K D S D T E K Q G P E R  
 676 AAGAGGATTAAGAAGGAGCCTGTACCCGGAAGGCCGGGCTGCTGTTGGCATGGGGCTGTCTGGAATCCGAGCC  
 83 K R I K K E P V T R K A G L L F G M G L S G I R A  
 751 GGCTACCCCTCTCCGAGCGCCAGCAGGTGGCCCTTCTCATGCAGATGACGGCCGAGGAGTCTGCCAACAGCCCA  
 108 G Y P L S E R Q Q V A L L M Q M T A E E S A N S P  
 826 GTGGACACAACACCAAAGCACCCTCCAGTCTACAGTGTGTGAGAAGGGAACGCCCAACTCTGCCTCAAAAACC  
 133 V D T T P K H P S Q S T V C Q K G T P N S A S K T  
 901 AAAGATAAAGTGAACAAGAGAAACGAGCGTGGAGAGACCCGCTGCACCGAGCCGCCATCCGCGGGGACGCCCCG  
 158 K D K V N K R N E R G E T R L H R A A I R G D A R  
 976 CGCATCAAAGAGCTCATCAGCGAGGGGGCAGACGTCAACGTCAAGGACTTCGCAGGCTGGACGGCGCTGCACGAG  
 183 R I K E L I S E G A D V N V K D F A G W T A L H E  
 1051 GCCTGTAACCGGGGCTACTACGACGTGCGAAGCAGTGTGGCTGCAGGTGCGGAGGTGAACACCAAGGCCTA  
 208 A C N R G Y Y D V A K Q L L A A G A E V N T K G L  
 1126 GATGACGACAGCCTTTGCACGACGTGCCAACACGGGCACTACAAGTGGTGAAGCTGCTGCTGCGGTACGGA  
 233 D D D T P L H D A A N N G H Y K V V K L L L R Y G  
 1201 GGAACCCGCGAGCAGCAACAGGAAAGGCGAGACGCCGCTGAAAGTGGCCAACCTCCCCACGATGGTGAACCTC  
 258 G N P Q Q S N R K G E T P L K V A N S P T M V N L  
 1276 CTGTTAGGCAAAGGCACTTACACTTCCAGCGAGGAGAGCTCGACCGAGAGCTCAGAAGAGGAAGACGCCACCATCC  
 283 L L G K G T Y T S S E E S S T E S S E E E D A P S  
 1351 TTCGCACCTTCCAGTTCAGTGCAGCGCAACAACACGACTCCGAGTTCGAAAAAGGCCCTCAAGCACAAGGCCAAG  
 308 F A P S S S V D G N N T D S E F E K G L K H K A K  
 1426 AACCCAGAGCCACAGAAGGCCACGGCCCCCGTCAAGGACGAGTATGAGTTT<sup>11</sup>GATGAGGACGACGAGCAGGACAGG  
 333 N P E P Q K A T A P V K D E Y E F D E D D E Q D R  
 1501 GTTCTCCGGTGGACGACAAGCACCTATTGAAAAAGGACTACAGAAAAAGAAACGAAATCCAATAGTTTATCTCT  
 358 V P P V D D K H L L K K D Y R K E T K S N S F I S  
 1576 ATACCCAAAATGGAGGTTAAAAGTTACACTAAAAATAACACGATTGCACCAAAGAAAGCGTCCCATCGTATCCTG  
 383 I P K M E V K S Y T K N N T I A P K K A S H R I L  
 1651 TCAGACACGTCCGACGAGGAGGACGCGAGTGTCCCGTGGGGACAGGAGAGAAGCTGAGACTCTCGGCACATACG  
 408 S D T S D E E D A S V T V G T G E K L R L S A H T  
 1726 ATATTGCCTGGTAGTAAGACACGAGAGCCTTCTAATGCCAAGCAGCAGAAGGAAAAAATAAAGTAAAAAGAAAG  
 433 I L P G S K T R E P S N A K Q Q K E K N K V K K K



1801 CGAAAGAAAGAAACAAAAGGCAGAGAGGTTCGCTTCGGAAAGCGGAGCGACAAGTTCTGCTCCTCGGAGTCCGGAG  
458 R K K E T K G R E V R F G K R S D K F C S S E S E

1876 AGCGAGTCTCAGAGAGTGGGGAGGATGACAGGGACTCTCTGGGGAGCTCTGGCTGCCTCAAGGGGTCCCCGCTG  
483 S E S S E S G E D D R D S L G S S G C L K G S P L

1951 GTGCTGAAGGACCCCTCCCTGTTCAGCTCCCTCTCTGCCTCCTCCACCTCGTCTCACGGGAGCTCTGCCGCCAG  
508 V L K D P S L F S S L S A S S T S S H G S S A A Q

2026 AAGCAGAACCCAGCCACACAGACCAGCACACCAAGCACTGGCGGACAGACAATTGGAAAACCATTTCTTCCCCG  
533 K Q N P S H T D Q H T K H W R T D N W K T I S S P

2101 GCTTGGTCAGAGGTCAAGTCTTTATCAGACTCCACAAGGACGAGACTGACAAGCGAGTCTGACTACTCCTCTGAG  
558 A W S E V S S L S D S T R T R L T S E S D Y S S E

2176 GGCTCCAGTGTGGAATCGCTGAAGCCAGTGAGGAAGGCGAGGACAGGAAGCGAGCCTCCCTGTCCGAGAAG  
583 G S S V E S L K P V R K R Q E H R K R A S L S E K

2251 AAGAGCCCCTTCTGTCCAGCGCGGAGGGCGCTGTCCCAAACCTGGACAAGGAGGGGAAAGTTGTCAAAAAACAT  
608 K S P F L S S A E G A V P K L D K E G K V V K K H

2326 AAAACAAAACACAAAACAAAACAAAGGAGAAGGGACAGTGTTCATCAGCCAAGAGCTGAAGTTGAAAAGTTTT  
633 K T K H K H K N K E K G Q C S I S Q E L K L K S F

2401 ACTTACGAATATGAGGACTCCAAGCAGAAGTCAGATAAGGCTATACTGTTAGAGAATGATCTTTCCACTGAAAAC  
658 T Y E Y E D S K Q K S D K A I L L E N D L S T E N

2476 AAGCTAAAAGTGTAAAGCACGATCGCGACCCTTTAAAAAGAAGAGAAAAGTTAGCAAAATGAAATTAGAAGAA  
683 K L K V L K H D R D H F K K E E K L S K M K L E E

2551 AAAGAATGGCTCTTTAAAGATGAAAAATCACTGAAGAGAATCAAAGACACAAACAAAGACATCAGCAGGTCTTTC  
708 K E W L F K D E K S L K R I K D T N K D I S R S F

2626 CGAGAAGAGAAAGACCGTTTCAATAAAAGCAGAAAAGGAGAGATCGCTGAAGGAAAAGTCTCCGAAAAGAAGAAAA  
733 R E E K D R S N K A E K E R S L K E K S P K E E K

2701 CTGAGACTGTACAAGAGGAGAGAAAAGAAGAAATCAAAGACCCGCCCTCAAATTAGAGAAGAAGAATGATTTA  
758 L R L Y K E E R K K K S K D R P S K L E K K N D L

2776 AAAGAGCACAAAATTTCAAAGAGAAGGAGAAGATTTTAAAGAAGATAAAGAAAAGTCAAAGAAAAGGTT  
783 K E D K I S K E K E K I F K E D K E K L K K E K V

2851 TATAGGGAAGATCTGCTTTTGACGAATATTGTAACAAAATCAGTTTCTGGAGAATGAAGACACCAAATTTAGC  
808 Y R E D S A F D E Y C N K N Q F L E N E D T K F S

2926 CTTTCTGACGATCAGCGAGATCGGTGGTTTTCTGACTTGTCCGATTCATCCTTTGATTTCAAAGGGGAGGACAGC  
833 L S D D Q R D R W F S D L S D S S F D F K G E D S

3001 TGGGACTCGCCAGTGACAGACTACAGGGACATGAAGAGCGACTCTGTGGCCAAGCTCATCTTGAGACGGTGAAG  
858 W D S P V T D Y R D M K S D S V A K L I L E T V K

3076 GAGGACAGCAAGGAGAGGAGGCGGGACAGCCGGGCCCCGGGAGAAGCGAGACTACAGAGAGCCCTTCTTCCGAAAG  
883 E D S K E R R R D S R A R E K R D Y R E P F F R K

3151 AAGGACAGGGACTATTTGGATAAAAACTCTGAGAAGAGGAAAGAGCAGACCGAAAAGCATAAAAGTGTCCCTGGC  
908 K D R D Y L D K N S E K R K E Q T E K H K S V P G

3226 TACCTTTCCGAAAAGGACAAGAAGAGGAGAGAGTCCGCAAAGGCCGGGCGGGACAGAAAAGGACGCCCTGGAGAGC  
933 Y L S E K D K K R R E S A K A G R D R K D A L E S

3301 TGCAAGGAGCGCAGGACCGGCAGGGCCAAGCCCCGAGGAGGCGCACCAGGAGGAGCTGAAGGAGTGTGGCTGCAAG  
958 C K E R R D G R A K P E E A H R E E L K E C G C K

3376 AGTGGCTTCAAGGACAAGTCCGACGGCGACTTTGGGAAGGGCCTGGAGCCGTGGGAACGGCACCACCCAGCACGA  
983 S G F K D K S D G D F G K G L E P W E R H H P A R

3451 GAGAAGGAGAAGAAGGATGGCCCCGATAAGGAAAGGAAGGAGAAGACAAAACCAGAAAGATACAAAGAGAAATCC

1008 E K E K K D G P D K E R K E K T K P E R Y K E K S  
**3526** AGTGACAAGGACAAAAGTGAGAAATCAATCCTGGAAAAATGTCAGAAGGACAAAGAATTTGATAAATGTTTTAAA  
1033 S D K D K S E K S I L E K C Q K D K E F D K C F K  
**3601** GAGAAAAAGATACCAAGGAAAAACATAAAGACACACATGGCAAAGACAAAGAAAGGAAAGCGTCTCTCGACCAA  
1058 E K K D T K E K H K D T H G K D K E R K A S L D Q  
**3676** GGGAAAGAGAAGAAGGAGAAGGCTTCCCTGGGATCATCTCAGAAGACTTCTCTGAAAAAAGATGACAAGAAA  
1083 G K E K K E K A F P G I I S E D F S E K K D D K K  
**3751** GGCAAAGAGAAAAGCTGGTACATCGCAGACATCTTACAGATGAGAGTGAGGACGACAGAGACAGCTGCATGGGG  
1108 G K E K S W Y I A D I F T D E S E D D R D S C M G  
**3826** AGCGGGTTCAAGATGGGAGAGGCCAGCGACTTGCCGAGGACGGACGGCCTCCAGGAGAAGGAGGAAGGACGGGAG  
1133 S G F K M G E A S D L P R T D G L Q E K E E G R E  
**3901** GCCTATGCCTCCGACAGACACAGGAAGTCTTCTGACAAGCAGCACCCCTGAGAGGCAGAAGGACAAGGAGCCAGA  
1158 A Y A S D R H R K S S D K Q H P E R Q K D K E P R  
**3976** GACAGGAGAAAGGACCGAGGGGCTGCCGACGCGGGGAGAGACAAAAAAGAGAAAGTCTTTGAAAAGCACAAGGAG  
1183 D R R K D R G A A D A G R D K K E K V F E K H K E  
**4051** AAGAAGGATAAAGAGTCCACAGAAAAGTACAAGGACAGGAAGGACAGAGCCTCAGTGGACTCCACGCAAGATAAG  
1208 K K D K E S T E K Y K D R K D R A S V D S T Q D K  
**4126** AAAAATAAACAGAAGCTCCCCGAGAAGGCTGAAAAGAAGCACGCTGCCGAAGACAAGGCTAAAAGCAAACCTCAA  
1233 K N K Q K L P E K A E K K H A A E D K A K S K L K  
**4201** GAGAAGTCCGACAAAAGAACATTCAAAGGAGAGGAAGTCTCGAGAAGTCCGACGCGGAAAAAGCCCTGCTTGA  
1258 E K S D K E H S K E R K S S R S A D A E K A L L E  
**4276** AAGTTGGAAGAAGAGGCTCTCCATGAGTACAGAGAAGACTCCAACGATAAAAATCAGCGAGGTCTCCTCTGACAGC  
1283 K L E E E A L H E Y R E D S N D K I S E V S S D S  
**4351** TTCACGACCGAGGGCAGGAGCCGGGGCTGACTGCCTTCTGGAGGTCTCTTTCACGGAGCCACCTGGAGACGAC  
1308 F T D R G Q E P G L T A F L E V S F T E P P G D D  
**4426** AAGCCGAGGGAGAGCGCCTGCCTCCCTGAGAAGCTGAAAGAGAAGGAGAGGCACAGACACTCCTCATCTTCATCC  
1333 K P R E S A C L P E K L K E K E R H R H S S S S S  
**4501** AAGAAGAGCCACGACCGAGAGCGAGCCAAGAAAGAGAAGGCCGAGAAGAAAGAGAAGGGCGAAGATTACAAGGAG  
1358 K K S H D R E R A K K E K A E K K E K G E D Y K E  
**4576** GGCGGTAGCAGGAAGGACTCCGGCCAGTACGAAAAGGACTTCTGGAGGCGGATGCTTACGGAGTTTCTTACAAC  
1383 G G S R K D S G Q Y E K D F L E A D A Y G V S Y N  
**4651** ATGAAAGCTGACATAGAAGATGAGCTAGATAAAACCATTGAATTGTTTTCTACCGAAAAGAAAAGATAAAAATGAT  
1408 M K A D I E D E L D K T I E L F S T E K K D K N D  
**4726** TCCGAGAGAGAACCTTCCAAGAAAATAGAAAAGGAACTAAAGCCTTATGGATCTAGTGCCATCAACATCCTAAAA  
1433 S E R E P S K K I E K E L K P Y G S S A I N I L K  
**4801** GAGAAGAAGAAGAGAGAGAAAACACAGGGAGAAATGGAGAGACGAGAAGGAGAGGCACCGGGACAGGCATGCGGAT  
1458 E K K K R E K H R E K W R D E K E R H R D R H A D  
**4876** GGGCTGCTGCGGCATCACAGGGACGAGCTCCTGCGGCATCACAGGGACGAGCAGAAGCCCGCCACCAGGGACAAG  
1483 G L L R H H R D E L L R H H R D E Q K P A T R D K  
**4951** GACAGCCCGCCCCGCTGCTCAAAGACAAGTCCAGGGACGAGGGCCCCGAGGCTCGGCGATGCCAAACTGAAGGAG  
1508 D S P P R V L K D K S R D E G P R L G D A K L K E  
**5026** AAATTCAAGGACGGTGCAGAGAAAAGAAAAGGGCGACCCAGTGAAGATGAGCAACGGGAATGATAAGGTAGCGCCA  
1533 K F K D G A E K E K G D P V K M S N G N D K V A P  
**5101** TCCAAAGACCCAGGCAAGAAAGACGCCAGGCCAGGGAGAAGCTCCTGGGGGACGGCGACCTGATGATGACCAGC  
1558 S K D P G K K D A R P R E K L L G D G D L M M T S

5176 TTCGAGAGGATGCTGTCCAGAAAGACCTGGAGATCGAGGAGCGCCACAAGCGGCACAAGGAGAGGATGAAGCAA  
1583 F E R M L S Q K D L E I E E R H K R H K E R M K Q

5251 ATGGAGAAGCTGAGGCACCGGTCCGGAGACCCCAAGCTCAAGGAGAAGGCCGAAGCCGGCAGACGACGGGCGGAAG  
1608 M E K L R H R S G D P K L K E K A K P A D D G R K

5326 AAGGGTCTGGACATTCTGTCTAAGAAACCGCCGGGGCTGGACCCTCCATTTAAAGACAAAAAGCTCAAAGAGTCCG  
1633 K G L D I P A K K P P G L D P P F K D K K L K E S

5401 ACTCCTATTCACCTGCCGCGGAAAATAAGCTACACCCAGCATCAGGTGCAGACTCCAAAGACTGGCTGGCAGGC  
1658 T P I P P A A E N K L H P A S G A D S K D W L A G

5476 CCTCACATGAAAGAGGTCTGCCTGCGTCCCCAGGCCTGACCAGAGCCGGCCACTGGCGTGCCACCCTACG  
1683 P H M K E V L P A S P R P D Q S R P T G V P T P T

5551 TCGGTGCTATCCTGCCCCAGCTACGAGGAGGTGATGCACACGCCAGGACCCCGTCTGCAGCGCCGATGACTAC  
1708 S V L S C P S Y E E V M H T P R T P S C S A D D Y

5626 GCGGACCTCGTGTTCGACTGCGCCGACTCGCAGCACTCCACGCCCGTGCCACCCTCCACCAGCGCTGCTCC  
1733 A D L V F D C A D S Q H S T P V P T A P T S A C S

5701 CCCTCCTTTTTTCGACAGGTTCTCCGTGGCTTCAAGTGGGCTTTTCGAAAACGCCAGCCAGGCTCCTGCCAGGCCT  
1758 P S F F D R F S V A S S G L S E N A S Q A P A R P

5776 CTCTCCACAAACCTTTACCGCTCGGTCTCTGTTCGACATTAGGAGGACCCCGAGGAAGAATTAGCGTCCGAGAC  
1783 L S T N L Y R S V S V D I R R T P E E E F S V G D

5851 AAGCTCTTCAGGCAGCAGAGCGTTCTGCTGCCTCCAGCTACGACTCTCCATGCCACCCTCGATGGAAGACAGG  
1808 K L F R Q Q S V P A A S S Y D S P M P P S M E D R

5926 GCGCCCCTGCCCCGTTCCCGCGGAGAAGTTTGCCTGCTTGTTCGCCAGGGTACTACTCCCAGACTATGGCCTC  
1833 A P L P P V P A E K F A C L S P G Y Y S P D Y G L

6001 CCGTCGCCCCAAGTCGACGCTTTCGACTGCCACCAGGCTGCCGTTGTCACTGTACCCCGTCTCCAGAGGGCGTC  
1858 P S P K V D A L H C P P A A V V T V T P S P E G V

6076 TTCTCAAGTTTACAAGCAAAACCTTCCCCTTCCCCAGAGCCGAGCTGCTGGTTCCTTCCCTCGAAGGGGCCCTT  
1883 F S S L Q A K P S P S P R A E L L V P S L E G A L

6151 CCCCCGACCTGGACACCTCCGAGGACCAGCAGGCGACGGCCGCCATCATCCCCCGGAGCCCAGCTACCTGGAG  
1908 P P D L D T S E D Q Q A T A A I I P P E P S Y L E

6226 CCGCTGGACGAGGGTCCCTTCAGCGCCGTCATCACCGAGGAGCCCGTTGAGTGGGCCACCCTCCGAGCAGGGC  
1933 P L D E G P F S A V I T E E P V E W A H P S E Q A

6301 CTTGCCTCTAGCCTGATCGGGGGCACCTCTGAAAACCTGTGAGCTGGCCTGTGGGCTCGGACCTCCTGCTGAAG  
1958 L A S S L I G G T S E N P V S W P V G S D L L L K

6376 TCTCCACAGAGATTCCCCGAGTCCCCAAAGCGTTTCTGCCCCGCGACCCCTCCACTCTGCCGCCCCAGGGCCC  
1983 S P Q R F P E S P K R F C P A D P L H S A A P G P

6451 TTCAGCGCCTCGGAGGCGCCGTACCCCGCCCTCCCGCCTCTCCTGCCCCGTACGCTCTGCCCGTCTGAGCCG  
2008 F S A S E A P Y P A P P A S P A P Y A L P V A E P

6526 GGGCTGGAGGACGTCAAGGACGGAGTGGACGCCGTCCCCGCGCCATCTCCACCTCAGAGGCGGCTCCCTACGCC  
2033 G L E D V K D G V D A V P A A I S T S E A A P Y A

6601 CCTCCCTCCGGGCTGGAGTCTTCTTCAGCAACTGCAAGTCACTTCCGGAAGCCCCGCTGGACGTGGCCCCGAG  
2058 P P S G L E S F F S N C K S L P E A P L D V A P E

6676 CCCGCTGTGTAGCCGCTGTGGCTCAGGTGGAGGCTCTGGGGCCCCCTGAAAATAGCTTCTTGACGGCAGCCG  
2083 P A C V A A V A Q V E A L G P L E N S F L D G S R

6751 GGCCTGTCTCACCTCGGCCAGGTGGAGCCGGTGCCTGGGCGGACGCCTTCGCCGCCCCGAGGACGACCTGGAC  
2108 G L S H L G Q V E P V P W A D A F A G P E D D L D

6826 CTGGGGCCCTTCTCCCTGCCGAGCTTCCCCTGCAGACTAAAGATGCCGAGATGGTGAAGCGGAACCCGTGAA  
2133 L G P F S L P E L P L Q T K D A A D G E A E P V E

**6901** GAAAGTCTTGCTCCTCCAGAAGAGATGCCTCCAGGGGCCCCGGGGTCATAAACGGTGGGGATGTTTCCACCGTA  
 2158 E S L A P P E E M P P G A P G V I N G G D V S T V

**6976** GTGGCTGAGGAGCCGCCGGCACTGCCTCCTGACCAGGCCTCCACCCGGCTCCCTGCAGAGCTCGAGCCTGAGCCC  
 2183 V A E E P P A L P P D Q A S T R L P A E L E P E P

**7051** TCAGGGGAGCCAAAGCTGGACGTGGCTCTAGAAGCTGCGGTGGAGGCGGAGACGGTGGCCGAAGAGAGGGCCCGT  
 2208 S G E P K L D V A L E A A V E A E T V P E E R A R

**7126** GGGGATCCGGACTCCAGCGTGGAGCCCGCGCCCGTCCCCCAGAACAGCGCCACTGGGGAGCGGAGACCAGGGG  
 2233 G D P D S S V E P A P V P P E Q R P L G S G D Q G

**7201** GCTGAGGCTGAAGGCCCCCGCGCGTCCCTCTGTGCCCTGACGGCCCCGCCCCGAACACTGTGGCACAAGCT  
 2258 A E A E G P P A A S L C A P D G P A P N T V A Q A

**7276** CAGGCCGACAGCGGTGCCGGCCCCGAGGACGACACTGAGGCCTCCCGTGGCGCCGCCAGCCGAAGGCCCTCCT  
 2283 Q A A D G A G P E D D T E A S R A A A P A E G P P

**7351** GGCGGCATCCAGCCGAAGCCGCGAGAACCAAAACCCACGGCCGAAGCCCCGAAGGCCCCCCGAGTGGAGGAGATC  
 2308 G G I Q P E A A E P K P T A E A P K A P R V E E I

**7426** CCTCAGCGCATGACCAGGAACCGGGCGCAGATGCTCGGAACCAGAGCAAGCAGGGCCCCCCCCCTCCGAGAAG  
 2333 P Q R M T R N R A Q M L A N Q S K Q G P P P S E K

**7501** GAGTGGCCCCCACCCCTGCCCGGTACCAGGGCCAAGGCCCGGGCTCCGAGGACGACGACGCCAGGCCAG  
 2358 E C A P T P A P V T R A K A R G S E D D D A Q A Q

**7576** CATCCGCGCAAACGCCGCTTTTCAGCGCTCCACCCAGCAGCTGCAGCAGCAGCTGAACACGTCCACGCAGCAGACG  
 2383 H P R K R R F Q R S T Q Q L Q Q Q L N T S T Q Q T

**7651** CGGGAGGTGATCCAGCAGACGCTGGCCGCCATCGTGGACGCCATCAAGCTGGATGCCATCGAGCCCTACCACAGC  
 2408 R E V I Q Q T L A A I V D A I K L D A I E P Y H S

**7726** GACAGGGCCAACCCCTACTTTCGAATACCTGCAGATCAGGAAGAAGATCGAGGAGAAGCGCAAGATCCTGTGCTGT  
 2433 D R A N P Y F E Y L Q I R K K I E E K R K I L C C

**7801** ATCACGCCGAGGCGCCCCAGTGCTACGCCGAGTACGTACCTACACGGGCTCCTACCTCCTGGACGGCAAGCCG  
 2458 I T P Q A P Q C Y A E Y V T Y T G S Y L L D G K P

**7876** CTCAGAAGCTCCACATCCCCGTGATCGCACCCCCCTCCCTCCCTGGCGGAGCCCCTGAAGGAGCTGTTTCAGCGAG  
 2483 L S K L H I P V I A P P P S L A E P L K E L F R Q

**7951** CAGGAGGCCGTCCGGGAAAGCTGCGTCTACAGCACAGCATCGAGAGGAGAAGCTGATCGTATCCTGTGAGCAG  
 2508 Q E A V R G K L R L Q H S I E R E K L I V S C E Q

**8026** GAGATTCTGCGGGTTCACTGCCGGGCGGCCAGGACCATCGCCAACCAGGCAGTGCCATTGACGACCTGCACGATG  
 2533 E I L R V H C R A A R T I A N Q A V P F S T C T M

**8101** CTGCTGGACTCCGAGGTCTACAACATGCCCTGGAGAGCCAGGGTACGAGAACAAGTCAGTGGCGGACCCGTTTC  
 2558 L L D S E V Y N M P L E S Q G D E N K S V R D R F

**8176** AACGCCCAGTTCATCTCCTGGCTCCAGGACGTGGATGACAAGTATGACCGCATGAAGACTTGCCTCCTCATG  
 2583 N A R Q F I S W L Q D V D D K Y D R M K T C L L M

**8251** CGGCAGCAGCACGAGGCCGCGGCCCTGAACGCCGTGCAGAGGATGGAGTGGCAGCTGAAGGTGCAGGAACCTGGAC  
 2608 R Q Q H E A A A L N A V Q R M E W Q L K V Q E L D

**8326** CCCGCCGGGCACAAGTCCCTGTGCGTGAACGAGGTGCCCTCCTTCTACGTGCCCATGGTCGACGTCAACGACGAC  
 2633 P A G H K S L C V N E V P S F Y V P M V D V N D D

**8401** TTTGTATTGTTGCCGGCATTGACACCCGCGGGACGGCCGACGAGGCGAGGGCCGCACGGCTGCCAGGACTG  
 2658 F V L L P A \*

**8476** CTGCTGAGCCCCAGGGGCGGAGGAGGAGCGCCCTGTCCACCCGGGCGGGGAGAGACCCCGAGAGAGACTGCACT  
**8551** TGGCCACAGCGCCTCCATCCCTTCCAGACTGGTCCAGACGTGCGGGAGGTGAAACACGTCTCTTCTTACCAGCT  
**8626** GCCGCGCGGGGGCAAAGCCCCAGAGCCTCACTGCCCCGGCGGGATGAGGAGACCATGCCGGCGCGCACACA  
**8701** CGGCAGCTTCTGTTGAAAACGGCGACTTCGGGCCCTTTTCTCCAGCAACTGCAGAGGCATTTTCAGGAGTTGGA

8776 GAACGGAGTACATTTTTAAATTTTTGTTCCATTCTGATCAACTAAAGAAAATTAAGGTGTCCACCTCTCACAA  
8851 ACAGTGACCTGCAGGTCGGAGGGGCAGGAGCCCATCCCAGCCGTGGTTCTGTTGCCGCCGAGCTGCGACGGCCT  
8926 CAGACTGGGCTTCAGACCTGGTGC GGAGCGAGGCGGAGGGACACAGAGCCAGGACTCCCAGCCGTATTGAAATGG  
9001 AGTCAAATCCGCGTGGTTTGTATCTTCTTTCTTTAATCTTGGGCATTTTGTCTCTTTCTGAGAACGTAACTT  
9076 GAAACACTACAGAGCCAATAATCATATAAAAAGTGTATCCGTAAACGCTGTACAGTTTATATACCTTCCAATG  
9151 TACTTTTGCTGCCTTTTAAATAATTTATTTTGCAACCCATTACTGCCCAGTTGTAAATAACTTATGTACTGTAAC  
9226 ATCACTTCAGTTATGTGTAAAAAATAATAAATTAATTTAAATGCAAAAAAAAAA 9282

sequence of 485 amino acids. Such a high degree of homology of this mouse protein to T13 indicated that it might be derived from a mouse homologue of human T13.

Sub-cellular localization of T13 protein was also predicted utilizing a program PSORT. Eight bipartite nuclear localization signals (NLS) were identified within the region between amino acids 457 and 1,658. These NLS were located at residues 456-475, 906-923, 1184-1201, 1208-1236, 1358-1377, 1427-1444, 1460-1479, and 1640-1657. Further *in silico* protein sequence analysis did not detect any other functional motifs or specific sequences that could suggest a possible function of the T13 protein.

### 5.3.8 T13 Gene Structure

The genomic structure of *T13* was characterized using the available genomic sequence information of two overlapping PAC clones, 74015 and 168P16, encompassing this region and overlapping with BAC 561E17 (**Figure 5.3**). Alignment of this genomic sequence with the sequence of *T13* full-length cDNA allowed the identification of T13 gene structure. The *T13* gene consists of 13 exons ranging in size from 85 bp to 6.53 kb, and spanning the genomic region of approximately 220 kb. The smallest exon is exon 2 (85 bp) while the largest is exon 9 of 6.53 kb. The sequences of the splice junctions of each exon, summarised in **Table 5.4**, conform to the published consensus sequence of donor and acceptor splicing sites (Shapiro and Senapathy, 1987). Both ATG initiation codons (section 5.3.6) are contained within exon 3. The ankyrin repeat motif was encoded by the sequence of exons 6, 7, and 8, while all eight NLSs were coded for by the sequence of exon 9. Exon 13 contained the 3' end of the ORF, the stop codon as well as the complete 863 bp 3' UTR (**Figure 5.8**). The sequence of a partial pseudogene on the X chromosome that was homologous to part of the transcript sequence of *T13* (section 5.3.4.1) included the sequence from part of exon 9 to exon 13.

---

***Figure 5.9*** : Genomic Organization of *T13*. Diagram represents the regional physical map of the chromosome 16 at the band 16q24.3. The map consists of the overlapping cosmids, BAC, and PACs, corresponding with those shown in *Figure 5.3*, and encompasses the *T13* gene. The restriction map of *EcoRI*, **E**, and *EagI*, **Ea**, is indicated in “kb”. Based on the physical map, the T13 gene expands over 170 kb of the genomic region, 5’ to 3’ from telomeric to centromeric direction. Exons 2 and 3 are separated by the largest intron of approximately 100 kb. The map locations of the trapped products, *filled circles*, and their corresponding exons are also indicated.

← **CEN**

**TEL** →



*T13 Sequence*

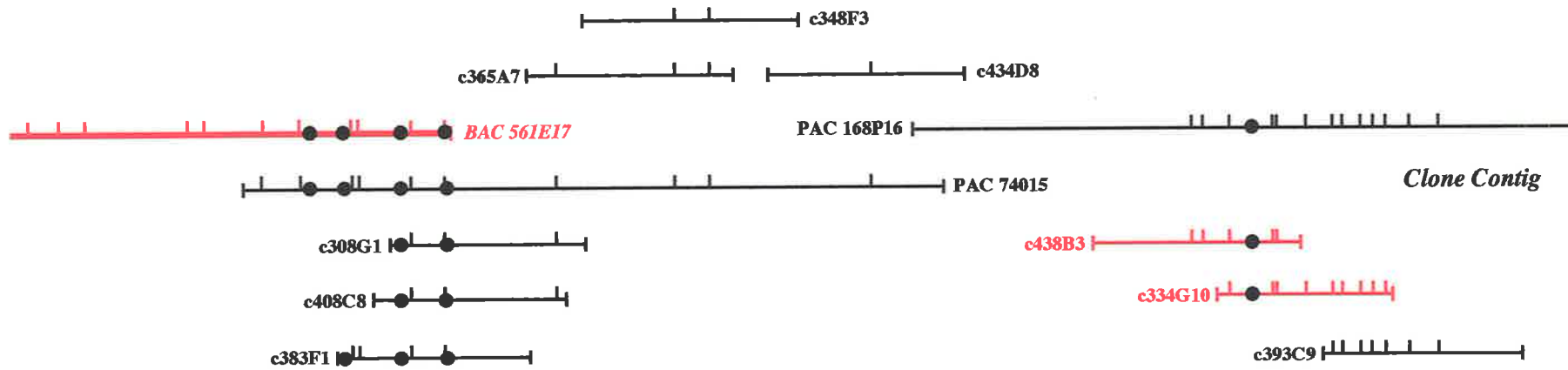
(ex12) (ex9) (ex8) (ex5)  
561-4 561-13 561-72 561-114

(ex2)  
ET24.35

*Trapped Exons*



*Restriction Map*



*Clone Contig*



**Table 5.4**  
**Nucleotide Sequence of the *T13* Gene Exon and Intron boundaries**

<i>Exon</i>	<i>3' Splice site (intron/exon)</i>	<i>Exon size (bp)</i>	<i>5' Splice site (exon/intron)</i>	<i>Intron size (kb)</i>
1	5' GAGGGCCTTTGCCGGGA	1-285	TCGGAGCACGgtgagaggcgc	71.936
2	cttgtttcagAATATAACA	286-370	TATGGAATCGgtaaataactca	101.349
3	tcttttgaagCAGACGGTTG	371-516	TGGGAAAAAGgtaatgctgcc	11.577
4	tcttttctagGATAAAGATA	517-655	TCGGACACAGgtaccagcccg	14.013
5	cgcctcacagAGAAGCAGGG	656-826	AACAGCCCAGgtgagcccggc	0.179
6	tggcctgcagTGGACACAAC	827-1030	GACTTCGCAGgtgggcaccct	1.935
7	ctgtttccagGCTGGACGGC	1031-1173	GCACTACAAGgttggcatccc	2.334
8	tgtcttgtagGTGGTGAAGC	1174-1321	AGCTCGACGGgtaagtcacgg	0.397
9	cctccaacagAGAGCTCAGA	1322-7899	CATCCCCGTGgtgagtgcggg	3.880
10	tgccctgtcagATCGCACCCC	7900-7998	CATCGAGAGGgtaagtgggct	0.132
11	cctccttcagGAGAAGCTGA	7999-8142	GGAGAGCCAGgtagggcccgt	3.898
12	gtttccacagGGTGACGAGA	8143-8235	CCGCATGAAGgtcgggtgttgc	2.148
13	tgtcccgcagACTGCCTCC	8236-9272	CGGCATGACA <sup>3' UTR</sup> --TTAAAATGC 3'	3' UTR
<i>Splicing consensus</i>	<u>ttctttncagG</u> ----- cctccc		-----AGgt <u>aagt</u> g	

**Note:** The sequence of "intron 1" of 71.9 kb was obtained from the complete human genome sequence of this region, available recently, and therefore is the second largest intron instead of intron 4 (section 5.3.8)

The trapped exon ET24.35 was trapped from exon 2, while 561-114 and 561-72 were trapped from exons 5 and 8 respectively. As described in sections 5.3.2 and 5.3.4.1, clone 561-114, size 268 bp, was likely trapped from two adjacent exons. The first exon was the sequence of 97 bp from 5' end of 561-114, which appeared only in some cDNA clones. It was located and separated from the second, 171-bp *T13* exon 5, by the intervening sequence of 497 bp. This extra exon was assigned as exon 5E and was bound at both ends by the sequences corresponding with the acceptor and donor splicing consensus. The trapped product 561-13 was isolated from the 3' part of exon 9, but in the reverse orientation that included part of its intron (described in 5.3.2 and 5.3.4.1). Exon 12 was represented by the trapped product 561-4.

The *T13* gene contains an unusual large intron between exons 2 and 3, which is approximately 90 kb across the genomic sequence. Intron 4 appears to be the second largest intron of *T13*, spanning the genomic sequence of 14 kb. The genomic structure of *T13* and its organization in the region extending from the published physical map of 16q24.3 (Whitmore *et al.*, 1998a) is shown in **Figure 5.9**.

### **5.3.9 T13 Alternative Splicing and Overlapping Transcripts**

During the construction of the extended physical map of 16q24.3 (Whitmore 1999, unpublished data), two groups of partial cDNA sequences, designated as *T16* and *T17* were mapped to the same region where *T13* exons were isolated. With the available genomic sequence of *T13*, both *T16* and *T17* were precisely located. The *T16* was localized within intron 4 of *T13* along with *T13* exon 5E, and trapped product 561-33, while *T17* situated within intron 7 (**Figure 5.10**). Sequence analysis of cDNA clone os05c06 representing *T16* defined that this partial transcript sequence was not part of *T13*. Instead, it appeared to be transcribed in the opposite direction with that of *T13*. The trapped exon 561-33 also

showed an ORF and splicing signal at the boundary sequence that suggested its transcription in the opposite direction with *T13*. To test if 561-33 was part of *T16*, RT-PCR analysis using primer pairs designed from both *T16* and 561-33 sequence was carried out. This, however, failed to link this trapped exon to *T16* sequence. Both *T16* and 561-33 might, therefore belong to the different genes overlapping with *T13*. This was omitted from further study.

*T17* is a partial 0.9-kb cDNA sequence generated from overlapping ESTs. The majority of EST sequences are derived from the 3' end of the transcript and contain the polyadenylation signal. The location of this polyadenylation signal indicates that *T17* is likely to be transcribed in the same direction as *T13*. The *T17* might be part of *T13* as an alternative isoform or be an independent gene that shares some exons with *T13*. To further investigate this relationship, RT-PCR experiment was undertaken utilizing various combinations of primers designed to both *T13* exons and *T17* sequences. The overall results of sequence analysis of the RT-PCR products are summarised in **Figure 5.10**. *T17* appeared to be an alternative transcript or transcription isoform of *T13* that selectively used the 3' terminal exon situated within intron 7 of *T13*. Northern blot hybridization was also undertaken to confirm the existence of this isoform (*T17*) utilizing RT-PCR product generated from the assembled 0.9-kb sequence of *T17* cDNA. As shown in **Figure 5.7**, the *T17* probe did not detect the 9.5 kb mRNA band of *T13* but did detect the smaller mRNA bands similar to those detected by a cDNA probe generated from exons 6, 7, 8 and 9 of *T13*.

Alternative inclusion of exon 5E (section 5.3.8) was also confirmed by RT-PCR either in the full-length *T13* transcript or in its *T17* isoform, though the quantity of 5E inclusion products was lower when compared with the products that excluded this exon. Alternative splicing was also detected between exons 2 and 3, which generated the product that either

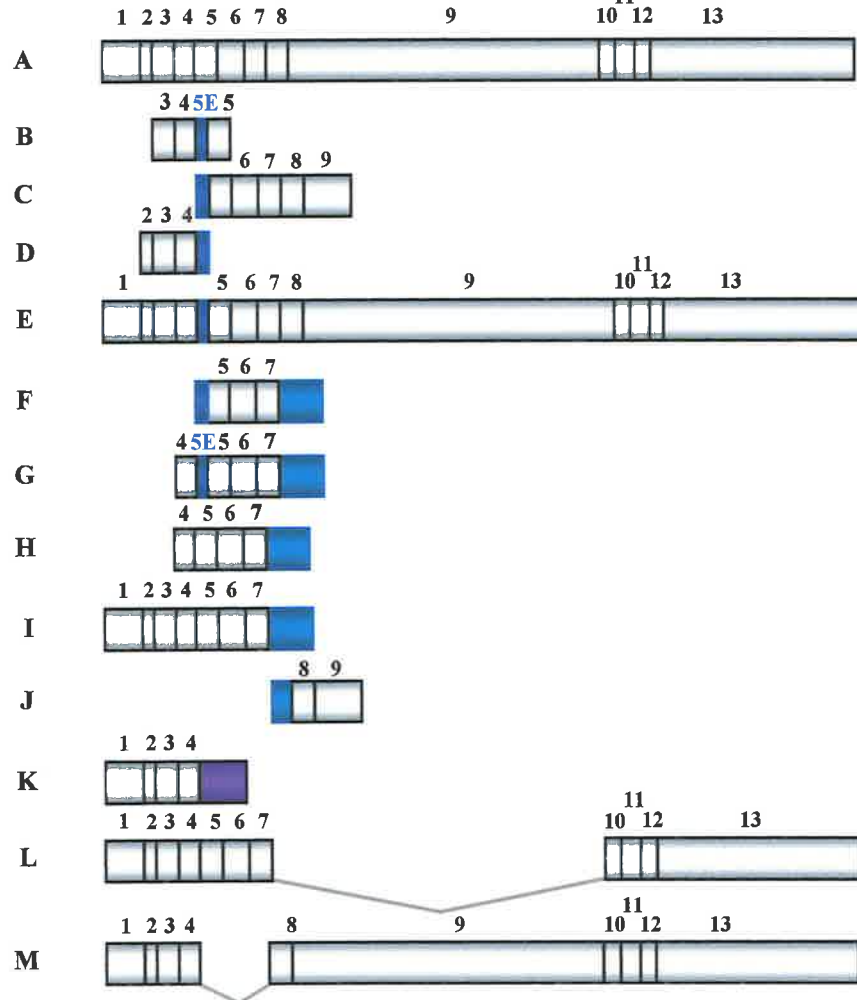
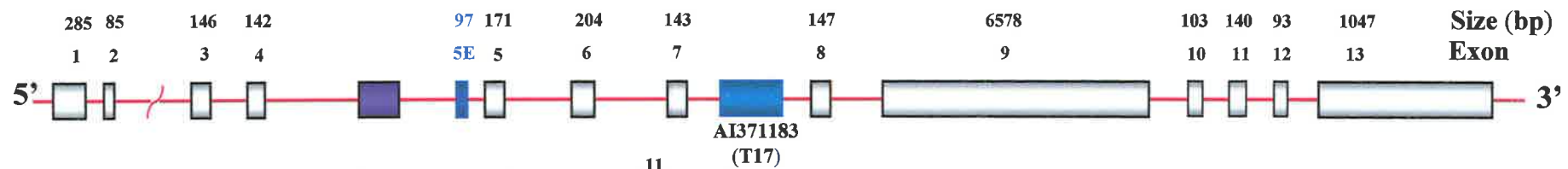
included or excluded the CAG tri-nucleotide. Analysis of the genomic sequence flanking exon 3 identified two tandem CAG sequences at the 5' splicing boundary of the exon. These tandem CAG sequences could cause the selective usage of 5' splice acceptor and result in the inclusion or exclusion of CAG between exons 2 and 3. Other splicing variants were also detected during the RT-PCR experiment, but with low quantity. These, however, were probably the minor products at least in three tissue sources of mRNA used in this study: Jurkat T-cell, fetal heart, and mammary gland.

Additional transcription isoforms of *T13* existed that were represented by cDNA sequences in the GenBank cDNA database. The diagram showing these cDNA isoforms comparing with that of *T13* is given in **Figure 5.10**. The first was an isoform represented by a 1460-bp cDNA clone, gene identification no. 20562097 (accession no. XM-085409). Sequence analysis revealed the transcript sequence identical to that of *T13* from exon 1 to exon 4, then followed by the new exon situated within intron 4. This cDNA sequence completed with an ORF of 124 amino acids, utilizing the same ATG start codon as that of *T13*. The second isoform, which was represented by a partial cDNA sequence (EST accession no. BI836550), showed the sequence of *T13* that joined exon 7 with exon 10 by skipping exons 8 and 9. This transcription product maintained an ORF of *T13* from start to stop codons except part of two skipped exons. It was predicted to code for a protein of 421 amino acids that shares the first 248 residues with that of *T17* and includes ankyrine repeats region. The third isoform was represented by a partial cDNA sequence (accession no. AI684375) of 509 bp isolated from leukemic B-cell library. This clone showed the sequence that is homologous to *T13* from exon 1 to exon 4 and joins to exon 8 by skipping exons 5, 6, and 7.

### 5.3.10 Mutation Study of T13 in Sporadic Breast Cancer

---

**Figure 5.10 : Alternative Splicing of T13.** Diagrams represent the results of RT-PCR analysis of T13 and T17 alternative splicing as well as the alternative isoforms found in cDNA database. The top diagram represents the organization of T13 exons with the exon number and size indicated above each exon. The size of exons and introns was not drawn to scale. The EST AI371183, situated within intron 7, represents a partial cDNA sequence of T17. Exon 5E was part of the trapped product 561-114 (exon 5) situated within intron 4, xx bp proximal to the exon 5. RT-PCR products were obtained by using various combination of primers and are represented by the diagrams B, C, D, F, G, H, and J. Each product represents the derivative of *T13* cDNA which also indicates the including exon numbers. The identity of the primers used are given on the right hand side of each product. The products B, C, and D represent the cDNA that include *Exon 5E*. This can be drawn as the T13 transcript with exon 5E, diagram E. *T17*, by RT-PCR analysis, product H, was shown that it is the *T13* derivative. Likewise, the cDNA products F and G represent the *T17* isoform that include exon 5E. Product J was obtained by using T17-forward primer and primer from T13 exon 9. This product may represent the isoform that contain part of T17 in between exons 7 and 8. The cDNA sequence XM085409, BI836550, and AI684375 were obtained from EST database. XM085409 contains T13 sequence from exons 1-4 and the sequence at the 3' part from within intron 4, diagram K. The cDNA BI836550 contains part of T13, exon 6, 7, 10, and 11, showing exons 8-9 skipping. This cDNA may represent the T13 isoform with exons 8-9 skipping (diagram L). EST AI684375 represents the T13 derivative with exons 5-7 skipping (diagram M).



**T13 Full-length Transcript**

T13-1(*F*), exon 3 / T13-8(*R*), exon 5

114E-F, exon 5E / T13-16@, exon 9

24.35F, exon 2 / 114E-R, exon 5E

**T13 Full-length Transcript, include exon 5E**

114E-F, exon 5E / T17-5089(*R*)

251293-T3(*F*), exon 4 / T17-5089 (*R*)

**Predicted T17 transcript**

T17-3F / T13-16(*R*)

**XM085409**

**BI836550**

**AI684375**

Study of *T13* for its mutation in sporadic breast cancer was performed by other members of the chromosome 16 / breast cancer group (Lower, K.M., Whitmore, S.A., Crawford, J., McKenzie, O., and Bais, A.J.). PCR-SSCP analysis of all *T13* exon and subsequent sequencing of the products that show mobility band shift were used to screen for *T13* mutation in 48 breast cancer samples showing either restricted 16q24.3 LOH or complex LOH of 16q that includes 16q24.3. In addition screening was also conducted in 24 breast cancer cell lines. The overall results shown that by the current methods there were no mutations detected in these breast cancer samples as well as cell lines.

## **5.4 Discussion**

The principle aim of the present study was to (i) explore the transcribed sequence within the proximal region extending from the previously reported 16q24.3 LOH physical map, (ii) characterize a gene or genes that may be considered to be a candidate for breast cancer related tumorigenesis. Within this region, BAC 561E17 was the largest genomic clone and was chosen for this study.

### **5.4.1 Exon trapping**

Initially there was no genomic sequence of this BAC available, exon amplification procedure was therefore utilized to provide the basis for gene identification. Exon amplification or “exon trapping” involves sub-cloning of the restriction-digested fragments of genomic DNA into an exon trapping vector, pSPL3B. Recombinant clones are subsequently isolated and transfected into the mammalian cell line, COS7. The multiple cloning site of the exon trapping vector is designed such that it is flanked by a short sequence characteristic of exons to provide both “5’ donor” and “3’ acceptor” splicing signals (*Figure 5.1*). Upon transfection into the COS7 cell line, the primary RNA

transcribed from the recombinant vector is processed by the splicing machinery of the host cell and allows splicing between the flanking exons and any potential exon contained within the DNA insert. If there is no potential exon sequence within the DNA insert, splicing will only occur between the flanking exons of the multiple cloning site and thus result in the “vector-only” product. Finally, the trapped exon sequences are amplified by RT-PCR from total RNA isolated from the host cell. The RT-PCR primers used are generated from the flanking exon sequences to ensure the amplified products are free from host transcript sequences. These PCR products can subsequently be isolated, by cloning, for further analysis. The advantage of exon trapping is that it can be used to isolate potential transcribed sequences contained within genomic DNA in the absence of any prior knowledge of tissue specific gene expression.

A total of 14 individual exons were identified from the BAC 561E17. The size of this BAC is approximately 140 kb, therefore exons were trapped at an average of 1 exon for each 10 kb. This success rate was similar to previous exon trapping from the 900-kb physical map of 16q24.3 (Whitmore *et al.*, 1998a) as well as those from other chromosomes (Burn *et al.*, 1996; Chen *et al.*, 1996; Church *et al.*, 1997). Due to the large number of clones obtained, only one to three clones representing a particular size group were chosen for sequence analysis. It would be likely that analysis of more clones within each size group would result in identification of additional different exons.

The proportion of “vector-only” products detected by colony PCR was approximately 30 %. These are mostly generated when the genomic insert does not contain an exon. In addition, vector-only products are also generated when the genomic insert contain an “incomplete exon”. A complete or “functional” exon refers to an intact exon sequence with flanking 5’ splice donor and 3’ splice acceptor sites. Restriction digestion of genomic clones that eliminated either of these splice sites from the exon could lead to an



incomplete exon sequence. In both cases splicing will occur between the flanking exons of the multiple cloning site. This will create a restriction *Bst*XI site (*Figure 5.1*) in the vector-only product. Indeed, the trapping procedure is designed to eliminate such products by the *Bst*XI digestion (section 5.2.1.9). It is possible that DNA digestion with *Bst*XI was not complete. This was probably due to a high percentage of these type of insert as well as some impurity that interfered the digestion reaction. This could be improved by purification of primary PCR products prior to digestion with the *Bst*XI enzyme. *font*

Two individual exons were trapped from the *CDH15* gene, which has previously been assigned to this physical region of 16q24.3 through somatic cell hybrid mapping (Kaupmann *et al.*, 1992; Kremmidiotis *et al.*, 1998). The mapping of these exons to the BAC 561E17, in the present study, therefore provided a refined physical location of this gene. The genomic structure information of *CDH15* was not available at this time. Alignment of the *CDH15* cDNA sequence with the available genomic sequence of this region showed that these two trapped exons corresponded to the exons 2 and 3 of this gene.

A group of partial cDNA sequences represented by clone IMAGE: 1098126 was also mapped to this BAC and was later identified to be the 2.3 kb full length transcript of *T18* gene (chapter 6). However, no exons were trapped from the *T18*. Since the restriction enzymes used in this exon trapping are *Bam*HI, *Bgl*II, and *Pst*I. The distribution of these restriction enzyme sites within this gene may be a result that lead to the generation of either large restriction fragments, inefficiently ligated into the pSPL3B vector, or small restriction fragments containing incomplete exon sequences. Therefore, exon trapping would not result in the identification of exons from *T18* gene.

A number of exons were trapped as a result of alternative usage of incorrect splice sites. The first example was exon 561-88. This exon was trapped from the *CDH15* gene corresponding with the nucleotides 120-278 of the cDNA (exon 2) but also included part of

the adjacent 5' intron. Similarly, exon 561-17 was trapped from *CDH15* exon 3, nucleotides 279-434, that also included part of its flanking 3' intron. The inclusion of intron sequences seen in both trapped products appears to interrupt the open reading frame of the transcript and are therefore likely to be a splicing error due to the presence of alternative splice signals. Alternative splice signals were also utilized in clone 561-13. This clone was a result of cloning of genomic fragment that included part of *T13* exon 9 in the reverse orientation.

#### **5.4.2 Characterization of Transcription unit 13 (T13)**

Identification of the transcribed sequence from the BAC 561E17 by exon trapping was based on sequence homology to expressed sequence in the NCBI database and initially identified five potential transcription units, including that which formed part of *CDH15*. Two of these, transcripts B and C, were identified from two independent overlapping sets of expressed sequences from both human and mouse. These transcripts were mapped in close proximity at the distal region of BAC 561E17. They were subsequently identified as forming part of the transcription unit 13 (*T13*). The transcript B was characterized as the 3' part of *T13*. Finally, a combination of these two transcripts (B and C), and computerized analysis of the genomic sequence encompassing this region, lead to the assembly of full-length transcript sequence of 9,282 bp of *T13*.

During the course of *T13* transcript sequence assembly, a large number of expressed sequences from the GenBank database that matched *T13* were found, mostly within the first 5 kb of *T13* 5' sequence. Analysis of the *T13* transcript sequence identified a number of long adenosine nucleotide (A) tracts distributed within the region between 2 and 5 kb from the 5' end. It is suggested that these poly (A) tracts allowed binding of the oligo dT primer generally used in the reverse-transcription step for cDNA library construction.

Accordingly, a large number of cDNA derivatives of T13 may have been reverse-transcribed from these long A tracts. Furthermore, due to the large transcription size of T13 it would be unlikely that reverse-transcription from the poly(A) tail would generate a full-length transcript. Together, these could result in the lack of homologous ESTs of the region down-stream from the first 5 kb of *T13*.

The *T13* transcript was predicted to code for a large protein of 2663 amino acid residues. T13 protein contains three tandem ankyrin repeats at the region close to the N-terminus. Ankyrin repeats are one of the most common protein sequence motifs, and consist of about 33 residues (Sedgwick and Smerdon, 1999). They have been found in various functionally diverse proteins, such as signal transduction and transcriptional regulators, Cdk inhibitors, cytoskeletal organizer, developmental regulators, and toxins. These protein motifs mediate “protein-protein interactions” in a number of different biological systems, from microbes to humans (Sedgwick and Smerdon, 1999). The presence of ankyrin repeat motifs within the T13 protein molecule therefore provides the primary evidence that the biological function of this protein requires interaction with an as yet unknown protein partner or partners. In addition to these ankyrin motifs, T13 protein also possesses eight bipartite nuclear localization signals (NLS) suggesting it is a protein located in the nucleus. Together, these indicate that T13 protein may be involved in transcriptional processes or in the regulation of other nuclear activities by interacting with other nuclear proteins.

Partial sequence homology was detected between T13 protein and two known human proteins, BRCA1-associated RING domain protein 1 (BARD1) and a protein member of nuclear factor kappa *B* (NF- $\kappa$ B) inhibitors, IkappaBR (I-kBR). Interestingly, homology was between the ankyrin domain region of these three proteins, with all three ankyrin repeats showing homology. It is possible that homologous ankyrin domains may allow

prediction of these protein that involve in the specificity of protein-protein interaction or be found among the functionally related protein. For instance, BARD1 possess three ankyrin repeats that mediates its BRCA1-independent function. BARD1 interacts with NF- $\kappa$ B, via Bcl-3, and modulates its transcriptional activity (Dechend *et al.*, 1999). This interaction involves binding of BARD1 ankyrin domain to Bcl-3. The inhibitors of NF- $\kappa$ B (I- $\kappa$ B) on the other hand inactivates NF- $\kappa$ B by binding to and preventing nuclear localization of NF- $\kappa$ B and hence DNA binding (Chen *et al.*, 2001). Binding of I- $\kappa$ B to NF- $\kappa$ B is also mediated by the ankyrin repeat region. As a member of I- $\kappa$ B family, I- $\kappa$ BR has also been demonstrated to interact with the p65 subunit of NF- $\kappa$ B and partially inhibit NF- $\kappa$ B DNA binding activity (Ray *et al.*, 1995). Given that the T13 protein possesses three ankyrin repeats homologous to those of BARD1 and I- $\kappa$ BR, it may be functionally related to or involve in the regulation of NF- $\kappa$ B activity. This, however, is speculative and requires further investigation.

A protein of unknown function, KIAA0874, was the only human protein in the database that was similar to T13. Similarity was high at both the N-terminal and C-terminal regions of these two proteins. The remaining regions, although showing a lower degree of homology, also contained a number of nuclear localization signals (NLS) similar to that of T13. KIAA0874 is a high molecular weight protein of 2062 amino acids. Because of these similarities KIAA0874 is likely to be paralogous to T13 and may be functionally related to the T13 protein. Protein sequence homology was also detected between the first 343 amino acids of T13 and LZ16, a protein sequence that is in the NCBI database as a “nasopharyngeal carcinoma susceptibility gene product” (Protein database accession no. NP-037407). However, analysis of the 1603-bp *LZ16* transcript sequence (GenBank accession no. AF121775) showed that *LZ16* was likely to be an incomplete *T13* cDNA that was reverse transcribed from the poly (A) tract between nucleotides position 1778 and

1809 of *T13*. A similar situation was also detected in another version of *LZ16* mRNA of 2557 bp (GenBank accession no. BC017437). This *LZ16* appeared to utilize the poly (A) tract between position 2512 and 2556 of *T13*. In addition both *LZ16* mRNAs did not possess polyadenylation signal characteristic of independent mRNAs, therefore confirming that both were derivatives of the *T13* transcript.

The *T13* gene consists of 13 exons and encompasses a genomic region of approximately 120 kb. Intron 2 and intron 4 appear to be the first and the second largest intron, while exon 9 is the largest exon. Alternative splicing that generated a number of transcription isoforms or aberrant transcript of both major and minor appeared to involve the sequence within these two introns. These were likely the result of cryptic splice sites within this region. Many isoforms, however, gave no ORF and were likely to be the products of splicing artifacts.

Three transcription isoforms, XM-085409, *T17*, and BI836550 (**Figure 5.10**) are likely to be functional based on the present of ORFs by utilizing the same ATG start, in exon 3, as that of full-length *T13*. These transcripts could code for proteins that are in common by their amino terminus with that of *T13*, though varying in length. The predicted protein sequence of 124 amino acids of XM-085409 shares only 75 amino acids with *T13* protein and excludes the ankyrin domain. The protein products of *T17* and BI836550 on the other hand contain the N-terminal part of *T13* protein that includes the ankyrin repeat domains. The latter also shares part of the carboxy terminal region with that of full-length *T13*. Together, these protein products may likely function in association with the full-length *T13* protein.

Failure to detect any sequence mutation of *T13* either in 48 breast cancer samples showing LOH of chromosome 16 at q24.3 region or 24 breast cancer cell lines might indicate that *T13* was not likely a gene involved in breast cancer pathogenesis.

Alternatively, T13 mutations may be involve in pathogenesis of other cancers because LOH at this chromosome region have also been reported in cancer of other organs such as hepatocellular carcinoma (Tsuda *et al.*, 1990), prostatic carcinoma (Carter *et al.*, 1990; Bergerheim *et al.*, 1991; Kunimi *et al.*, 1991), and ovarian cancer (Sato *et al.*, 1991b). Mechanism other than sequence changes exist that inactivates gene function. Hyper- or aberrant methylation of the CpG island at the promotor region of the gene has also been reported to reduce or suppress gene expression in many genes that behave as tumor suppressors (Baylin *et al.*, 2001; Karpf and Jones, 2002; Plass, 2002). Such mechanism may also involve inactivation of *T13* in breast carcinogenesis. This, however, requires further investigation.

The *T13* gene product is a large molecule although only one functional domain could be defined by *in silico* analysis. T13 protein may also possess other functional domains, yet to be identified, that are important for its activity. BRCA1, for instance, is also a large molecule that contains only two known functional domains, Ring and BRCT domains. However, BRCA1 has been reported to interact with several proteins for its biological functions (section 1.5.1.1). For the T13 protein a search for proteins that interact with the ankyrin domain and other parts of the T13 protein may provide an understanding of its biological function.

## Chapter 6

# Characterization of Transcription Unit

18 (T18)

---

## ***Table of Contents***

	<b>Page</b>
<b>6.1 Introduction</b>	180
<b>6.2 Methods</b>	180
6.2.1 T18 transcript sequence assembly	180
6.2.2 Expression Analysis	181
6.2.3 Sequence Homology and <i>In Silico</i> Protein Analysis	181
6.2.4 Characterization of the Genomic Structure	182
<b>6.3 Results</b>	182
6.3.1 Identification of T18 transcript sequence	182
6.3.2 Transcript Sequence and Expression Analysis	183
6.3.3 Protein Sequence Analysis	183
6.3.4 Sequence Homologies	184
6.3.4 T18 gene structure	185
<b>6.4 Discussion</b>	186



## **6.1 Introduction**

Transcription unit 18 (*T18*) was assigned to the cDNA clone IMAGE:1098126 which was mapped to the proximal region of the BAC clone 561E17 (section 5.1). This BAC forms part of a genomic contig that extended approximately 140 kb from the proximal boundary of the physical map established at chromosome region 16q24.3 (Whitmore *et al.*, 1998a). As part of the aim <sup>was</sup> to study genes within this extended region, in addition to *T13*, *T18* was also selected for gene characterization. The cDNA clone IMAGE:1098126 was isolated from a cDNA library derived from breast tumor tissue. The partial sequence of this cDNA clone, EST AA603773, was from its 3' end. This EST belongs to the group of partial cDNA sequences collectively refer to the UniGene cluster **Hs.123661**. The EST sequence members of this cluster are derived from cDNA libraries of various tissue sources, suggesting a ubiquitous expression of this transcript, which may indicate the essential function of its protein product.

This chapter described the isolation of the T18 transcript sequence, analysis of its sequence and pattern of expression, and subsequently characterization of the putative function of its predicted protein. Finally there is discussion of its putative protein function and whether this gene is likely to be a breast cancer associated tumor suppressor or possibly associated with other human genetic disorders.

## **6.2 Methods**

### **6.2.1 T18 transcript sequence assembly**

A combination of RT-PCR experiments and direct sequencing of the cDNA clone IMAGE:1098126 were carried out to obtain the full-length transcribed sequence of T18

The RT-PCR method was performed as described in section 2.2.13, using a primer set generated from the partial cDNA sequences, AA251134 and THC124415. The sequence of

primers and their corresponding location in the transcribed sequence of *T18* are presented in **Table 6.1**. The tissue sources of mRNA used for RT-PCR are fetal brain and fetal heart.

The cDNA clone IMAGE:1098126 was purchased, DNA isolated and sequenced by the Dye-terminator method (2.2.16.1) on an ABI 373 DNA sequencer utilizing vector primers flanking the inserted cDNA and the primers designed to the sequence generated from the RT-PCR products (**Table 6.1**).

### 6.2.2 Expression Analysis

RT-PCR products generated from fetal brain and fetal heart mRNA utilizing primer set as shown in **Table 6.1** were separated by agarose-gel electrophoresis and cDNA bands were purified from the gel (section 2.2.14.2). These purified cDNA fragments were then used to hybridize to a multiple tissue Northern membrane, commercially available from Clontech, containing 2 µg of poly(A)+ mRNA from six different human tissues. The tissue sources of mRNA are heart, brain, placenta, lung, liver, skeletal muscle, kidney, and pancreas. The hybridization condition was performed according to the manufacturer's recommendation in "ExpressHyb" hybridization solution (Clontech).

### 6.2.3 Sequence Homology and *In Silico* Protein Analysis

Nucleotide sequence homology tool, BLASTN, was used to perform homology searches against nucleotide databases: dbEST, non-redundant database, and genome sequence database. BLASTP was used to search for homology <sup>ours</sup> sequences from various protein databases. Prediction of protein function was assisted by, (i) protein domain identification tools, Pfam, SMART, and Prosite, (ii) sub-cellular localization prediction program, PSORTII, (iii) comparing the sequence with both unknown and known functional

proteins by multiple sequence alignment tools, CLUSTALW, available at GenomeNet server (Kyoto Center, Japan).

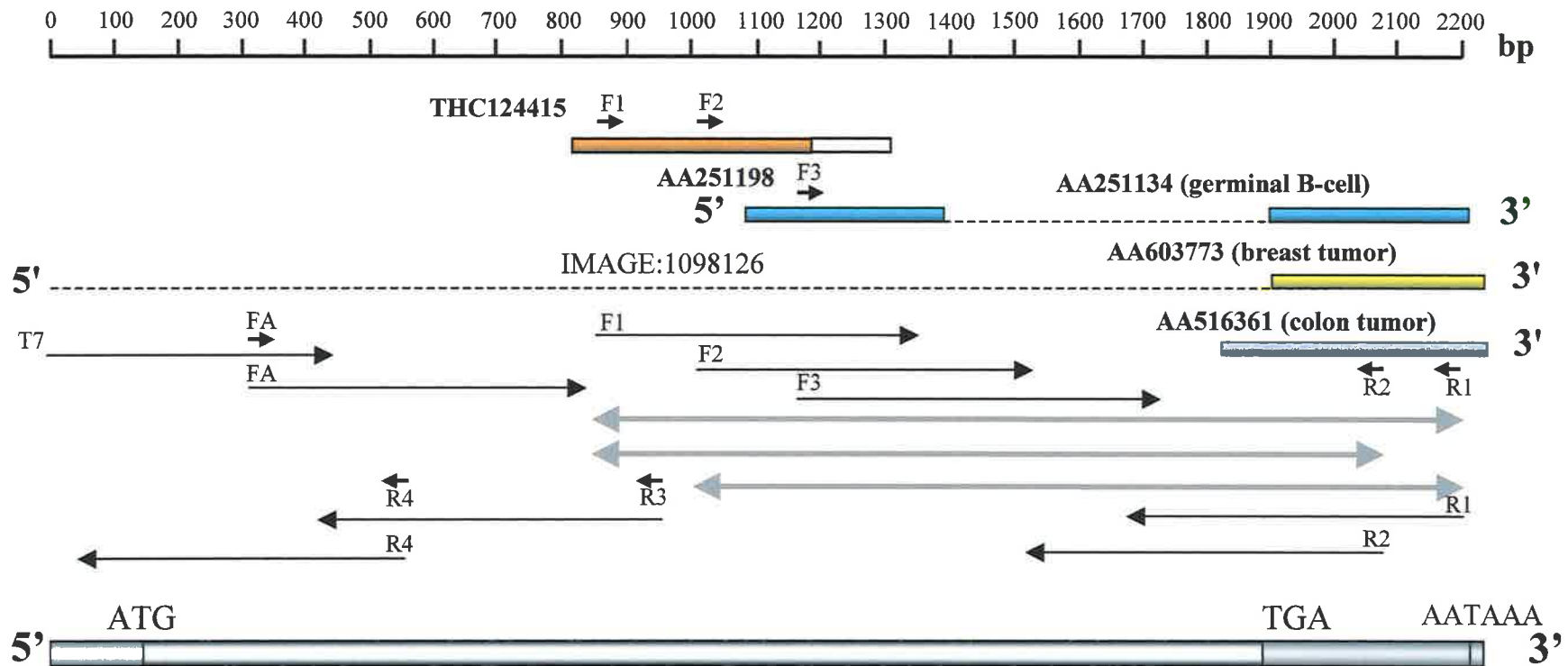
#### 6.2.4 Characterization of the Genomic Structure

The gene structure of *T18* was determined by comparing the complete cDNA sequence with genomic sequence encompassing this region, available from GenBank database. The transition of exon to intron (exon boundary) was identified from the genomic sequence that diverges from that of cDNA.

### 6.3 Results

#### 6.3.1 Identification of *T18* transcript sequence

The sequence of AA603773 was initially used to search the GenBank EST database. A number of homologous EST sequences, all from the 3' end of cDNA, were identified and most contained the consensus polyadenylation signal, AATAAA. Two of these ESTs, AA251134 and AA516361, were assembled to form a sequence that represented the 3' end of *T18*. The sequence of EST AA251134 is derived from the 3' end of the cDNA clone IMAGE:684139 of unknown insert size. However, the 5' sequence of this cDNA was also available as the database entry AA251198 (**Figure 6.1**). The sequence of AA251198 was then used to perform homology searches, and an assembled cDNA sequence, THC124415, was identified from the TIGR database. THC124415 extended the sequence 5' from AA251198 by a further 290 bp. To determine the entire transcribed sequence, primers were designed to the cDNA sequences of AA251134 and THC124415 for RT-PCR experiments (**Figure 6.1** and **Table 6.1**). Three RT-PCR products were obtained (**Figure 6.2**) and DNA sequence determined sequentially from both strands.

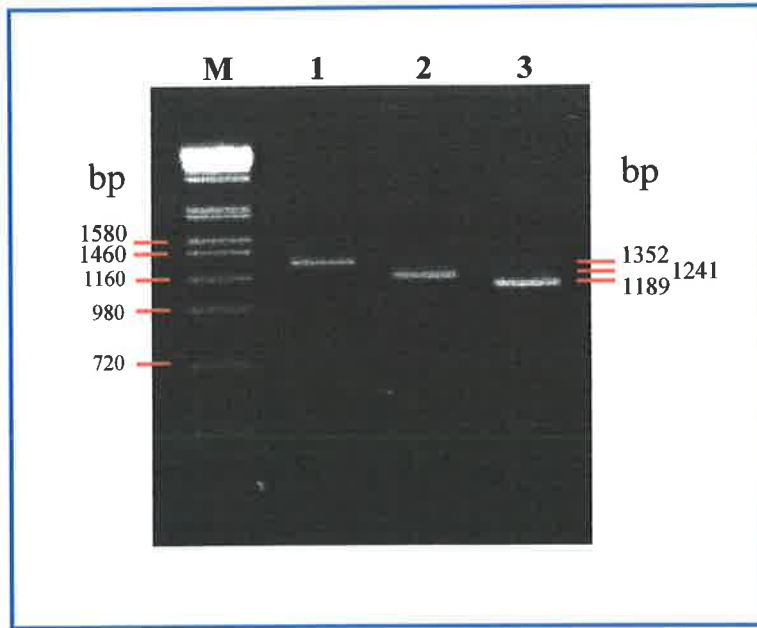


**Figure 6.1** : Determination of *T18* transcript sequence. Three ESTs (AA251134, AA603773, AA516361) that belong to a UniGene cluster Hs.123661 were chosen and performed sequence alignment before the primers, R1 and R2, were designed. AA251198 is the sequence from the 5' end of the same cDNA clone of AA251134. THC124415 is the assembled cDNA sequence from TIGR database that overlaps with AA251198. Primers F1 and F2 were designed from THC124415 sequence. Clone IMAGE: 1098126 is the cDNA clone origin of AA603773 that was isolated for sequence determination. Thick lines with arrow heads at both ends denote the RT-PCR products using three combination of primers: F1/R1, F1/R2, and F2/R1. Lines with single arrow head indicate the extend of sequencing of RT-PCR products (primers R1, R2, F1, F2, F3) and of cDNA clone IMAGE:1098126 (primers T7, FA, R3, and R4). The 2.26 kb transcript sequence of T18 was assembled from these sequencing products and revealed an ORF, completed with start, ATG, and stop, TGA, codons and a polyadenylation signal, AATAAAA.

**Table 6.1**

**Sequences and Nucleotide positions of the primers used for the Identification of *T18* transcript sequence**

<b><i>No</i></b>	<b><i>Primer</i></b>	<b><i>Nucleotide sequence (5'→3')</i></b>	<b><i>Neucleotide Position in cDNA</i></b>
1	T18-FA	aga atc gcc ctg gtt gac c	313-331
2	T18-F1	gac caa aga cga cgt gat cc	843-862
3	T18-F2	atc aat gtc ttt atg gca gtg c	1006-1027
4	T18-F3	cag tgc tgg aga agt gga ag	1160-1179
5	T18-R1	gca act gca ggg cac tgt g	2194-2176
6	T18-R2	gga cgg tct cct tga gct gc	2083-2065
7	T18-R3	cat cat cac aca ggt ggc tc	948-929
8	T18-R4	cgc aga tga cat act cca gc	565-546
9	T7-primer	taa tac gac tca cta tag gg	Vector primer



---

**Figure 6.2** : Agarose gel electrophoresis of T18 RT-PCR products. The cDNA was amplified from the total RNA isolated from fetal heart. The products contain in gel, lane 1, 1352 bp; lane 2, 1241 bp; and lane 3, 1189 bp were obtained using primer pairs T18-F1/-R1, T18-F1/-R2, and T18-F2/-R1 respectively. The DNA marker is loaded in lane M and size-fragments are indicated in bp.

---

The cDNA clone IMAGE:1098126 was also sequenced utilizing the vector primers flanking the cDNA insert and the primers designed to the sequence from the RT-PCR products mentioned above (**Figure 6.1**). The cDNA sequences from RT-PCR products and those from sequencing of cDNA clone IMAGE:1098126 were finally assembled to form a full-length cDNA of 2,263 bp.

### 6.3.2 Transcript Sequence and Expression Analysis

Analyses of the *T18*, 2,263-bp transcript sequence (**Figure 6.3**) identified an ORF of 1728 bp with the start codon, ATG, and stop codon, TGA, at nucleotide positions 149-151 and 1877-1879 respectively. The ORF was followed by a 384-bp 3' UTR. The sequence surrounding an ATG start (AGTGCAATGC), though, weakly conformed to the Kozak consensus of an initiation start site (GCCA/GCCATGG) (Kozak, 1989). It was, however, preceded by an in-frame stop codon, TAG, 39 bp upstream from the initiation start site. The ORF was predicted to code for a protein of 576 amino acids. A sequence characteristic of polyadenylation signal, AATAAA, was identified at the 3' terminal region and situated 19 bp upstream from a polyadenylation site.

Northern hybridization using RT-PCR products of *T18* (section 6.2.2) as probes identified a prominent mRNA band of approximately 2.5 kb in all tissues tested, though with varying degrees of expression (**Figure 6.4**). This mRNA band corresponded with the size of the identified transcript sequence of *T18*. With a longer exposure, faint bands of 1.3 and 4.4 kb were also detected. Expression of *T18* as determined by Northern analysis was equally high in liver, kidney, and pancreas. Medium degree of expression was seen in heart and skeletal muscle, while expression in brain, placenta, and lung was fairly low.

### 6.3.3 Protein Sequence Analysis

---

**Figure 6.3** : The cDNA transcript and deduced amino acid sequence of *T18*. The transcript is 2262 bp from the first nucleotide to the polyadenylation site (A). The start, ATG, and stop, TGA, codons are located at positions 148-150 and positions 1876-1878 respectively, and are highlighted. Also indicated, an inframe stop codone, TAG, that preceded the translation start site. The sequence of polyadenylation signal, AATAAA, at positions 2238-2243 is underlined. Within the amino acid sequence, the sequence characteristic of “AMP-binding domain” is indicated at residues 198-211 (underlined) and the consensus residues are highlighted.

---



gcaagaggccccggccggaacccggccccgacccccggcgcgcgcgcgaggagacgaggaagagttgtggcgaggcagatc **80**  
 ctgccccgtggccgcgccgtctcgt**tag**ctcgccccaggaggctccccgggagcgctgtcagtgca**ATGCCGCCCATG** **160**  
M P P H **4**

TGGTGCTCACCTTCCGGCGCCTGGGCTGCGCCTTGCCCTCCTGCCGCTGGCGCCTGCGAGACACAGAGGAAGTGGTCTT **240**  
 V V L T F R R L G C A L P S C R L A P A R H R G S G L **31**

CTGCACACAGCCCCAGTGGCCCGCTCGGACAGGAGCGCCCCGGTGTTCACCCGTGCCCTGGCCTTTGGGGACAGAATCGC **320**  
 L H T A P V A R S D R S A P V F T R A L A F G D R I A **58**

CCTGGTTGACCAGCACGGCCGCCACAGTACAGGGAGCTTTATTCCCGCAGCCTTCGCCTGTCCCAGGAGATCTGCAGGC **400**  
 L V D Q H G R H T Y R E L Y S R S L R L S Q E I C R **84**

TCTGCGGGTGTGTCGGCGGGGACCTCCGGGAGGAGAGGGTCTCCTTCTATGTGCTAACGACGCCTCCTACGTCGTGGCC **480**  
 L C G C V G G D L R E E R V S F L C A N D A S Y V V A **111**

CAGTGGGCCTCATGGATGAGCGGGTGTGGCAGTACCCCTCTACAGGAAGCATCCCGCGGCCAGCTGGAGTATGTCAT **560**  
 Q W A S W M S G G V A V P L Y R K H P A A Q L E Y V I **138**

CTGCGACTCCCAGAGCTCTGTGGTCTTGCCAGCCAGGAGTACCTGGAGCTCCTGAGCCCGGTGGTCAGGAAGCTGGGGG **640**  
 C D S Q S S V V L A S Q E Y L E L L S P V V R K L G **164**

TCCCCTGCTGCCGCTCACACCAGCCATCTACACTGGAGCAGTAGAGGAACCGGCAGAGGTCCCCTGCCAGAGCAGGGA **720**  
 V P L L P L T P A I Y T G A V E E P A E V P V P E Q G **191**

TGGAGGAACAAGGGCGCCATGATCATCTACACCAGTGGGACCACGGGGAGGCCCAAGGGCGTGTGAGCACGCACCAAAA **800**  
 W R N K G A M I I Y T S G T T G R P K G V L S T H Q N **218**

CATCAGGGCTGTGGTGACCGGGCTGGTCCACAAGTGGGCATGGACCAAAGACGACGTGATCCTTCACGTGCTCCCCTGC **880**  
 I R A V V T G L V H K W A W T K D D V I L H V L P L **244**

ACCACGTCCATGGTGTGGTCAACGCGCTGCTCTGTCTCTTGGGTGGGAGCCACCTGTGTGATGATGCCTGAGTTCAGC **960**  
 H H V H G V V N A L L C P L W V G A T C V M M P E F S **271**

CCTCAGCAGGTTTGGGAAAAGTTCTTAAGTCTGAAACCGCCGGATCAATGTCTTTATGGCAGTGCCTACAATATACAC **1040**  
 P Q Q V W E K F L S S E T P R I N V F M A V P T I Y T **298**

CAAGCTGATGGAGTACTACGACAGGCATTTTACCCAGCCGACGCCCAGGATTTCTTGCCTGCAGTTTGTGAAGAAAAA **1120**  
 K L M E Y Y D R H F T Q P H A Q D F L R A V C E E K **324**

TTAGGCTGATGGTCTCAGGCTCAGCTGCCCTGCCCTCCCAGTGTGGAGAAGTGGAGAACATCACGGGCCACACCTG **1200**  
 I R L M V S G S A A L P L P V L E K W K N I T G H T L **351**

CTGGAGCGGTATGGCATGACCAGATCGGCATGGCTCTGTCCGGGCCCTGACCACTGCCGTGCGCCTGCCAGGTTCCGT **1280**  
 L E R Y G M T E I G M A L S G P L T T A V R L P G S V **378**

GGGGACCCACTGCCTGGAGTACAGGTGCGCATTGTCTCAGAAAACCCACAGAGGGAAGCCTGCTCCTACACCATCCAG **1360**  
 G T P L P G V Q V R I V S E N P Q R E A C S Y T I H **404**

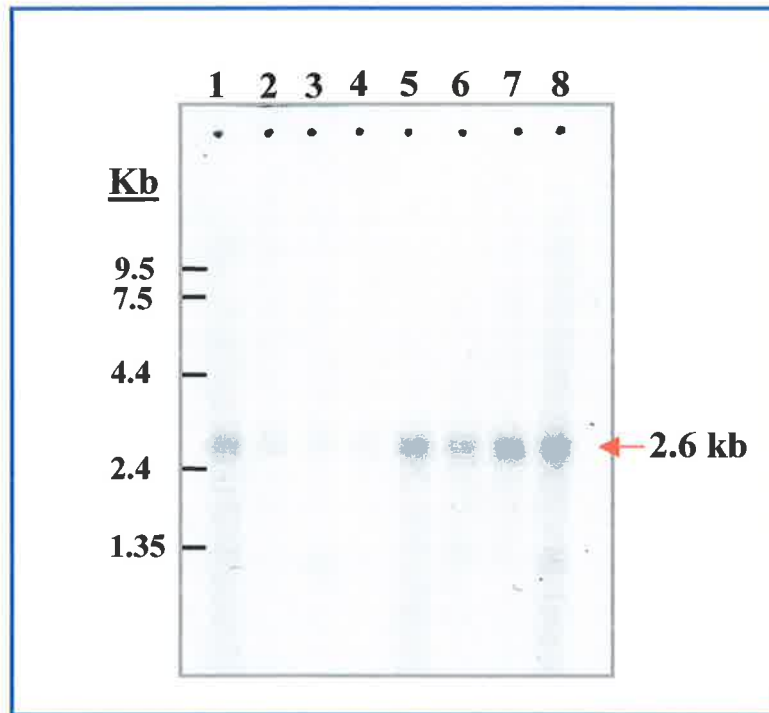
CAGAGGGAGACGAGAGGGGGACCAAGGTGACCCAGGGTTTGAAGAAAAGGAGGGGGAGCTGCTGGTGAGGGGACCCTCC **1440**  
 A E G D E R G T K V T P G F E E K E G E L L V R G P S **431**

GTGTTTCGAGAACTACTGGAATAAACAGAAAGAACTAAGAGTGCATTACCCCTGGATGGCTGGTTTAAAGACAGGGGACAC **1520**  
 V F R E Y W N K P E E T K S A F T L D G W F K T G D T **458**

CGTGGTGTTTAAGGATGGCCAGTACTGGATCCGAGGCCGGACCTCAGTGGACATCATCAAGACTGGAGGCTACAAGGTCA **1600**  
 V V F K D G Q Y W I R G R T S V D I I K T G G Y K V **484**

GCGCCCTGGAGGTGGAGTGGCACCTGCTGGCCACCCAGCATCACAGATGTGGCTGTGATTGGAGTCCGGATATGACA **1680**  
 S A L E V E W H L L A H P S I T D V A V I G V P D M T **511**





**Figure 6.4** : Northern Blot Analysis of *T18*. The multiple tissue Northern membrane (Clontech) was hybridized with  $^{32}\text{P}$ -labelled cDNA probe derived from T18 RT-PCR products. A prominent band of 2.6 kb was obtained with varying degree of intensity in various tissues. The tissue sources of mRNA are 1 heart, 2 brain, 3 placenta, 4 lung, 5 liver, 6 skeletal muscle, 7 kidney, 8 pancreas.

Analysis of the deduced amino acid sequence of *T18*, utilizing three domain-prediction programs: Pfam, Prosite, and SMART, identified the region between residues 67 and 504 as a sequence domain characteristic of an AMP-binding enzyme. This suggested that *T18* protein belonged to this enzyme family. Within this region a conserved “AMP-binding domain” was also detected between amino acid residues 198-211 (**Figure 6.3**). By utilizing the PSORT program, the *T18* protein was predicted to be localized to the cytoplasmic compartment and be a membrane-bound protein. The most likely anchoring membrane was endoplasmic reticulum (ER).

#### 6.3.4 Sequence Homologies

The nucleotide and predicted amino acid sequences of *T18* were compared with the GenBank and EMBL databases utilizing BLASTN and BLASTP algorithms. By non-redundant database searches, the transcript sequence of *T18* showed a significant homology to mRNA sequences of both human and mouse. A human mRNA LOC197322 of 873 bp was completely identical to the sequence of *T18* from nucleotide 1386 to its 3' end, suggesting that mRNA LOC197322 was a partial sequence and represented the 3' half of *T18*. Sequence homology to mouse mRNA was detected to LOC234844, a 2225-bp mRNA, and to MGC:37904, an mRNA of 1096 bp. Significant homology between *T18* and mouse mRNA sequences was only within their ORF (**Figure 6.5**). The nucleotide sequence of LOC234844 was then retrieved and the protein sequence deduced, utilizing program “ORF finder” at NCBI. The predicted mouse protein of 583 amino acids was then compared with that of *T18*. Significant homology, 80% identical, was demonstrated between these two proteins (**Figure 6.6**). With the similar size of both mRNA and protein as well as the protein sequence homology suggested that the mouse LOC234844 was an orthologue of human *T18*, and at this stage was assigned as *mt18*.

---

**Figure 6.5 : Comparison of the Human and Mouse *T18* cDNA sequences.** Sequence homology is shown only between the coding region of both cDNAs. Both transcripts utilize the equivalent translation start site. Also indicated are the 5'upstream stop codons preceding the start sites on both transcripts. The termination codon in mouse transcript (**tag**) is situated downstream from the site equivalent to the position of termination codon (**tga**) of the human transcript. The exon boundaries were equivalent in human and mouse genes, and are indicated by the vertical lines with double-head horizontal arrows and the numbers of exon. The boundaries of exons 2-3 and 3-4 in human *T18* was only predicted based on the sequence homology to its mouse homologue and are indicated by dash lines. The boundary of exon 1-2 was not able to be predict because there is no significant homology between human and mouse nucleotide sequence at this region.

---

HumT18: 100 gtctcgtagctcggccccaggaggtccccgggagcgctgtcagtgcaatgcccggccat 159  
 | | | | | | | | | | 1 ←→ 2 | | | | | | | | | | | | | | |  
 MusT18: 80 cacccaaccccgcctgaccccagcccagccagtggtttgtcgggtgtgatgcccacctcac 139

HumT18: 160 gtggtgctcaccttccggcgcctgggctgcgcttgcctcctgccggctggcgccctgcg 219  
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 musT18: 140 ttggcactgcccttcaggcgctctcttctggtccttggcctccagtcagctgattcccaga 209

HumT18: 220 agacacagaggaagtggctcttctgcacacagccccagtgggcccgtcggacaggagcgcc 279  
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 musT18: 200 agacaccgaggacatagcctcctgcctaccaccccagaggccacacggatgggagtgtc 259

HumT18: 280 ccggtgttcacccgtgccctggccttggggacagaatcgccctggttgaccagcacggc 339  
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 musT18: 260 ccggtattcatccgtgccctggccttggggacaggattgccctattgacaaatatggc 319

HumT18: 340 cgccacacgtacagggagctttattcccgcagccttgcctgtcccaggagatctgcagg 399  
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 musT18: 320 catcacacctacaggaactgtatgatgcagccttgtctggcccaggagatctgcagg 379

HumT18: 400 ctctgcgggtgtgtcggcggggacctccgggaggagagggtctccttcctatgtgctaac 459  
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 musT18: 380 cttcagggttgtaaagtggggacctccaggaagaaagggtctccttcctgtgctccaat 439

HumT18: 460 gacgcctcctacgtcgtggcccagtgggcctcatggatgagcggcggtgtggcagtacc 519  
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 musT18: 440 gacgtctcctacgtttagcacagtgggcctcctggatgagtggtggtgtgtgccca 499

HumT18: 520 ctctacaggaagcatcccgcggcccagctggagtagtcatctgcgactcccagagctct 579  
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 musT18: 500 ctgtactggaagcacccctgaggcccagctggagtagtattcatccaggactcccggagctct 559

HumT18: 580 gtggtccttgccagccaggagtacctggagctcctgagccccggtggtcaggaagctgggg 639  
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 musT18: 560 ttggtggtggttggccaggagtagtggagcgggctaagtccactggctcagaggtcggga 619

HumT18: 640 gtcccgtgtgtgccgctcacaccagccatctacactggagcagtagaggaaccggcagag 699  
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 musT18: 620 gtccctcttctgccgctcagcctgccgtctaccatggagcaacagagaagcccacagag 679

HumT18: 700 gtcccgggtcccagagcagggatggaggaacaaggcgccatgatcatctacaccagtgg 759  
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 musT18: 680 cagccagttgaagagagtggtggcgtgaccgggggtgcatgatcttctacaccagcggg 739

HumT18: 760 accacggggaggccaagggcgtgctgagcacgcacaaaaacatcagggctgtggtgacc 819  
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 2 ←→ 3  
 musT18: 740 accacagggaggccaaggggtgactgagcaccaccgtaacctagccgctgtggtgact 799

HumT18: 820 gggctgggtccacaagtgggcattggacaaagacgacgtgatccttcacgtgctcccgtg 879  
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 musT18: 800 gggctgggtccactcatggcgtggactaaaaacgatgtgatccttcacgtcctcccgtg 859

HumT18: 880 caccacgtccatgggtgtggtcaacgcgctgctctgtcctctctgggtgggagccacctgt 939  
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 musT18: 860 caccatgtccacggcgtgggtcaacaagctgctctgtccactctgggtaggagccacctgt 919

**HumT18: 940** gtgatgatgcctgagttcagccctcagcaggtttgggaaaagtcttaagtctgaaacg **999**  
 || ||| |||| | |||| |||| | |||| |||| | |||| |||| | |||| |||| | |||| |||| |

**musT18: 920** gtcattgctgcccggagttcagtgctcagcaggtttgggaaaaattcttgagttctgaagcc **979**  
 || ||| |||| | |||| |||| | |||| |||| | |||| |||| | |||| |||| | |||| |||| |

**HumT18:1000** ccgcgatcaatgtctttatggcagtgccctacaatacaccaagctgatggagtactac **1059**  
 || | ||| | || || | |||| |||| | || | |||| | |||| | |||| | |||| |

**musT18: 980** ccacagattaccgtattcatggcagtgcccactgtctacagcaagctgctggactactat **1039**  
 || | ||| | || || | |||| |||| | || | |||| | |||| | |||| | |||| |

**HumT18:1060** gacaggcattttaccagccgcagcccaggatttcttgctgagtttgtgaagaaaa **1119**  
 |||| | |||| | || |||| || | |||| |||| | |||| |||| | |||| |||| |

**musT18:1040** gacaagcatttcacgcagccccatgtccaggatttctgctgagtttgtaaagaaaga **1099**  
 || |||| |||| | |||| |||| | |||| |||| | |||| |||| | |||| |||| |

**HumT18:1120** attaggctgatggctcaggtcagctgccctgccctcccagtgctggagaagtggag **1179**  
 || |||| |||| | |||| |||| | |||| |||| | |||| |||| | |||| |||| |

**musT18:1100** atcaggctgatggctcaggttcggccgctctgctgctcccactgctggagaagtggagg **1159**  
 || |||| |||| | |||| |||| | |||| |||| | |||| |||| | |||| |||| |

**HumT18:1180** aacatcacgggccacaccctgctggagcggatggcatgaccgagatcggcatggctctg **1239**  
 | | |||| |||| | |||| |||| | |||| |||| | |||| |||| | |||| |||| |

**musT18:1160** agcggccacgggccacacgctgctggagcgtatggcatgacagagatcggcatggctctg **1219**  
 || | |||| |||| | |||| |||| | |||| |||| | |||| |||| | |||| |||| |

**HumT18:1240** tccgggcccctgaccactgccgtgcgctgccaggttccgtggggaccccactgcctgga **1299**  
 ||| |||| |||| | || |||| |||| | |||| |||| | |||| |||| | |||| |||| |

**musT18:1220** tccaaccccctgacggaggct---cgcgctgccaggttctgtggggaccccattgccagga **1276**  
 ||| |||| |||| | || |||| |||| | |||| |||| | |||| |||| | |||| |||| |

**HumT18:1300** gtacaggtgctcattgtctcagaaaaccacagaggggaagcctgctcctacaccatccac **1359**  
 || | ||| |||| | |||| |||| | |||| |||| | |||| |||| | |||| |||| |

**musT18:1277** gtggaagtacgcatcatctcagaaaaccgcagaaggggtccc---cctacatcatccat **1333**  
 || | ||| |||| | |||| |||| | |||| |||| | |||| |||| | |||| |||| |

**HumT18:1360** gcagagggagacgagagggggaccaaaggtgaccccagggttgaagaaaaggagggggag **1419**  
 |||| |||| | |||| |||| | |||| |||| | |||| |||| | |||| |||| |

**musT18:1334** gcagagggaaaacgagagggggacaaaggtgactccagggttcaggaaaaggaagggag **1393**  
 |||| |||| | |||| |||| | |||| |||| | |||| |||| | |||| |||| |

**HumT18:1420** ctgctggtgaggggaccctccgtgttccgagaatactggaataaaccagaagaaactaag **1479**  
 |||| |||| | |||| |||| | |||| |||| | |||| |||| | |||| |||| |

**musT18:1394** ctgctggttaggggaccctccgtgttccgagaatactgggataagccagaagaaacgaaa **1453**  
 |||| |||| | |||| |||| | |||| |||| | |||| |||| | |||| |||| |

**HumT18:1480** agtgcattcaccctggatggctggtttaaagacaggggacaccgtggtggttaaggatggc **1539**  
 |||| |||| | |||| |||| | |||| |||| | |||| |||| | |||| |||| |

**musT18:1454** agtgcattcacttccgatggctggttcagaacagggagacaccgccgtgttcaaggatgct **1513**  
 |||| |||| | |||| |||| | |||| |||| | |||| |||| | |||| |||| |

**HumT18:1540** cagtactggatccgaggccggacctcagtgacatcatcaagactggaggctacaaggtc **1599**  
 |||| |||| | || || || |||| |||| | |||| |||| | |||| |||| | |||| |||| |

**musT18:1514** aggtactggatccgaggccgcacgtctgtggacatcatcaagactggaggctataaagtc **1573**  
 |||| |||| | || || || |||| |||| | |||| |||| | |||| |||| | |||| |||| |

**HumT18:1600** agegccctggaggtggagtggcacctgctggcccaccccagcatcacagatgtggctgtg **1659**  
 |||| |||| | || || |||| |||| | |||| |||| | |||| |||| | |||| |||| |

**musT18:1574** agegccctggagatcgagcggcacctgctggcccaccccagcatcacagatgtcgctgta **1633**  
 |||| |||| | || || |||| |||| | |||| |||| | |||| |||| | |||| |||| |

**HumT18:1660** attggagttccggatatgacatggggccagcgggtcactgctgtggtgaccctccgagaa **1719**  
 |||| |||| | |||| |||| | |||| |||| | |||| |||| | |||| |||| |

**musT18:1634** attggagttcctgatatgacatggggccagcgggtcacagctgtggtgctctcaagaa **1693**  
 |||| |||| | |||| |||| | |||| |||| | |||| |||| | |||| |||| |

**HumT18:1720** ggacactcactgtcccacagggagctcaaagagtgggccagaaatgtcctggccccgtac **1779**  
 |||| |||| | |||| |||| | |||| |||| | |||| |||| | |||| |||| |

**musT18:1694** ggacattcactgtcccacaggggatctcaaggagtgggccagaggtgtcctggccccctat **1753**  
 |||| |||| | |||| |||| | |||| |||| | |||| |||| | |||| |||| |



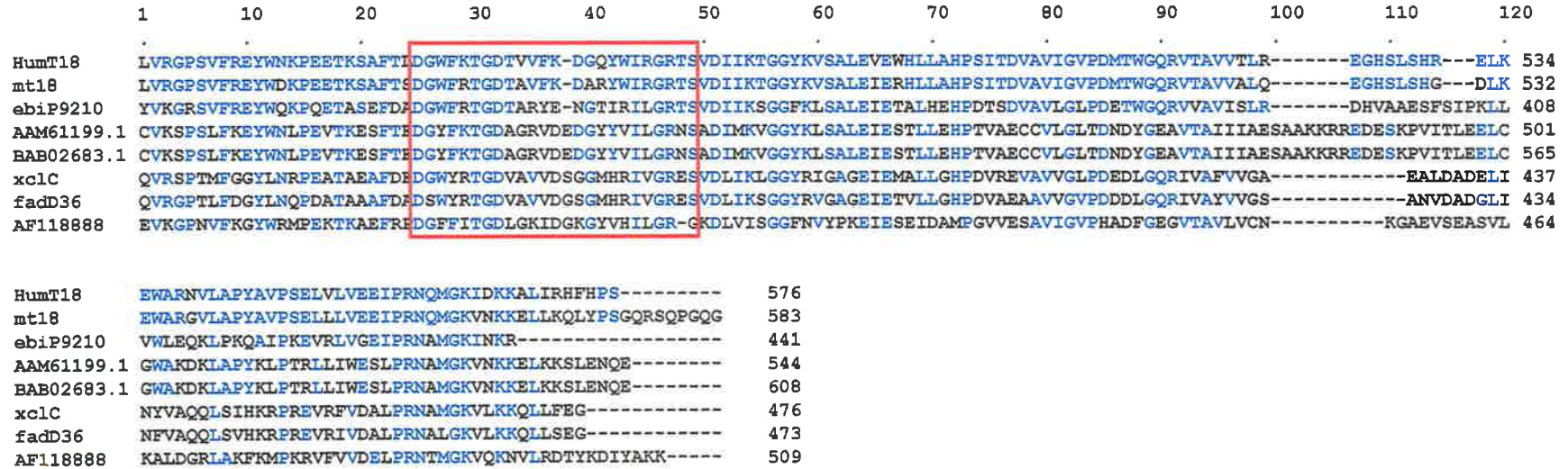


There was no significant homology of the T18 protein sequence to any known human protein in the database. Instead, sequence similarity was identified to a number of proteins from other species. These proteins also belong to the AMP-binding protein family. The representative protein sequences were ebiP9210, an incomplete protein sequence from *Anopheles gambiae*; AY084636, a putative long-chain acyl-CoA synthetase, and AB012247, a long-chain fatty acid CoA ligase-like protein from *Arabidopsis thaliana*; AF118888, malonyl CoA synthetase from *Bradyrhizobium japonicum*; U15181 (xcLC), an acyl-CoA synthase from *Mycobacterium leprae*; and Z93777 (fadD36), a substrate-CoA ligase from *Mycobacterium tuberculosis*. The overall sequence similarity was between 35-44% identity. Showing in **Figure 6.6** is multiple sequence alignment of human T18 protein with its mouse homologue and similar proteins from other species. T18 and all of these proteins shared the consensus sequence of “AMP-binding domain”. In addition to this domain, at least six other regions with high degree of sequence homology with respect to T18 were also detected between residues 236-250, 326-336, 347-359, 376-385, 427-522, and 556-570. Furthermore, the sequence motif which conforms to the “*fatty acyl-CoA synthetase (FACS) signature*” (Black *et al.*, 1997) was detected between residues 450 and 473 of T18. FACS signature motif was also seen in all of these similar proteins. Together with the across species sequence similarity and the presence of two specific sequence motifs, the AMP-binding domain and FACS signature, T18 was considered to be a new member of the “*fatty acyl-CoA synthetase*” sub-family.

#### 6.3.4 T18 gene structure

The gene structure of T18 was determined by comparing its transcript sequence with the genomic sequence available from human genomic sequence database. Since the genomic sequence encompassing T18 region was not complete (as of December, 2002), the gene

	1	10	20	30	40	50	60	70	80	90	100	110	120		
HumT18	-----MPPHVLTFRRLGCALPSCRLAPARHRGSGLLHTAPVARSDRSAPVFFRALAFGDRIALVDQHRHTYRELYSRSLRLSQEICRLCGCVGG-----												91		
mt18	-----MPPHLALPFRRLFWSLASSQLIPRRHRGHSLLPTTPEAHTDGSVPVFRALAFGDRIALIDKYGHHTYRELYDRSLCLAQEICRLQCKVGG-----												91		
ebiP9210	-----												0		
AAM61199.1	-----MEVFKAAFSEASNSCDRIAIAKADGKSYSYGQLTSALRISKFLKDDTTNGGQETK												56		
BAB02683.1	MTATTTLKSFN	YLSLINHRLNHNSYAILSSPLLPR	SHPASTSFSNSGFRFFQSNHLFSSQSGSLMEVFKAAFSEASNSCDRIAIAKADGKSYSYGQLTSALRISKFLKDDTTNGGQETK										120		
xc1C	-----MLLASLNPSVVTATDIADAVCVDDGVVLSRSGLVGAATSAERVGG-----												45		
fadD36	-----MLLASLNPAVVSAAADIADAVRIDGVLRSDDLVAATSAERVAG-----												45		
AF118888	-----MNRAANANLFSRLFDGLDDPKRLAIETHD GARISYGD LIARAGOMANVLVARG-----												53		
HumT18	-----DLREERVSFLCANDASYVVAQWASWMSGGAVVPLYRKHPAAQLEYVICDSQSSVVLASQEYELLELSPVVRKLGVPPLPLTPAIYTG-----												AVEEPAEVFPVEQGWNRKGA	197	
mt18	-----DLQEEERVSFLCNSDVSYYVAQWASWMSGGAVVPLWYKHPAAQLEYFIQDSRSSLVVVGQEYLERLSPLAQRLGVPLPLTPAVYHG-----												ATEKPTQPVESGWRDRGA	197	
ebiP9210	-----AVPLNAKYPADLLEYIYIKDSASLLLTPEFLPLAEPLAAKQKPLLVVNHDLIHTNGENLLQLNDTLVVE SALNGEFYRDANA												84		
AAM61199.1	KYEGFGSLKGARIGIVAKPSAEFVAGVLGTWFGGAVVPLALSYPEAELLHMNDSDISLLSTEDHSETMKTIAAKSGARFHLLIPVVNSTS---												ETVACNQFODDSFEAEKGFLLDDPA	173	
BAB02683.1	KYEGFGSLKGARIGIVAKPSAEFVAGVLGTWFGGAVVPLALSYPEAELLHMNDSDISLLSTEDHSETMKTIAAKSGARFHLLIPVVNSTS---												ETVACNQFODDSFEAEKGFLLDDPA	237	
xc1C	-----ARLVAVLATPTASTVLAITGCLLIAGVFPVVPADIGVVERRHMLTDSGAQAWLGPGPSADSAFDPGLPHIPVQLDARSWNRYPEP-----												SPDATA	135	
fadD36	-----AHRVAVLATPTASTVLAITGCLLIAGVFPVVPADVGVTERRHMLTDSGVQAWLGPLP---												DDPAGLPHIPVRTHARSWHRYPEP-----	SPGATA	132
AF118888	-----VKPGDRVAVQVEKSVANIVLVLATVRAGAVYLPNTAYTLNELDYFIGDAEPSLVVCDPSKAELGAPLAAKVAGVETLGPDKGKSLT-----												EADKASSAFTTVPRENDL	162	
HumT18	MIIYTSGTTGRPKSVLSTHQNIRAVVTGLVHKWAWTKDDVILHVLPLHHVHGVMNALLCPLVWGATCVMPEFSQQVWEKFLSS-----												ETPRINVFMAVPTIYTKLMEYYDRHFTQ	310	
mt18	MIFYTSGTTGRPKSALSTHRNLAAVVTGLVHVSVAWTKNDVILHVLPLHHVHGVMNALLCPLVWGATCVMPEFSQQVWEKFLSS-----												EAPQITVFMAVPTIVYSKLLDYDKHFTQ	310	
ebiP9210	LIIYTSGTTGPKSVVLSYANLDAQLRALAHAWQVSATDSVLHTLPLNHVHGTINALNPLAVGAKCVMPLPKFSSSSVWSYLLNVNMT---												TKERVNVFMGVPTMYGLLIREYDVSFGK	200	
AAM61199.1	LIVYTSGTTGPKSVVHTHNSINSQVRMLTEAWEYTSADHFLHCLPLHHVHGLFNALFAPLYARSLVEFLPKFVSVSGIWRWRRESYPVNDKTNDSITVFTGVPTMYTRLIQGYEAMDKE												293		
BAB02683.1	LIVYTSGTTGPKSVVHTHNSINSQVRMLTEAWEYTSADHFLHCLPLHHVHGLFNALFAPLYARSLVEFLPKFVSVSGIWRWRRESYPVNDKTNDSITVFTGVPTMYTRLIQGYEAMDKE												357		
xc1C	MVIYTSGTTGPKSVLLSRRALAVDLDALAQAWQWTADDVLVHGLPLFHVHGLVLGLLGSLRIGNRFVHTGKPTPTAYAQCACEAGG-----												SLYFGVPTVWSRLVADEA-----	240	
fadD36	MVVYTSGTTGPKSVQLSRRALADLDALAEAWQWTAEDVLVHGLPLFHVHGLVLGLLGSLRIGNRFVHTGKPTPAGYAQACEYAHG-----												TLFFGVPTVWSRVAADQA-----	237	
AF118888	AIIYTSGTTGRSKSAMLTHDNLASNSLSLVGYWRFDDKDVLIHALPIYHTHGLFVATNVILFSRASMIPLPKLDPDLIIKLMARA-----												TVLMGVPTFYTRLLQNAALS---	267	
HumT18	-PHAQDFLRVAVCEKIRLMVSGSAAALPLVLEKWKNI TGH TLLERYGMTEIGMALSGLTAVR----												LPGSVGTPPLGVQVRIVSENPOREACSYTIHAEGERGKVTVPGFEEKEGEL	425	
mt18	-PHVQDFVRVAVCKERIRLMVSGSAAALPVLLEKWSATGHTLLERYGMTEIGMALSGLTAVR-R----												VPGSVGTPPLGVQVRIVSENPOK-GSPYIIHAEGERGKVTVPGFEEKEGEL	423	
ebiP9210	NARMCDYVVKTHCKNKIRLMISGSAPLPGNIFARWNETIGHRLLERYGMTEIGMAISNPYVQDGEQSRKQGCVMPLPVTD-----												SSSKQOQETIAGQL	296	
AAM61199.1	MQDSS----AFAARKLRMLMSGSSALPRPVMHQWESITGHRLLERYGMTEFVMAMSNPLRGARK----												AGTVGKPLPGVEAK-----	IKEDENDANGVEI	381
BAB02683.1	MQDSS----AFAARKLRMLMSGSSALPRPVMHQWESITGHRLLERYGMTEFVMAMSNPLRGARN----												AGTVGKPLPGVEAK-----	IKEDENDANGVEI	445
xc1C	-----VARALRPARLLVSGSASLPVVFDRLAHLTGHRPIERYGSTEISLITLSTLDGERR-----												AGWVGLPLAGVQTRLVDES-----	GGFVYDGETVGR	329
fadD36	-----AAGALKPARLLVSGSAAALPVVFDKLVQLTGHREPVERYGASESLITLSTLDGERR-----												PGWVGLPLAGVQTRLVDDD-----	GGEVPHDGETVGR	326
AF118888	-----RETTTRHMLFISGSAPLLAETHREWSARTGHAVLERYGMTEINMNTSNPYDGERVP-----												GAVGFPLPGVSLRVTD-----	ETGKELPREEIGMI	354



**Figure 6.6 : T18 protein sequence similarity.** Comparison of human T18 amino acid sequence with those of mouse orthologous (mt18) protein and similar proteins from other species: ebiP9210, an incomplete protein sequence from *Anopheles gambiae*; AY084636, a putative long-chain acyl-CoA synthetase, and AB012247, a long-chain fatty acid CoA ligase-like protein from *Arabidopsis thaliana*; AF118888, malonyl CoA synthetase from *Bradyrhizobium japonicum*; U15181 (xcLC), an acyl-CoA synthase from *Mycobacterium leprae*; and Z93777 (fadD36), a substrate-CoA ligase from *Mycobacterium tuberculosis*. Multiple sequence alignment was performed using CLUSTAL-W program. Identical amino acids, with respect to T18, are highlighted in blue. Gaps are introduced to optimize alignment and are indicated by broken line. Numbers on the right hand side of the sequences indicate the amino acid positions. Two functional domains, the “AMP-binding domain”(consensus: *IxYTSGTTGRPK*) and the 25-residue “fatty acyl-CoA synthetase (FACS) signature” motif (consensus: *DGWLHTGDIGxWxPxGxLKIDRRKK*) are boxed.

structure of *T18* was therefore partially identified for only 7 exons from its 3' half (**Figure 6.5**). To explore a putative complete structure of *T18* gene, the mouse homologue, *mt18* cDNA sequence was then aligned with mouse genomic sequence (WGS supercontig Mm8\_WIFeb01\_194) from the mouse genome database at NCBI. The mouse gene, *mt18*, is organized into 10 exons encompassing the genomic sequence of 42 kb. The exons are ranging in size from 103 to 690 bp. Exon 2 is the largest with 690 bp in size. Both start codon, ATG, and AMP-binding domain are encoded by exon 2. The FACS signature is coded for by the sequence within exons 7 and 8. When compared the partial *T18* gene structure with that of its mouse homologue, *mt18*, they appeared to utilize the equivalent exon boundaries for all available exons (**Figure 6.5**). This may suggest that *T18* would also separate into 10 exons similar to *mt18*.

#### **6.4 Discussion**

The purpose of this part of the study was to isolate and characterize a gene designated transcription unit 18 (*T18*) which was mapped to the proximal region of BAC 561E17. This BAC form part of the extended contig of the genomic clones of chromosome 16 at the region 16q24.3. As a result, this gene was predicted to code for a protein of 576 amino acids, which by *in silico* analysis was an ER-bound protein containing a sequence signature characteristic of an AMP-binding enzyme. *T18* showed a high degree of sequence homology of both the ORF and the protein sequence to its mouse homologue, which was identified as *mt18*. In addition, the *T18* protein also showed sequence similarity to proteins from invertebrate species. These homologies suggested a significant conservation of this gene, both functional and evolutionary.

The "AMP-binding motif" is shared by a number of proteins in both prokaryote and eukaryote. These proteins include luciferase (EC 1.13.12.7) in insects, alpha-aminoadipate

reductase (EC 1.2.1.31) from yeast, acetyl-CoA synthetase (EC 6.2.1.1), and long-chain fatty acid CoA ligase (EC 6.2.1.3). These proteins all exert their activity by ATP-dependent binding of AMP to their substrate. T18 also belongs to this AMP-binding protein superfamily. In addition, T18 contained a “FACS signature” motif that is specific for members of the fatty acyl-CoA synthetase sub-family. This FACS motif is found only in fatty acyl-CoA synthetases with specificity for long- and medium-chain fatty acids but not for very long-chain fatty acids. The FACS motif contributes to the enzyme function as an essential part of the substrate-binding pocket of the fatty acyl-CoA synthetase and also modulates fatty acid substrate specificity (Black *et al.*, 1997, 2000)

Fatty acyl-CoA synthetases (fatty acid:CoA ligase, AMP-forming; EC 6.2.1.3) are a multigene family. There are at least 6 human fatty acyl-CoA synthetase (*FACL*) genes and their chromosomal locations have been characterized, namely *FACL1* at 3q13 (Stanczak *et al.*, 1992), *FACL2* at 4q34-q35 (Minoshima *et al.*, 1991; Cantu *et al.*, 1995), *FACL3* at 2q34-q35 (Minekura *et al.*, 1997), *FACL4* at Xq23 (Piccini *et al.*, 1998; Cao *et al.*, 1998), *FACL5* at 10q25.1-q25.2 (Yamashita *et al.*, 2000), and *FACL6* at 5q31 (Yagasaki *et al.*, 1999). Alignment of the amino acid sequence of T18 with the sequence of these six *FACL* proteins indicated that they all share two common consensus domains, the AMP-binding motif and FACS signature (**Figure 6.7**). The overall protein sequence homology outside of these two domains is low. Therefore, *T18* is likely to be a novel member of the human *FACL* gene family that is located at chromosome region 16q24.3. It is likely that each member of this gene family has different tissue specificity as there is variation in their degree of expression and tissue distribution.

Fatty acyl-CoA synthetases play a central role in intermediary metabolism by catalyzing the formation of fatty acyl-CoA from fatty acid, ATP, and coenzyme A (CoA). Fatty acyl-CoA represent bioactive compounds that are involved in a wide variety of cellular

AMP-binding

FACS signature

```
T18:  MIIYTSGETTGRPKG-----LDGWFKTGDIV-VFKDQYWIRGRTS
FACL1: VICFTSGTTGNPKG-----KDGWLHTGDIGKWLPNGLTKIIDRKK
FACL2: VICFTSGTTGNPKG-----KDGWLHTGDIGKWLPNGLTKIIDRKK
FACL3: VIMYTSGSTGLPKG-----GQRWLCTGDIGEFEPDGCLKIIDRKK
FACL4: IVMYTSGSTGRPKG-----GQRWFCTGDIGEFHPDGCLQIIDRKK
FACL5: VICFTSGTTGDPKG-----SDGWLHTGDIGRWLPNGLTKIIDRKK
FACL6: IVCFTSGTTGNPKG-----SDGWLHTGDIGKWLPAGLTKIIDRKK
```

---

**Figure 6.7** : Sequence alignment of the *AMP binding domain* and the *fatty acyl CoA synthetase* (FACS) signature, comparing between those of T18 and of FACLs. The identical amino acids based on T18 sequence are highlight

processes including protein transport (Glick and Rothman, 1987; Pfanner *et al.*, 1989), enzyme activation or deactivation (Fulceri *et al.*, 1995; Lai *et al.*, 1993; Yamashita *et al.*, 1995), protein modification (Gordon *et al.*, 1991; McLaughlin and Aderem, 1995; Nimchuk *et al.*, 2000), cell signaling (Korchak *et al.*, 1994; Shrago *et al.*, 1995), and transcriptional control (DiRusso *et al.*, 1999; Hertz *et al.*, 1998; Raman *et al.*, 1997) in addition to serving as substrates for  $\beta$ -oxidation and phospholipid biosynthesis. Given their multiple roles, it is clear that fatty acyl-CoA synthetases occupy a pivotal role in cellular homeostasis.

Although the map location of the *T18* gene is within the region showing LOH in breast cancer, by its putative protein function this gene was considered unlikely to be a tumor suppressor. Therefore it has not been screened for breast cancer specific mutations. By analogy with the other characterized members of the *FACL* gene family, T18 is suggested to be part of the essential cellular machinery that serves for cell survival and is unlikely to be involved in growth suppression.

Whether *T18* is associated with any human disease is not known. However, other members of the fatty acyl-CoA synthetase family have been reported to be associated with a number of diseases. Deletion of the X chromosome region q22.3-q23 has been found in Alport syndrome, elliptocytosis, and mental retardation (Piccini *et al.*, 1998). This deletion involves at least four genes, one of which is the gene for fatty acyl-CoA synthetase 4 (*FACL4*). A subsequent study has indicated that mutations of *FACL4* is involved in nonspecific X-linked mental retardation (Meloni *et al.*, 2002). Up-regulation of *FACL4* has been reported in colon adenocarcinoma and was found synergistically with cyclooxygenase-2 (COX-2) to inhibit apoptosis (Cao *et al.*, 2001). A recurrent t(5;12)(q31;p13) translocation has been found in three types of hematologic disorders: refractory anemia with excess blasts and basophilia, acute myelogenous leukemia (AML) with eosinophilia, and acute eosinophilic leukemia (AEL) (Yagasaki *et al.*, 1999). This translocation involves the fusion

of the *FACL2* gene with *TEL/ETV6*. *FACL5* has been reported to show increased expression, compared with normal brain, in primary brain tumors, glioma of grade IV malignancy (Yamashita *et al.*, 2000). Such associations of the *FACL* gene family with human disorders could suggest that *T18* may also have a role, however at this stage there is no known association with any human disease. Since *T18* was highly expressed in liver, kidney, and pancreas, this gene may have critical function in these tissues. This, however, requires further investigation.



# **Chapter 7**

## **General Discussion**

---

## General Discussion

Carcinogenesis is a multistage process in both morphological changes and genetic alterations. Although several causative agents namely physical, chemical, and biological, are known to <sup>be</sup> involve<sup>d</sup> in the initiation and progression of the tumor, the common underlying lesion <sup>is</sup> a non-lethal DNA damage. Subsequent DNA replication and cell division lead to the stage of genetic instability, which is widely accepted to be the cause of many genetic alterations seen in cancers. One such alteration is the loss of heterozygosity (LOH), which is considered to be a somatic mechanism associated with the loss of tumor suppressor function (discussed in section 1.4.2).

The ultimate goal for the study of genetic diseases including cancer is to identify the genes, mutations of which associate with disease phenotypes. The common path for the study of <sup>a</sup> disease-associated gene is, firstly, by mapping of the disease-associated locus. The following step is the identification of a candidate gene that exhibits disease-causing mutation. Finally, confirmation of such candidate gene is likely by studying its disease-causing <sup>phenotype</sup> in the animal model.

Most of the works described in this thesis were initially undertaken before the first draft of the human genome sequence was available. Therefore, identification of the gene was based on the positional cloning approach utilizing the physical and transcription map constructed at the candidate region for the breast cancer tumor suppressor (Whitmore *et al.*, 1998a), <sup>at</sup> the chromosome band 16q24.3. Exon amplification technique and bioinformatics were utilized to facilitate the isolation of the transcribed sequence of the genes. Four genes were isolated and their nucleotide and protein sequences were characterized for their possible function. Only two of these characterized genes were shown to have function likely to be a candidate tumor

suppressor. The gene designated “*T3*”, by analysis of its protein sequence, was shown to have function likely to be involved in transcriptional regulation. *T13* on the other hand was predicted to code for a large protein containing sequence motifs of nuclear localization signal, suggesting the protein function in nuclear activity. Subsequent gene mutation study was utilized to determine if any of these genes was involved in breast carcinogenesis. With the current method for the detection of nucleotide changes in the coding sequence, there were no mutations identified that suggest the association of these genes in breast carcinogenesis. However, other mechanisms of gene mutation could not be ruled out.

Mutations that cause the gene sequence changes may result in either the production of aberrant protein or the reduction of mRNA expression. Mutation of a single nucleotide that change a codon usage and subsequent amino acid transition (missense mutation) could lead to the production of aberrant protein if such amino acid substitution affect the native secondary structure of that protein. The mutation that converts a functional codon into a stop signal (nonsense mutation) could lead to the premature degradation of mRNA by the mechanism called “nonsense-mediated mRNA decay” (Culbertson, 1999; Frischmeyer and Dietz, 1999). In addition mutation at the promoter region of the gene may disrupt the promoter function and lead to the aberrant expression of mRNA (Giedraitis *et al.*, 2001). Other mechanism, the promoter hypermethylation has also been reported to play a role in the down-regulation of tumor suppressor genes (Wieland *et al.*, 2000; Yuan *et al.*, 2001).

Other two genes, *T12* and *T18* that were also characterized and were excluded as a candidate tumor suppressor based on their possible function. Furthermore one of these genes, *T12*, later found to be identical to *SPG7*, a gene involves in recessive type hereditary spastic

paraplegia (HSP). Subsequent determination of the *SPG7* gene structure was undertaken and was utilized in the mutation study of a large number of HSP patients from unrelated families. This study had confirmed that frequency of *SPG7* mutations was low in a general diverse population in contrast to the previous report, studied in the specific ethnic groups (Casari *et al.*, 1998). The *T18* on the other hand was shown by its protein sequence identity and functional motifs to be a member of long chain fatty acyl CoA ligase (*FACL*) sub-family. When compared the *T18* sequence with those of other *FACL* proteins, it was suggested that *T18* is likely to be a novel member of *FACL*.

In addition to the isolation and characterization of *T3*, *T12*, *T13*, and *T18*, <sup>the candidate</sup> Candidate had also involved in the collaborative LOH study of this 16q24.3 region in breast cancer (Cleton-Jansen *et al.*, 2001) and the study of *PISSLRE* as one candidate gene (Crawford *et al.*, 1999). The detail information of these studies can be found in the manuscript of both publications <sup>region</sup> given in the appendix of the thesis. The LOH study was initially <sup>ed</sup> aim to narrow down the LOH to the smallest overlap region of 16q24.3 LOH. By using the high-density mapped polymorphic markers in this region, <sup>the</sup> LOH study was conducted in <sup>using</sup> three panels of breast cancer tissue samples and their matched peripheral blood leukocytes. During the course of this LOH study, it appeared that interpretation of the LOH data was complicated due to the large number of complex LOH, non-informativeness of the markers, and the contamination of normal tissue. Other factor such as genetic heterogeneity of the tumor could also be taken into account. In breast cancer, genetic divergence has been reported by the changes in pattern of LOH in a number chromosomal loci (Fujii *et al.*, 1996; Lichy *et al.*, 2000), with the progressive stage of tumor, <sup>ing</sup> exhibited more discordant LOH. Furthermore LOH has also been

identified in normal tissue adjacent to tumor (Deng *et al.*, 1996).

Study of the *PISSSLRE* gene was conducted since this gene was mapped to the current physical map of 16q24.3 LOH region (Whitmore *et al.*, 1998a). The possible function of *PISSSLRE* gene product as a cyclin-dependent kinase (CDK) had drawn attention that this gene might be a good candidate tumor suppressor target by LOH. CDK in association with cyclin as CDK/cyclin complex involved in various stages of the cell cycle. The presence of the *PISSSLRE* protein is necessary for the progression from G2 to M phase, and overexpression of the gene leads to growth suppression (Li *et al.*, 1995). Furthermore expression of *PISSSLRE* gene is high in terminally differentiated cells, which are withdrawn from the cell cycle (Brambilla and Draetta, 1994; Grana *et al.*, 1994). Despite this possible growth suppression function, mutation analysis in breast cancer samples was failed to detect any mutation of this gene that can be considered to be pathogenic for breast carcinogenesis.

At the initial stage of this study, the rate-limiting step was the isolation of the transcript sequence of the genes in the candidate region. With the laboratory technique available at that time isolation of the full-length transcript sequence was time consuming. However, when the large scale sequencing program of this genomic region was established, identification of the transcribed sequence of the gene was facilitated as shown in the characterization of *T3* and *T13* transcript as well as their gene structure determination.

The progression of the Human Genome Project to the point where the complete and finished human genome sequence is available could accelerate the discovering of the genes associated with human diseases including breast cancer. Indeed a number of projects has

benefited in term of gene identification and functional prediction through a number of programs associated with the human genome project. These include the accumulation of biological data in the form of computerized databases and the development of computer-assisted tools to access, archive, and analysis<sup>o</sup> this data. The present study had also used such bioinformatics to assist in the analysis of the gene<sup>e</sup> in both DNA and the predicted protein sequence. The success of the Human Genome Project could change the way of gene identification from positional cloning approach to position candidate analysis to selective candidate gene approach. The later based on a known or a putative function of unknown gene that can be derived *in silico* from the genomic information. To this point computational analysis of the gene is a critical step that can provide clues to the molecular basis of pathogenesis and invaluable insights for further experimental analysis.

Finally, although the present study was unable to identify a breast cancer tumor suppressor gene from this chromosome region, 16q24.3, the characterized genes could be the candidates for other human diseases or other types of cancer, yet to be identified. In addition to the LOH seen in breast and other cancers (mention in section 1.5.4), the presence of the genes associated with human diseases such as fanconi anemia, the *FAA* gene, and hereditary spastic paraplegia, the *SPG7*, provides an example of the involvement of this chromosome region as one hot spot for the disease associated gene await for identification and characterization.

## References

---

## References

- Aaltonen, L.A., Peltomaki, P., Leach, F.S., Sistonen, P., Pylkkanen, L., Mecklin, J.P., Jarvinen, H., Powell, S.M., Jen, J., Hamilton, S.R., et al. (1993) Clues to the pathogenesis of familial colorectal cancer. *Science* **260**: 812-816.
- Aasland, R., Gibson, T.J., and Stewart, A.F. (1995) The PHD finger: implications for chromatin-mediated transcriptional regulation. *Trends Biochem. Sci.* **20**: 56-59.
- Abbott, D.W., Thompson, M.E., Robinson-Benion, C., Tomlinson, G., Jensen, R.A., and Holt, J.T. (1999) BRCA1 expression restores radiation resistance in BRCA1-defective cancer cells through enhancement of transcription-coupled DNA repair. *J. Biol. Chem.* **274**: 18808-18812.
- Aberle, H., Bauer, A., Stappert, J., Kispert, A., and Kemler, R. (1997)  $\beta$ -catenin is a target for the ubiquitin-proteasome pathway. *EMBO J.* **16**: 3797-3804.
- Adams, S.M., Helps, N.R., Sharp, M.G.F., Brammar, W.J., Walker, R.A., and Varley, J.M. (1992) Isolation and characterization of a novel gene with differential expression in benign and malignant human breast tumors. *Hum. Mol. Genet.* **1**: 91-96.
- Albarosa, R., Colombo, B.M., Roz, L., Magnani, L., Polo, B., Cirenei, N., Giani, C., Conti, A.M.F., DiDonato, S., and Finocchiaro, G. (1996) Deletion mapping of gliomas suggests the presence of two small regions for candidate tumor-suppressor genes in a 17-cM interval on chromosome 10q. *Am. J. Hum. Genet.* **58**: 1260-1267.
- Aldaz, C.M., Chen, T., Sahin, A., Cunningham, J., and Bondy, M. (1995) Comparative allelotype of *in situ* and invasive human breast cancer: high frequency of microsatellite instability in lobular breast carcinomas. *Cancer Res.* **55**: 3976-3981.
- Alevizopoulos, K., Vlach, J., Hennecke, S., and Amati, B. (1997) Cyclin E and c-Myc promote cell proliferation in the presence of p16INK4a and of hypophosphorylated retinoblastoma family proteins. *EMBO J.* **17**: 5322-5333.
- Ali, I.U., Campbell, G., Lidereau, R., and Callahan, R. (1988) Lack of evidence for the prognostic significance of c-rebB2 amplification in human breast carcinoma. *Oncogene Res.* **3**: 139-146.
- Allred, D.C., Clark, G.M., Molina, R., Tandon, A.K., Schnitt, S.J., Gilchrist, K.W., Osborne, C.K., Tormey, D.C., and McGuire, W.L. (1992) Overexpression of *HER-2/neu* and its relationship with other prognostic factors change during the progression of *in situ* to invasive breast cancer. *Hum. Pathol.* **23**: 974-979.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**: 3389-3402.
- Amara, S.G., Jonas, V., and Rosenfeld, M.G. (1982) Alternative RNA processing in calcitonin gene expression generates mRNAs encoding different polypeptide products.



*Nature* **298**: 240-244.

Antequera, F., and Bird, A. (1993) Number of CpG islands and genes in human and mouse. *Proc. Natl. Acad. Sci. U.S.A.* **90**: 11995-11999.

Aprelikova, O.N., Fang, B.S., Meissner, E.G., Cotter, S., Campbell, M., Kuthiala, A., Bessho, M., Jensen, R.A., and Liu, E.T. (1999) BRCA-associated growth arrest is RB dependent. *Proc. Natl. Acad. Sci. U.S.A.* **96**: 11866-11871.

Arlt, H., Tauer, R., Feldmann, H., Neupert, W., and Langer, T. (1996) The YTA10-12-complex, an AAA protease with chaperone-like activity in the inner membrane of mitochondria. *Cell* **85**: 875-885.

Attwood, T.K., Croning, M.D., Flower, D.R., Lewis, A.P., Mabey, J.E., Scordis, P., Selley, J.N., Wright, W. (2000) PRINTS-S: the database formerly known as PRINTS. *Nucleic Acids Res.* **28**: 225-227.

Auch, D., and Reth, M. (1990) Exon trap cloning: using PCR to rapidly detect and clone exons from genomic DNA fragments. *Nucleic Acids Res.* **18**: 6743-6744.

Auerbach, A.D. (1993) Fanconi anemia diagnosis and the diepoxybutane (DEB) test. *Exp. Hematol.* **21**: 731-733.

Austruy, E., Cohen-Salmon, M., Antignac, C., Beroud, C., Henry, I., Nguyen, V.C., Brugieres, L., Junien, C., and Cornelisse, C. (1993) Isolation of a kidney cDNA down-expressed in Wilms' tumor, by a subtractive hybridization approach. *Cancer Res.* **53**: 2888-2894.

Baens, M., Aerssens, J., van Zand, K., van den Berghe, H., and Marynen, P. (1995) Isolation and regional assignment of human chromosome 12p cDNA. *Genomics* **29**: 44-52.

Balaban, G., Gilbert, F., Nichols, W., Meadows, A.T., and Shields, J. (1982) Abnormalities of chromosome 13 in retinoblastoma from individuals with normal constitutional karyotypes. *Cancer Genet. Cytogenet.* **6**: 213-221.

Ballabio, A. (1993) The rise and fall of positional cloning? *Nature Genet.* **3**: 277-279.

Banfi, S., Bassi, M.T., Andolfi, G., Marchitello, A., Zanotta, S., Ballabio, A., Casari, G., and Franco, B. (1999) Identification and Characterization of *AFG3L2*, a Novel Paraplegin-Related Gene. *Genomics* **59**: 51-58.

Banin, S., Moyal, L., Shieh, S., Taya, Y., Anderson, C.W., Chessa, L., et al. (1998) Enhanced phosphorylation of p53 by ATM in response to DNA damage. *Science* **281**: 1674-1677.

Bartkova, J., Lukas, J., Muller, H., Lutzht, D., Strauss, M., and Bartek, J. (1994) Cyclin D1 protein expression and function in human breast cancer. *Intl. J. Cancer* **57**: 353-361.

Baskaran, R., Wood, L.D., Whitaker, L.L., Canman, C.E., Morgan, S.E., Xu, Y., et al. (1997) Ataxia telangiectasia mutant protein activates c-Abl tyrosine kinase in response to ionizing radiation. *Nature* **387**: 516-519.

- Bateman, A., Birney, E., Durbin, R., Eddy, S.R., Howe, K.L., and Sonnhammer, E.L. (2000) The Pfam protein families database. *Nucleic Acids Res.* **28**: 263-266.
- Baumann, P., Benson, F.E., and West, S.C. (1997) Human RAD51 protein promotes ATP-dependent homologous pairing and strand transfer reactions in vitro. *Cell* **87**: 757-766.
- Baylin, S.B., Esteller, M., Rountree, M.R., Bachman, K.E., Schuebel, K., and Herman, J.G. (2001) Aberrant patterns of DNA methylation, chromatin formation and gene expression in cancer. *Hum. Mol. Genet.* **10**: 687-692.
- Bednarek, A.K., Laflin, K.J., Daniel, R.L., Liao, Q., Hawkins, K.A., and Aldaz, C.M. (2000) WWOX, a novel WW domain-containing protein mapping to human chromosome 16q23.3-24.1, a region frequently affected in breast cancer. *Cancer Res.* **60**: 2140-2145.
- Behrens, J., Jerchow, B.A., Wurtele, M., Grimm, J., Asbrand, C., Wirtz, R., Kuhl, M., Wedlich, D., and Birchmeier, W. (1998) Functional interaction of an axin homolog, conductin, with  $\beta$ -catenin, APC, and GSK3 $\beta$ . *Science* **280**: 596-599.
- Bellefroid, E.J., Marine, J.C., Matera, A.G., Bourguignon, C., Desai, T., Healy, K.C., Bray-Ward, P., Martial, J.A., Ihle, J.N., and Ward, D.C. (1995) Emergence of the ZNF91 Kruppel-associated box-containing zinc finger gene family in the last common ancestor of anthropoidea. *Proc. Natl. Acad. Sci. U.S.A.* **92**: 10757-10761.
- Berns, E.M.J.J., Klijn, J.G.M., van Putten, W.L.J., van Staveren, I.L., Portengen, H., and Foekens, J.A. (1992) *c-Myc* amplification is a better prognostic factor than *HER2/neu* amplification in primary breast cancer. *Cancer Res.* **52**: 1107-1113.
- Bertwistle, D., Swift, S., Marston, N.J., Jackson, L.E., Crossland, S., Crompton, M.R., Marshall, C.J., and Ashworth, A. (1997) Nuclear location and cell cycle regulation of the BRCA2 protein. *Cancer Res.* **57**: 5485-5488.
- Berx, G., Cleton-Jansen, A-M., Nollet, F., de Leeuw, W.J., van de Vijver, M., Cornelisse, C., and van Roy, F. (1995a) E-cadherin is a tumor/invasion suppressor gene mutated in human lobular breast cancers. *EMBO J.* **14**: 6107-6115.
- Berx, G., Cleton-Jansen, A-M., Strumane, K., de Leeuw, W.J., Nollet, F., van Roy, F., and Cornelisse, C. (1996) E-cadherin is inactivated in a majority of invasive human lobular breast cancers by truncation mutation throughout its extracellular domain. *Oncogene* **13**: 1919-1925.
- Berx, G., Staes, K., van Hengel, J., Molemans, F., Bussemakers, M.J., van Bokhoven, A., and van Roy, F. (1995b) Cloning and characterization of the human invasion suppressor gene E-cadherin (CDH1). *Genomics* **26**: 281-289.
- Bhattacharyya, N.P., Skandalis, A., Ganesh, A., Groden, J., and Meuth, M. (1994) Mutator phenotypes in human colorectal carcinoma cell lines. *Proc. Natl. Acad. Sci. U.S.A.* **91**: 6319-6323.
- Black, P.N., DiRusso, C.C., Sherin, D., MacColl, R., Knudsen, J., and Weimar, J.D. (2000)

- Affinity labeling fatty acyl-CoA synthetase with 9-p-azidophenoxy nonanoic acid and the identification of the fatty acid-binding site. *J. Biol. Chem.* **275**: 38547-38553.
- Black, P.N., Zhang, Q., Weimar, J.D., and DiRusso, C.C. (1997) Mutational analysis of a fatty acyl-Coenzyme A synthetase signature motif identifies seven amino acid residues that modulate fatty acid substrate specificity. *J. Biol. Chem.* **272**: 4896-4903.
- Bleck, O., McGrath, J.A., and South, A.P. (2001) Searching for candidate genes in the new millennium. *Clin. Exp. Dermatol.* **26**: 279-283.
- Bird, A.P. (1986) CpG-rich islands and the function of DNA methylation. *Nature* **321**: 209-213.
- Bird, A.P. (1987) CpG islands as gene markers in the vertebrate nucleus. *Trends Genet.* **3**: 342-347.
- Bochar, D.A., Savard, J., Wang, W., Lafleur, D.W., Moore, P., Côté, J., and Shiekhattar, R. (2000) A family of chromatin remodeling factors related to Williams syndrome transcription factor. *Proc. Natl. Acad. Sci. U.S.A.* **97**: 1038-1043.
- Bodmer, W., Bailey, C., Bodmer, J., Bussey, H., Ellis, A., Gorman, P., Lucibello, F., Murday, V., Rider, S., and Scambler, P. (1987) Localization of the gene for familial adenomatous polyposis on chromosome 5. *Nature* **328**: 614-616.
- Bonnefoy-Berard, N., Munshi, A., Yron, I., Wu, S., Collins, T.L., Deckert, M., Shalom-Barak, T., Giampa, L., Herbert, E., Hernandez, J., Meller, N., Couture, C., and Altman, A. (1996) Vav: function and regulation in hematopoietic cell signaling. *Stem Cells* **14**: 250-268.
- Borg, A., Baldetorp, B., Ferno, M., Killander, D., Olsson, H., and Sigurdsson, M. (1991) *ErbB2* amplification in breast cancer with rate of proliferation. *Oncogene* **6**: 137-143.
- Borg, A., Baldetorp, B., Ferno, M., Olsson, H., and Sigurdsson, H. (1992) c-Myc amplification is an independent prognostic factor in postmenopausal breast cancer. *Int. J. Cancer* **51**: 687-691.
- Bose, S., Wang, S.I., Terry, M.B., Hibshoosh, H., and Parsons, R. (1998) Allelic loss of chromosome 10q23 is associated with tumor progression in breast carcinomas. *Oncogene* **17**: 123-127.
- Bouchard, L., Lamarre, L., Tremblay, P.J., and Jolicoeur, P. (1989) Stochastic appearance of mammary tumors in transgenic mice carrying the *MMTV/c-NEU* oncogene. *Cell* **57**: 931-936.
- Brambilla, R., and Draetta, G. (1994) Molecular cloning of PISSLRE, a novel putative member of the cdk family of protein serine/threonine kinases. *Oncogene* **9**: 3037-3041.
- Brenner, A.J., and Aldaz, C.M. (1995) Chromosome 9p allelic loss and p16/CDKN2 in breast cancer and evidence of p16 inactivation in immortal breast epithelial cells. *Cancer Res.* **55**: 2892-2895.

- Brenner, A.J., and Aldaz, C.M. (1997) The genetics of sporadic breast cancer. *Prog. Clin. Biol. Res.* **396**: 63-82.
- Brinkman, A., van der Flier, S., Kok, E.M., and Dorssers, L.C.J. (2000) BCAR1, a human homologue of the adapter protein p130Cas, and antiestrogen resistance in breast cancer cells. *J. Natl. Cancer Inst.* **92**: 112-120.
- Broeks, A., Urbanus, J.H.M., Floore, A.N., Dahler, E.C., Klijju, J.G.M., Rutgers, E.J.T., Devilee, P., Russell, N.S., van Leeuwen, F.E., and van 't Veer, L.J. (2000) *ATM*-heterozygous germline mutations contribute to breast cancer-susceptibility. *Am. J. Hum. Genet.* **66**: 494-500.
- 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. U.S.A.* **88**: 4005-4009.
- Buetow, K.H., Weber, J.L., Ludwigsen, S., Scherpbier-Heddeme, T., Duyk, G.M., Sheffield, V.C., Wang, Z., and Murray, J.C. (1994) Integrated human genome-wide maps constructed using the CEPH reference panel. *Nature Genet.* **6**: 391-393.
- Bullrich, F., Maclachlan, T.K., Sang, N., Druck, T., Veronese, M.L., Allen, S.L., Chiorazzi, N., Koff, A., Heubner, K., Croce, C.M., and Giordano, A. (1995) Chromosomal mapping of members of the *cdc2* family of protein kinases, *cdk3*, *cdk6*, *PISSLRE*, and *PITALRE*, and a *cdk* inhibitor, *p27<sup>Kip1</sup>*, to regions involved in human cancer. *Cancer Res.* **55**: 1199-1205.
- Burke, D.T. (1991) The role of yeast artificial chromosome clones in generating genome maps. *Curr. Opin. Genet. Dev.* **1**:69-74.
- Burke, D.T., Carle, G.F., and Olson, M.V. (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* **236**:806-812.
- Burn, T.C., Connors, T.D., Klinger, K.W., and Landes, G.M. (1995) Increased exon-trapping efficiency through modifications to the pSPL3 splicing vector. *Gene* **161**: 183-187.
- Burn, T.C., Connors, T.D., Van Raay, T.J., Dackowski, W.R., Millholland, J.M., Klinger, K.W., and Landes, G.M. (1996) Generation of a transcriptional map for a 700-kb region surrounding the polycystic kidney disease type 1 (*PKD1*) and tuberous sclerosis type 2 (*TSC2*) disease genes on human chromosome 16p3.3. *Genome Res.* **6**: 525-537.
- Butturini, A., Gale, R.P., Verlander, P.C., Adler-Brecher, B., Gillio, A.P., and Auerbach, A.D. (1994) Hematologic abnormalities in Fanconi anemia : an International Fanconi Anemia Registry study. *Blood.* **48**:1650-1655.
- Cahill, D.R., Lengaver, C., Yu, J., et al. (1998) Mutation of mitotic checkpoint genes in human cancers. *Nature* **392**: 300-303.
- Cairns, J. (1975) Mutation selection and the natural history of cancer. *Nature* **255**: 197-200.

- Callebaut, I., and Mornon, J.P. (1997) From BRCA1 to RAP2: a widespread BRCT module closely associated with DNA repair. *FEBS Lett.* **400**: 25-30.
- Callen, D.F., Lane, S.A., Kozman, H., Kremmidiotis, G., Whitmore, S.A., Lowenstein, M., Doggett, N.A., Kenmochi, N., Page, D.C., Maglott, D.R., Nierman, W.C., Murakawa, K., Berry, R., Sikela, J.M., Houlgatte, R., Auffray, C., and Sutherland, G.R. (1995) Integration of transcript and genetic maps of chromosome 16 at near-1-Mb resolution: demonstration of a "hot spot" for recombination at 16p12. *Genomics* **29**: 503-511.
- Canman, C.E., Lim, D.S., Cimprich, K.A., Taya, Y., Tamai, K., Sakaguchi, K., et al. (1998) Activation of the ATM kinase by ionizing radiation and phosphorylation of p53. *Science* **281**: 1677-1679.
- Cantu, E.S., Sprinkle, T.J., Ghosh, B., and Singh, I. (1995) The human palmitoyl-CoA ligase (*FACL2*) gene maps to the chromosome 4q34-q35 region by fluorescence in situ hybridization (FISH) and somatic cell hybrid panels. *Genomics* **28**: 600-602.
- Cao, Y., Dave, K.B., Doan, T.P., and Prescott, S.M. (2001) Fatty acid CoA ligase 4 is up-regulated in colon adenocarcinoma. *Cancer Res.* **61**: 8429-8434.
- Cao, Y., Traer, E., Zimmerman, G.A., McIntyre, T.M., and Prescott SM. (1998) Cloning, expression, and chromosomal localization of human long-chain fatty acid-CoA ligase 4 (*FACL4*). *Genomics* **49**: 327-330.
- Carter, B.S., Ewing, C.M., Ward, W.S., Treiger, B.F., Aalders, T.W., Schalken, J.A., Epstein, J.I., and Isaacs, W.B. (1990) Allelic loss of chromosome 16q and 10q in human prostate cancer. *Proc. Natl. Acad. Sci. U.S.A.* **87**: 8751-8755.
- Casari, G., De Fusco, M., Ciarmatori, S., Zeviani, M., Mora, M., Fernandez, P., De Michele, G., Filla, A., Coccozza, S., Marconi, R., Durr, A., Fontaine, B., and Ballabio, A. (1998) Spastic paraplegia and OXPHOS impairment caused by mutations in paraplegin, a nuclear-encoded mitochondrial metalloprotease. *Cell* **93**: 973-983.
- Castagna, M. (1987) Phorbol esters as signal transducers and tumor promoters. *Biol. Cell* **59**: 3-13.
- Catteau, A., Harris W.H., Xu, C.F., and Solomon, E. (1999) Methylation of the *BRCA1* promoter region in sporadic breast and ovarian cancer: correlation with disease characteristics. *Oncogene* **18**: 1957-1965.
- Cavenee, W.K., Dryja, T.P., Phillips, R.A., Benedict, W.F., Godbout, R., Gallie, B.L., Murphree, A.L., Strong, L.C., and White, R.L. (1983) Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. *Nature* **305**: 779-784.
- Champème, M.H., Bièche, I., Lizard, S., and Lidereau, R. (1995) 11q13 amplification is involved in local recurrence of human primary breast cancer. *Genes Chromosomes Cancer* **12**: 128-133.
- Chen, F., Castranova, V., and Shi, X. (2001) New insights into the role of nuclear factor-kappaB in cell growth regulation. *Am. J. Pathol.* **159**: 387-397.

- Chen, J., Birkholtz, G.G., Lindblom, P., Rubio, C., and Lindblom, A. (1998a) The role of ataxia-telangiectasia heterozygotes in familial breast cancer. *Cancer Res.* **58**: 1376-1379.
- Chen, J., Silver, D.P., Walpita, D., Cantor, S.B., Gazdar, A.F., Tomlinson, G., Couch, F.J., Weber, B.L., Ashley, T., Livingston, D.M., and Scully, R. (1998b) Stable interaction between the products of the *BRCA1* and *BRCA2* tumor suppressor genes in mitotic and meiotic cells. *Mol. Cell* **2**: 317-328.
- Chen, T., Sahin, A., and Aldaz, C.M. (1996) Deletion map of chromosome 16q in ductal carcinoma *in situ* of the breast: refining a putative tumor suppressor gene region. *Cancer Res.* **56**: 5605-5609.
- Cher, M.L., Ito, T., Weidner, N., Carroll, P.R., and Jensen, R.H. (1995) Mapping of regions of physical deletion on chromosome 16q in prostate cancer cells by fluorescence *in situ* hybridization (FISH). *J. Urol.* **153**: 249-254.
- Chomczynski, P., and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**: 156-159.
- 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**: 98-105.
- Church, D.M., Yang, J., Bocian, M., Shiang, R., and Wasmuth, J.J. (1997) A high-resolution physical and transcript map of the Cri du chat region of human chromosome 5p. *Genome Res.* **7**: 787-801.
- Clarke, R.B., Howell, A., Potten, C.S., and Anderson, E. (1997) Dissociation between steroid receptor expression and cell proliferation in the human breast. *Cancer Res.* **57**: 4987-4991.
- Cleton-Jansen, A-M., Moerland, E.W., Kuipers-Dijkshoorn, N.J., Callen, D.F., Sutherland, G.R., Hansen, B., Devilee, P., and Cornelisse, C.J. (1994) At least two different regions are involved in allelic imbalance on chromosome arm 16q in breast cancer. *Genes Chromosomes Cancer* **9**: 101-107.
- Cleton-Jansen, A-M., Moerland, E.W., Pronk, J.C., van Berkel, C., Apostolou, S., Crawford, J., Savoia, A., Auerbach, A.D., Mathew, C.G., Callen, D.F., and Cornelisse, C.J. (1999) Mutation analysis of the Fanconi anemia A gene in breast tumours with loss of heterozygosity at 16q24.3. *Br. J. Cancer* **79**: 1049-1052.
- Coffer, P.J., Jin, J., and Woodgett, J.R. (1998) Protein kinase B (c-Akt): a multifunctional mediator of phosphatidylinositol 3-kinase activation. *Biochem J.* **335** ( Pt 1):1-13.
- Collins, F.S. (1995a) Positional cloning moves from perdictional to traditional. *Nature Genet.* **9**: 347-350.
- Collins, N., McManus, R., Wooster, R., Mangion, J., Seal, S., et al. (1995b) Consistent loss of the wild-type allele in breast cancers from a family linked to the *BRCA2* gene on

chromosome 13q12-13. *Oncogene* **10**: 1673-1675.

Collins, N., Wooster, R., and Stratton, M.R. (1997) Absence of methylation of CpG dinucleotides within the promoter of the breast cancer susceptibility gene *BRCA2* in normal tissues and in breast and ovarian cancers. *Br. J. Cancer* **76**: 1151-1156.

Cornelis, R.S., Neuhausen, S.L., Johansson, O., Arason, A., Kelsell, D., et al. (1995) High allele loss rates at 17q12-q21 in breast and ovarian tumor from *BRCA1*-linked families. *Genes Chromosomes Cancer* **13**: 203-210.

Coppola, M., Pizzigoni, A., Banfi, S., Bassi, M.T., Casari, G., and Incerti, B. (2000) Identification and characterization of *YME1L1*, a novel paraplegin-related gene. *Genomics* **66**: 48-54.

Cortez, D., Wang, Y., Qin, J., Elledge, S.J. (1999) Requirement of ATM-dependent phosphorylation of Brca1 in the DNA damage response to double-strand breaks. *Science* **286**: 1162-1166.

Cory, S. (1995) Regulation of lymphocyte survival by the bcl-2 gene family. *Annu. Rev. Immunol.* **13**: 513-543.

Couch, F.J., and Weber, B.L. (1996) Mutations and polymorphisms in the familial early-onset breast cancer (*BRCA1*) gene. Breast Cancer Information Core. *Hum. Mut.* **8**: 8-18.

Coussens, L., Yang-Feng, T.L., Liao, Y.C., Chen, E., Gray, A., McGrath, J., et al. (1985) Tyrosine kinase receptor with extensive homology to EGF receptor shares chromosomal location with *Neu* oncogene. *Science* **230**: 1132-1139.

Coutinho, P., Barros, J., Zemmouri, R., Guimaraes, J., Alves, C., Choro, R., Lourenco, E., Ribeiro, P., Loureiro, J.L., Santos, J.V., Hamri, A., Paternotte, C., Hazan, J., Silva, M.C., Prud'homme, J.F., and Grid, D. (1999) Clinical heterogeneity of autosomal recessive spastic paraplegias: analysis of 106 patients in 46 families. *Arch. Neurol.* **56**: 943-949.

Crawford, J., Ianzano, L., Savino, M., Whitmore, S.A., Cleton-Jansen, A-M., Settasatian, C., d'Apolito, M., Seshadri, R., Pronk, J.C., Auerbach, A.D., Verlander, P.C., Mathew, C.G., Tipping, A.J., Doggett, N.A., Zelante, L., Callen, D.F., and Savoia, A. (1999) The *PISLRE* gene: Structure, exon skipping, and exclusion as tumor suppressor in breast cancer. *Genomics* **56**: 90-97.

Croce, C.M., and Nowell, P.C. (1985) Molecular basis of human B-cell neoplasia. *Blood* **65**: 1-7.

Culbertson, M.R. (1999) RNA surveillance. Unforeseen consequences for gene expression, inherited genetic disorders and cancer. *Trends Genet.* **15**: 74-80.

Dahia, P.L. (2000) *PTEN*, a unique tumor suppressor gene. *Endocr. Relat. Cancer* **7**:115-129.

Dasika, G.K., Lin, S.J., Zhao, S., Sung, P., Tomkinson, A., and Lee, E.Y-H.P. (1999) DNA damage-induced cell cycle checkpoints and DNA strand break repair in development and

tumorigenesis. *Oncogene* **18**: 7883-7899.

Datson, N.A., Duyk, G.M., van Ommen, G.J., and den Dunnen, J.T. (1994) Specific isolation of 3'-terminal exons of human genes by exon trapping. *Nucleic Acids Res.* **22**: 4148-4153.

Dechend, R., Hirano, F., Lehmann, K., Heissmeyer, V., Ansieau, S., Wulczyn, F.G., Scheidereit, C., and Leutz, A. (1999) The Bcl-3 oncoprotein acts as a bridging factor between NF-kappaB/Rel and nuclear co-regulators. *Oncogene* **18**: 3316-3323.

De Klein, A., Geurts van Kessel, A., Grosveld, G., Bartram, C.R., Hagemeijer, A., Bootsma, D., Spurr, N.K., Heisterkamp, N., Groffen, J., and Stephenson, J.R. (1982) A cellular oncogene is translocated to the Philadelphia chromosome in chronic myelocytic leukemia. *Nature* **300**: 765-767.

Deng, C.X., and Scott, F. (2000) Role of the tumor suppressor gene Brcal in genetic stability and mammary gland tumor formation. *Oncogene* **19**: 1059-1064.

Deng, G., Lu, Y., Zlotnikov, G., Thor, A.D., and Smith, H.S. (1996) Loss of heterozygosity in normal tissue adjacent to breast carcinomas. *Science* **274**: 2057-2059.

De Michele, G., De Fusco, M., Cavalcanti, F., Filla, A., Marconi, R., Volpe, G., Monticelli, A., Ballabio, A., Casari, G., and Coccozza, S. (1998) A new locus for autosomal recessive hereditary spastic paraplegia maps to chromosome 16q24.3. *Am. J. Hum. Genet.* **63**: 135-139.

Devilee, P., and Cornelisse, C.J. (1994) Somatic genetic changes in human breast cancer. *Biochimica et Biophysica Acta* **1198**: 113-130.

DiRusso, C.C., Black, P.N., and Weimar, J.D. (1999) Molecular inroads into the regulation and metabolism of fatty acids, lessons from bacteria. *Prog. Lipid Res.* **38**: 129-197.

Dobrovic, A., and Simpfendorfer, D. (1997) Methylation of the *BRCA1* gene in sporadic breast cancer. *Cancer Res.* **57**: 3347-3350.

Donehower, L.A., and Bradley, A. (1993) The tumor suppressor p53. *Biochim. Biophys. Acta Rev. Cancer* **1155**: 181-205.

Dorion-Bonnet, F., Mautalen, S., Hostein, I., and Longy, M. (1995) Allelic imbalance study of 16q in human primary breast carcinomas using microsatellite markers. *Genes Chromosomes Cancer* **14**: 171-181.

Dorssers, L.C., van Agthoven, T., Dekker, A., van Agthoven, T.L., and Kok, E.M. (1993) Induction of antiestrogen resistance in human breast cancer cells by random insertional mutagenesis using defective retroviruses: identification of Bcar-1, a common integration site. *Mol. Endocrinol.* **7**: 870-878.

Driouch, K., Dorion-Bonnet, F., Briffod, M., Champeme, M-H., Longy, M., and Lidereau, R. (1997) Loss of heterozygosity on chromosome arm 16q in breast cancer metastases. *Genes Chromosomes Cancer* **19**: 185-191.



- Durbin, H., Novelli, M., and Bodmer, W. (1994) Detection of a 4-bp insertion (CACA) functional polymorphism at nucleotide 241 of the cellular adhesion regulatory molecule CMAR (formerly CAR). *Genomics* **19**: 181-182.
- Durbin, H., Novelli, M.R., and Bodmer, W.F. (1997) Genomic and cDNA sequence analysis of the cell matrix adhesion regulator gene. *Proc. Natl. Acad. Sci. U. S. A.* **94**: 14578-14583.
- Duyk, G.M., Kim, S., Myers, R.M., and Cox, D.R. (1990) Exon trapping: a genetic screen to identify candidate transcribed sequences in clone mammalian genomic DNA. *Proc. Natl. Acad. Sci. U.S.A.* **87**: 8995-8999.
- Easton, D.F. (1994) Cancer risks in A-T heterozygotes. *Int. J. Radiat. Biol.* **66**(Suppl 6): S177-S182.
- Easton, D.F., Bishop, D.T., Ford, D., Crockford, G.P., and Breast Cancer Linkage Consortium. (1993) Genetic linkage analysis in familial breast and ovarian cancer: results from 214 families. *Am. J. Hum. Genet.* **52**: 678-701.
- Edwards-Gilbert, G., and Milcarek, C. (1995) Regulation of poly(A) site use during mouse B-cell development involves a change in the binding of a general polyadenylation factor in a B-cell stage-specific manner. *Mol. Cell. Biol.* **15**: 6420-6429.
- El-Ashry, D., and Lippman, M.E. (1994) Molecular biology of breast carcinoma. *World J. Surg.* **18**: 12-20.
- Elo, J.P., Harkonen, P., Kyllonen, A.P., Lukkarinen, O., Poutanen, M., Vihko, R., and Vihko, P. (1997) Loss of heterozygosity at 16q24.1-q24.2 is significantly associated with metastatic and aggressive behavior of prostate cancer. *Cancer Res.* **57**: 3356-3359.
- Eppert, K., Scherer, S.W., Ozcelik, H., Pirone, R., Hoodless, P., Kim, H., Tsui, L.C., Bapat, B., Gallinger, S., Andrulis, I.L., Thomsen, G.H., Wrana, J.L., and Attisano, L. (1996) MADR2 maps to 18q21 and encodes a TGF beta-regulated MAD related protein that is functionally mutated in colorectal carcinoma. *Cell* **86**: 543-552.
- Escot, C., Theiller, C., Lidereau, R., Spyrtos, F., Champème, M.H., Gest, J., and Callahan, R. (1986) Genetic alteration of the *c-Myc* proto-oncogene (*MYC*) in human primary breast carcinomas. *Proc. Natl. Acad. Sci. U.S.A.* **83**: 4834-4838.
- Eshleman, J.R., Lang, E.Z., Bowerfind, G.K., Parsons, R., Vogelstein, B., Willson, J.K., Veigl, M.L., Sedwick, W.D., and Markowitz, S.D. (1995) Increased mutation rate at the *hprt* locus accompanies microsatellite instability in colon cancer. *Oncogene* **10**: 33-37.
- Esnault, C., Maestre, J., and Heidmann, T. (2000) Human LINE retrotransposons generate processed pseudogenes. *Nature Genet.* **24**: 363-367.
- Esteller, M., Silva, J.M., Dominguez, G., Bonilla, F., Matias-Guiu, X., Lerma, E., Bussaglia, E., Prat, J., Harkes, I.C., Repasky, E.A., Gabrielson, E., Schutte, M., Baylin, S.B., and Herman, J.G. (2000) Promoter hypermethylation and *BRCA1* inactivation in sporadic breast and ovarian tumors. *J. Natl. Cancer Inst.* **92**: 564-569.

- Ewen, M.E. (1994) The cell cycle and the retinoblastoma protein family. *Cancer Metastasis Reviews* **13**: 45-66.
- Fan, S., Wang, J.A., Yuan, R., Meng, O., Yuan, R.O., Ma, Y.X., Erdos, M.R., Pestell, R.G., Yuan, F., Auborn, K.J., Goldberg, I.D., and Rosen, E.M. (1999) BRCA1 Inhibition of Estrogen Receptor Signaling in Transfected Cells. *Science* **284**: 1354-1356.
- Fanconi Anaemia/Breast Cancer Consortium (1996) Positional cloning of the Fanconi anaemia group A gene. *Nature Genet.* **14**: 324-328.
- Fearon, E.R., and Vogelstein, B. (1990) A genetic model for colorectal tumorigenesis. *Cell* **61**: 759-767.
- Feinberg, A.P., and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132**: 6-13.
- Filippova, G.N., Fagerlie, S., Klenova, E., Myers, C., Dehner, Y., Goodwin, G.H., Neiman, P.E., Collins, S., and Lobanekov, V.V. (1996) An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian *c-myc* oncogenes. *Mol. Cell Biol.* **16**: 2802-2813.
- Filippova, G.N., Lindblom, A., Meincke, L.J., Klenova, E.M., Neiman, P.E., Collins, S.J., Doggett, N.A., and Lobanekov, V.V. (1998) A widely expressed transcription factor with multiple DNA sequence specificity, CTCF, is localized at chromosome segment 16q22.1 within one of the smallest regions of overlap for common deletions in breast and prostate cancers. *Genes Chrom. Cancer* **22**: 26-36.
- Fink, J.K., Wu, C.T., Jones, S.M., Sharp, G.B., Lange, B.M., Lesicki, A., Reinglass, T., Varvil, T., Otterud, B., and Leppert, M. (1995) Autosomal dominant familial spastic paraplegia: tight linkage to chromosome 15q. *Am. J. Hum. Genet.* **56**: 188-192.
- Fink, J.K. (1997) Advances in hereditary spastic paraplegia. *Curr. Opin. Neurol.* **10**: 313-318.
- Fishman, J., Osborne, M.P., and Telang, N.T. (1995) The role of estrogen in mammary carcinogenesis. *Ann. NY Acad. Sci.* **768**: 91-100.
- FitzGerald, M.G., Bean, J.M., Hegde, S.R., Unsal, H., MacDonald, D.J., Harkin, D.P., *et al.* (1997) Heterozygous *ATM* mutations do not contribute to early onset of breast cancer. *Nature Genet.* **15**: 307-310.
- FitzGerald, M.G., MacDonal, D.J., Krainer, M., Hoover, I., O' Neil, E., *et al.* (1996) Germline *BRCA1* mutations in Jewish and non-Jewish women with early-onset breast cancer. *N. Eng. J. Med.* **334**: 143-149.
- Foekens, J.A., Rio, M.C., Seguin, P., Van Putten, W.L., Fauque, J, Nap, M., *et al.* (1990) Prediction of relapse and survival in breast cancer patients by pS2 protein status. *Cancer Res.* **50**: 3832-3837.

- Foekens, J.A., Schmitt, M., Van Putten, W.L., Peters, H.A., Kramer, M.D., Janicke, F., et al. (1994) Plasminogen activator inhibitor-1 and prognosis in primary breast cancer. *J. Clin. Oncol.* **12**: 1648-1658.
- Fontaine, B., Davoine, C.S., Durr, A., Paternotte, C., Feki, I., Weissenbach, J., Hazan, J., and Brice, A. (2000) A new locus for autosomal dominant pure spastic paraplegia, on chromosome 2q24-q34. *Am. J. Hum. Genet.* **66**: 702-707.
- Francke, U., and Kung, F. (1976) Sporadic bilateral retinoblastoma and 13q-chromosomal deletion. *Med. Pediatr. Oncol.* **2**: 379-385.
- Friend, S.H., Bernards, R., Rogelj, S., Weinberg, R.A., Rapaport, J.M., Albert, D.M., and Dryja, T.P. (1986) A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature* **323**: 643-646.
- Frischmeyer, P.A., and Dietz, H.C. (1999) Nonsense-mediated mRNA decay in health and disease. *Hum. Mol. Genet.* **8**:1893-1990.
- Fujii, H., Marsh, C., Cairns, P., Sidransky, D., and Gabrielson, E. (1996) Genetic divergence in the clonal evolution of breast cancer. *Cancer Res.* **56**: 1493-1497.
- Fulceri, R., Gamberucci, A., Scott, H.M., Giunti, R., Burchell, A., and Benedetti, A. (1995) Fatty acyl-CoA esters inhibit glucose-6-phosphatase in rat liver microsomes. *Biochem. J.* **307**: 391-397.
- Futreal, P.A., Liu, Q., Shattuck-Eidens, D., Cochran, C., Harshman, K., et al. (1994) BRCA1 mutations in primary breast and ovarian carcinomas. *Science* **266**: 120-122.
- Garber, J.E., Goldstein, A.M., Kantor, A.F., Dreyfus, M.G., Fraumeni, J.F., Jr., and Li, F.P. (1991) Follow-up study of twenty-four families with Li-Fraumeni syndrome. *Cancer Res.* **51**: 6094-6097.
- Galaktionov, K., Chen, X., and Beach, D. (1996) Cdc25 cell-cycle phosphatase as a target of c-myc. *Nature* **382**: 511-517.
- Gatei, M., Scott, S.P., Filippovitch, I., Soronika, N., Lavin, M.F., Weber, B., and Khanna, K.K. (2000) Role for ATM in DNA damage-induced phosphorylation of BRCA1. *Cancer Res.* **60**: 3299-3304.
- Gatti, R.A., Boder, E., Vinters, H.V., Sparkes, R.S., Norman, A., and Lange, K. (1991) Ataxia-telangiectasia: an interdisciplinary approach to pathogenesis. *Medicine* **70**: 99-117.
- Gecz, J., Villard, L., Lossi, A.M., Millasseau, P., Djabali, M., and Fontes, M. (1993) Physical and transcriptional mapping of DXS56-PGK1 1 Mb region: identification of three new transcripts. *Hum. Mol. Genet.* **2**: 1389-1396.
- Gibbons, R.J., Bachoo, S., Picketts, D.J., Aftimos, S., Asenbauer, B., Bergoffen, J., Berry, S.A., Dahl, N., et al. (1997) Mutations in transcriptional regulator *ATR*X establish the functional significance of a PHD-like domain. *Nature Genet.* **17**: 146-148.

- Giedraitis, V., He, B., Huang, W.X., and Hillert, J. (2001) Cloning and mutation analysis of the human IL-18 promoter: a possible role of polymorphisms in expression regulation. *J Neuroimmunol.* **112**: 146-152.
- Gigli, G.L., Diomedi, M., Bernardi, G., Placidi, F., Marciani, M.G., Calia, E., Maschio, M.C., and Neri, G. (1993) Spastic paraplegia, epilepsy, and mental retardation in several members of a family: a novel genetic disorder. *Am. J. Med. Genet.* **45**: 711-716.
- Gillett, C., Fantl, V., Smith, R., Fisher, C., Bartek, J., Dickson, C., Barnes, D., and Peters, G. (1994) Amplification and overexpression of cyclin D1 in breast cancer detected by immunohistochemical staining. *Cancer Res.* **54**: 1812-1817.
- Glick, B.S., and Rothman, J.E. (1987) Possible role for fatty acyl-coenzyme A in intracellular protein transport. *Nature* **326**: 309-312.
- Godfrey, T.E., Cher, M.L., Chhabra, V., and Jensen, R.H. (1997) Allelic imbalance mapping of chromosome 16 shows two regions of common deletion in prostate adenocarcinoma. *Cancer Genet. Cytogenet.* **98**: 36-42.
- Go, R.C., King, M.C., Bailey-Wilson, J., Elston, R.C., Lynch, H.T. (1983) Genetic epidemiology of breast cancer and associated cancers in high-risk families. I. Segregation analysis. *J. Natl. Cancer Inst.* **71**: 455-461.
- Gordon, J.I., Duronio, R.S., Rudnick, D.A., Adams, S.P., and Gokel, G.W. (1991) xxxT18xx *J. Biol. Chem.* **266**: 8647-8650.
- Gowen, L.C., Avrutskaya, A.V., Latour, A.M., Koller, B.H., and Leadon, S.A. (1998) BRCA1 required for transcription-coupled repair of oxidative DNA damage. *Science* **281**: 1009-1012.
- Grady, W.M., Rajput, A., Myeroff, L., Liu, D.F., Kwon, K., Willis, J., and Markowitz, S. (1998) Mutation of the type II transforming growth factor- $\beta$  receptor is coincident with the transformation of human colon adenomas to malignant carcinomas. *Cancer Res.* **58**: 3101-3104.
- Grady, W.M., Myeroff, L.L., Swinler, S.E., Rajput, A., Thiagalingam, S., Lutterbaugh, J.D., et al. (1999) Mutational inactivation of transforming growth factor- $\beta$  receptor type II in microsatellite stable colon cancers. *Cancer Res.* **59**: 320-324.
- Grana, X., Claudio, P.P., De Luca, A., Sang, N., and Giordano, A. (1994) PISSLRE, a human novel CDC2-related protein kinase. *Oncogene* **9**: 2097-2103.
- Green, E.D., Riethman, H.C., Dutchik, J.E., and Olson, M.V. (1991) Detection and characterization of chimeric yeast artificial-chromosome clones. *Genomics* **11**: 658-669.
- Groden, J., Thliveris, A., Samowitz, W., Carlson, M., Gelbert, L., Albertsen, H., et al. (1991) Identification and characterization of the familial adenomatous polyposis coli gene. *Cell* **66**: 589-600.

- Gudas, J.M., Li, T., Nguyen, H., Jensen, D., Rauscher, F.J., and Cowan, K.H. (1996) Cell cycle regulation of BRCA1 messenger RNA in human breast epithelial cells. *Cell Growth Differ.* **7**: 717-723.
- Gumbiner, B.M. (1995) Signal transduction of beta-catenin. *Curr. Opin. Cell Biol.* **7**: 634-640.
- Gu, Y., Nakamura, T., Alder, H., Prasad, R., Canaani, O., *et al.* (1992) The t(4;11) chromosome translocation of human acute leukemias fuses the *ALL-1* gene, related to *Drosophila trithorax*, to the *AF-4* gene. *Cell* **71**: 701-708.
- Gyapy, G., Morisette, J., Dib, C., Fizames, C., Millasseau, P., Marc, S., Bernardi, G., Lathrop, M., and Weissenbach, J. (1994) The 1993-4 Genethon human genetic linkage map. *Nature Genet.* **7**: 246-339.
- Halbach, T., Scheer, N., and Werr, W. (2000) Transcriptional activation by the PHD finger is inhibited through an adjacent leucine zipper that binds 14-3-3 proteins. *Nucleic Acids Res.* **28**: 3542-3550.
- 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 familial breast cancer to chromosome 17q21. *Science* **250**: 1684-1689.
- Hanssen, A.M.N., and Fryns, J.P. (1995) Cowden syndrome. *J. Med. Genet.* **32**: 117-119.
- Harbour, J.W., Lai, S-L., Whang-Peng, J., Gazdar, A.D., Minna, J.D., and Kaye, F.J. (1988) Abnormalities in structure and expression of the human retinoblastoma gene in SCLC. *Science* **241**: 353-357.
- Harding, A.E. (1981) Hereditary 'pure' spastic paraplegia: a clinical and genetic study of 22 families. *J. Neurol. Neurosurg. Psychiat.* **44**: 871-883.
- Harkin, P.D., Bean, J.M., Miklos, D., Song, Y-H., Truong, V.B., Englert, C., Christians, F.C., Ellisen, L.W., Maheswaran, J., Oliner, J.D., and Harber, D.A. (1999) Induction of GADD45 and JNK/SARK-dependent apoptosis following inducible expression of BRCA1. *Cell* **97**: 575-586.
- Harris, H. (1988) The analysis of malignancy by cell fusion: the position in 1988. *Cancer Res.* **48**: 3302-3306.
- Hasenpusch-Theil, K., Chadwick, B.P., Theil, T., Heath, S.K., Wilkinson, D.G., and Frischauf, A-M. (1999) *PHF2*, a novel PHD finger gene located on human Chromosome 9q22. *Mamm. Genome* **10**: 294-298.
- Hayashi, K. (1991) PCR-SSCP: a simple and sensitive method for detection of mutations in the genomic DNA. *PCR Methods Appl.* **1**: 34-38.
- Hazan, J., Fonknechten, N., Mavel, D., Paternotte, C., Samson, D., Artiguenave, F., Davoine, C.S., Cruaud, C., Durr, A., Wincker, P., Brottier, P., Cattolico, L., Barbe, V., Burgunder, J.M., Prud'homme, J.F., Brice, A., Fontaine, B., Heilig, B., Weissenbach, J. (1999)

- Spastin, a new AAA protein, is altered in the most frequent form of autosomal dominant spastic paraplegia. *Nature Genet.* **23**: 296-303.
- Hazan, J., Fontaine, B., Bruyn, R.P., Lamy, C., van Deutekom, J.C., Rime, C.S., Durr, A., Melki, J., Lyon-Caen, O., Agid, Y., *et al.* (1994) Linkage of a new locus for autosomal dominant familial spastic paraplegia to chromosome 2p. *Hum. Mol. Genet.* **3**: 1569-1573.
- Hazan, J., Lamy, C., Melki, J., Munnich, A., de Recondo, J., and Weissenbach, J. (1993) Autosomal dominant familial spastic paraplegia is genetically heterogeneous and one locus maps to chromosome 14q. *Nature Genet.* **5**: 163-167.
- Hedera, P., Rainier, S., Alvarado, D., Zhao, X., Williamson, J., Otterud, B., Leppert, M., Fink, J.K. (1999) Novel locus for autosomal dominant hereditary spastic paraplegia, on chromosome 8q. *Am. J. Hum. Genet.* **64**: 563-569.
- He, T-C., Sparks, A.B., Rago, C., Hermeking, H., Zawel, L., da Costa, L.T., Morin, P.J., Vogelstein, B., and Kinzler, K.W. (1998) Identification of *c-MYC* as a target of the APC pathway. *Science* **281**: 1509-1512.
- Heisterkamp, N., Stam, K., Groffen, J., de Klein, A., and Grosveld, G. (1985) Structural organization of the *BCR* gene and its role in the Ph' translocation. *Nature* **315**: 758-761.
- Heisterkamp, N., Stephenson, J.R., Groffen, J., Hansen, P., de Klein, A., Bartram, C.R., and Grosveld, G. (1983) Localization of the *c-ABL* oncogene adjacent to a translocation breakpoint in chronic myelocytic leukemia. *Nature* **306**: 239-242.
- Hentati, A., Pericak-Vance, M.A., Hung, W.Y., Belal, S., Laing, N., Boustany, R.M., Hentati, F., Ben Hamida, M., Siddique, T. (1994a) Linkage of 'pure' autosomal recessive familial spastic paraplegia to chromosome 8 markers and evidence of genetic locus heterogeneity. *Hum. Mol. Genet.* **3**: 1263-1267.
- Hentati, A., Pericak-Vance, M.A., Lennon, F., Wasserman, B., Hentati, F., Juneja, T., Angrist, M.H., Hung, W.Y., Boustany, R.M., Bohlega, S., *et al.* (1994b) Linkage of a locus for autosomal dominant familial spastic paraplegia to chromosome 2p markers. *Hum. Mol. Genet.* **3**: 1867-1871.
- Hertz, R., Magenheimer, J., Berman, I., and Bar-Tana, J. (1998) Fatty acyl-CoA thioesters are ligands of hepatic nuclear factor-4alpha. *Nature* **392**: 512-516.
- Higgins, D.G., Thompson, J.D., and Gibson, T.J. (1996) Using CLUSTAL for multiple sequence alignments. *Methods Enzymol.* **266**: 383-402.
- Hiraguri, S., Godfrey, T., Nakamura, H., Graff, J., Collins, C., Shayesteh, L., Doggett, N., Johnson, K., Wheelock, M., Herman, J., Baylin, S., Pinkel, D., and Gray, J. (1998) Mechanisms of inactivation of E-cadherin in breast cancer cell lines. *Cancer Res.* **58**: 1972-1977.
- Hofmann, K., Bucher, P., Falquet, L., and Bairoch, A. (1999) The PROSITE database, its status in 1999. *Nucleic Acids Res.* **27**: 215-219.

- Hofmann, S.L., Das, A.K., Lu, J.Y., and Soyombo, A.A. (2001) Positional candidate gene cloning of CLN1. *Adv. Genet.* **45**: 69-92.
- Holt, J.T., Thompson, M.E., Szabo, C., Robinson-Benion, C., Arteaga, C.L., King, M.C., and Jensen, R.A. (1996) Growth retardation and tumour inhibition by BRCA1. *Nature Genet.* **12**: 223-225.
- Holt, J.T. (1997) Breast cancer genes: therapeutic strategies. *Ann. NY. Acad. Sci.* **833**:34-41
- Horowitz, J.M., Park, S-H., Bogenmann, E., Cheng, J-C., Yandell, D.W., Kaye, F.J., Minna, J.D., Dryja, T.P., and Weinberg, R.A. (1990) Frequent inactivation of the retinoblastoma anti-oncogene is restricted to a subset of human tumor cells. *Proc. Natl. Acad. Sci. U.S.A.* **87**: 2775-2779.
- Horwitz, K.B., Wei, L.L., Sedlacek, S.M., and D' Arville, C.N. (1985) Progestin action and progesterone receptor structure in human breast cancer: a review. *Recent Prog. Horm. Res.* **41**: 249-316.
- Hotfilder, M., Baxendale, S., Cross, M.A., and Sablitzky, F. (1999) Def-2, -3, -6, and -8, novel mouse genes differentially expressed in haemopoietic system. *Br. J. Haemato.* **106**: 335-344.
- Hsu, L.C. and White, R.L. (1998) BRCA1 is associated with the centrosome during mitosis. *Proc. Natl. Acad. Sci. U.S.A.* **95**: 12983-12988.
- Hugot, J.-P., Laurent-Puig, P., Gower-Rousseau, C., Olson, J. M., Lee, J. C., Beaugierie, L., Naom, I., Dupas, J.-L., Van Gossum, A., Groupe d'Etude Therapeutique des Affections Inflammatoires Digestives, Orholm, M., Bonaiti-Pellie, C., Weissenbach, J., Mathew, C. G., Lennard-Jones, J. E., Cortot, A., Colombel, J.-F., and Thomas, G. (1996) Mapping of a susceptibility locus for Crohn's disease on chromosome 16. *Nature* **379**: 821-823.
- Ichii, S., Horii, A., Nakatsuru, S., Furuyama, J., Utsunomiya, J., and Nakamura, Y. (1992) Inactivation of both *APC* alleles in an early stage of colon adenomas in a patient with familial adenomatous polyposis (FAP). *Hum. Mol. Genet.* **1**: 387-390.
- Iida, A., Isobe, R., Yoshimoto, M., Kasumi, F., Nakamura, Y., and Emi, M. (1997) Localization of a breast cancer tumor-suppressor gene to a 3-cM interval within chromosome region 16q22. *Br. J. Cancer* **75**: 264-267.
- Ilyas, M., and Tomlinson, I.P. (1997) The interactions of APC, E-cadherin, and  $\beta$ -catenin in tumor development and progression. *J. Pathol.* **182**: 128-137.
- Inskip, H.M., Kinlen, L.J., Taylor, A.M.R., Woods, C.G., and Arlett, C.F. (1999) Risk of breast cancer and other cancers in heterozygotes for ataxia-telangiectasia. *Br. J. Cancer* **79**: 1304-1307.
- Ioannou, P.A., Amemiya, C.T., Garnes, J., Kroisel, P.M., Shizuya, H., Chen, C., Batzer, M.A., and de Jong, P.J. (1994) A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nature Genet.* **6**: 84-89.

- Ionov, Y., Peinado, M.A., Malkhosyan, S., Shibata, D., and Perucho, M. (1993) Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature* **363**: 558-561.
- Irminger-Finger, I., Siegel, B.D., and Leung, W.C. (1999) The functions of breast cancer susceptibility gene 1 (BRCA1) product and its associated proteins. *Biol. Chem.* **380**: 117-128.
- Janin, N., Andrieu, N., Ossian, K., Laugé, A., Croquette, M.F., Griscelli, C., et al. (1999) Breast cancer risk in ataxia-telangiectasia (A-T) heterozygotes: haplotype study in French A-T families. *Br. J. Cancer* **80**: 1042-1045.
- Jardines, L., Weiss, M., Fowble, B., and Greene, M. (1993) Neu (c-erbB2/HER2) and the epidermal growth factor receptor (EGFR) in breast cancer. *Pathobiology* **61**: 268-282.
- Jensen, R.J., Thomson, M.E., Jetton, T.L., Szabo, C.I., Van deer Meer, R., et al. (1996) BRCA1 is secreted and exhibits properties of a granin. *Nature Genet.* **12**: 303-308.
- Jin, Y., Xu, X.L., Yang, M.C., Wei, F., Ayi, T.C., Bowcock, A.M., and Baer, R. (1997) Cell cycle-dependent colocalization of BARD1 and BRCA1 protein in discrete nuclear domains. *Proc. Natl. Acad. Sci. U.S.A.* **94**: 12075-12080.
- Jordan, V.C., and Murphy, C.S. (1990) Endocrine pharmacology of antiestrogens as antitumor agents. *Endocr. Rev.* **11**: 578-610.
- Jouet, M., Rosenthal, A., Armstrong, G., MacFarlane, J., Stevenson, R., Paterson, J., Metzberg, A., Ionasescu, V., Temple, K., Kenwick, S. (1994) X-linked spastic paraplegia (SPG1), MASA syndrome and X-linked hydrocephalus result from mutations in the L1 gene. *Nature Genet.* **7**: 402-407
- Juliano, R.L., and Haskill, S. (1993) Signal transduction from the extracellular matrix. *J. Cell Biol.* **120**: 577-585.
- Karpf, A.R., and Jones, D.A. (2002) Reactivating the expression of methylation silenced genes in human cancer. *Oncogene* **21**:5496-5503.
- Kashiwaba, M., Tamura, G., Suzuki, Y., Maesawa, C., Ogasawara, S., Sakata, K., and Satodate, R. (1995) Epithelial-cadherin gene is not mutated in ductal carcinomas of the breast. *Jpn. J. Cancer Res.* **86**: 1054-1059.
- Kastan, M.B., Onyekwere, O., Sidransky, D., et al. (1991) Participation of p53 protein in the cellular response to DNA damage. *Cancer Res.* **51**: 6304-6311.
- Kaupmann, K., Becker-Follmann, J., Scherer, G., Jockusch, H., and Starzinski-Powitz, A. (1992) The gene for the cell adhesion molecule M-cadherin maps to mouse chromosome 8 and human chromosome 16q24.1-qter and is near the E-cadherin (uvomorulin) locus in both species. *Genomics* **14**: 488-490.
- Kazanietz, M.G. (2000) Eyes wide shut: protein kinase C isoenzymes are not only the receptors for the phorbol ester tumor promoters. *Mol. Carcinog.* **28**: 5-11.



- Kim, H., Jen, J., Vogelstein, B., and Hamilton, S.R. (1994) Clinical and pathological characteristics of sporadic colorectal carcinomas with DNA replication errors in microsatellite sequences. *Am. J. Pathol.* **145**: 148-156.
- Kinzler, K.W., and Vogelstein, B. (1996) Lessons from hereditary colorectal cancer. *Cell* **87**: 159-170.
- Klein, G. (1989) Multiple phenotypic consequences of the *Ig/Myc* translocation in B-cell derived tumors. *Genes Chromosomes Cancer* **1**: 3-8.
- Klenova, E.M., Nicolas, R.H., Paterson, H.F., Carne, A.F., Heath, C.M., Goodwin, G.H., Neiman, P.E., and Lobanenkova, V.V. (1993) *CTCF*, a conserved nuclear factor required for optimal transcriptional activity of the chicken *c-myc* gene, is an 11-Zn-finger protein differentially expressed in multiple forms. *Mol. Cell Biol.* **13**: 7612-7624.
- Knudson, A.G. (1971) Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl. Acad. Sci. U.S.A.* **68**: 820-823.
- Knudson, A.G. (1993) Antioncogenes and human cancer. *Proc. Natl. Acad. Sci. U.S.A.* **90**: 10914-10921.
- Knudson, A.G., Meadows, A.T., Nichols, W.W., and Hill, R. (1976) Chromosomal deletion and retinoblastoma. *N. Engl. J. Med.* **295**: 1120-1123.
- Kobayashi, H., Hosoda, F., Maseki, N., Sakurai, M., Imashuku, S., Ohki, M., and Kaneko, Y. (1997) Hematologic malignancies with the t(10;11)(p13;q21) have the same molecular event and a variety of morphologic or immunologic phenotypes. *Genes Chromosomes Cancer* **20**: 253-259.
- Konishi, M., Kikuchi-Yanoshita, R., Kiyoko, T., Magatoshi, M., Onda, A., Okumura, Y., Kishi, N., Iwama, T., Mori, T., Koike, M., Ushio, K., Chiba, M., Nomizu, S., Konishi, F., Utsunomiya, J., and Miyaki, M. (1996) Molecular nature of colon tumors in hereditary nonpolyposis colon cancer, familial polyposis and sporadic colon cancer. *Gastroenterology* **111**: 307-317.
- Koonin, E.V., Altschul, S.F., and Bork, P. (1996) BRCA1 protein products: functional motifs. *Nature Genet.* **13**: 266-267.
- Korchak, H.M., Kane, L.H., Rossi, M.W., and Corkey, B.E. (1994) Long chain acyl coenzyme A and signaling in neutrophils. An inhibitor of acyl coenzyme A synthetase, triacsin C, inhibits superoxide anion generation and degranulation by human neutrophils. *J. Biol. Chem.* **269**: 30281-30287.
- Korinek, V., Barker, N., Morin, P.J., van Wichen, D., de Weger, R., Kinzler, K.W., Vogelstein, B., and Clevers, H. (1997) Constitutive transcriptional activation by a  $\beta$ -catenin-Tcf complex in APC<sup>-/-</sup> colon carcinoma. *Science* **275**: 1784-1787.
- Korn, B., Sedlacek, Z., Manca, A., Kioschis, P., Konecki, D., Lehrach, H., and Poustka, A. (1992) A strategy for the selection of transcribed sequences in the Xq28 region. *Hum.*

*Mol. Genet.* **1**: 235-242.

- Koyama, K., Emi, M., and Nakamura, Y. (1993) The cell adhesion regulator (CAR) gene, TaqI and insertion/deletion polymorphisms, and regional assignment to the peritelomeric region of 16q by linkage analysis. *Genomics* **16**:264-265.
- Kozak, M. (1989) The scanning model for translation: an update. *J. Cell Biol.* **108**: 229-241.
- Kozak, M. (1991) Structural features in eukaryotic mRNAs that modulate the initiation of translation. *J. Biol. Chem.* **266**: 19867-19870.
- Kremmidiotis, G., Baker, E., Crawford, J., Eyre, H.J., Nahmias, J., and Callen, D.F. (1998) Localization of human cadherin genes to chromosome regions exhibiting cancer-related loss of heterozygosity. *Genomics* **49**: 467-471.
- Kunimi, K., Bergerheim, U., Larsson, I-L., Ekman, P., and Collins, V.P. (1991) Allelotyping of human prostatic adenocarcinoma. *Genomics* **11**: 530-536.
- Kurokawa, M., Lynch, K., and Podolsky, D.K. (1987) Effects of growth factors on an intestinal epithelial cell line: transforming growth factor beta inhibits proliferation and stimulates differentiation. *Biochem. Biophys. Res. Comm.* **142**: 775-782.
- Lagios, M.D., Margolin, F.R., Westdahl, P.R., and Rose, M.R. (1989) Mammographically detection duct carcinoma *in situ*. Frequency of local recurrence following tylectomy and prognostic effect of nuclear grade on local recurrence. *Cancer* **63**(4):618-624.
- Lai, J.C.K., Liang, B.B., Jarri, E.J., Cooper, A.J.L., and Lu, D.R. (1993) Differential effects of fatty acyl coenzyme A derivatives on citrate synthase and glutamate dehydrogenase. *Res. Commun. Chem. Pathol. Pharmacol.* **82**: 331-338.
- Lammie, G.A., and Peters, G. (1991) Chromosome 11q13 abnormalities in human cancer. *Cancer Cells* **3**: 413-420.
- Lancaster, J.M., Wooster, R., Mangion, J., Phelan, C.M., Cochran, C., et al. (1996) BRCA2 mutations in primary breast and ovarian cancers. *Nature Genet.* **13**: 238-240.
- Larsen, F., Gundersen, G., Lopez, R., and Prydz, H. (1992) CpG islands as gene markers in the human genome. *Genomics* **13**: 1095-1107.
- Latif, F., Tory, K., Gnarr, J., Yao, M., Duh, F.M., Orcutt, M.L., et al. (1993) Identification of the von Hippel-Lindau disease tumor suppressor gene. *Science* **260**: 1317-1320.
- Latil, A., Cussenot, O., Fournier, G., Driouch, K., and Lidereau, R. (1997) Loss of heterozygosity at chromosome 16q in prostate adenocarcinoma: identification of three independent regions. *Cancer Res.* **57**: 1058-1062.
- Le Douarin, B., Nielsen, A.L., Garnier, J.M., Ichinose, H., Jeanmougin, F., Losson, R., and Chambon, P. (1996) A possible involvement of TIF1 $\alpha$  and TIF1 $\beta$  in the epigenetic control of transcription by nuclear receptors. *EMBO J.* **15**: 6701-6715.

- Levine, A.J. (1993) The tumor suppressor genes. *Annu.Rev. Biochem.* **62**: 623-651.
- Lee, E.Y-H.P., To, H., Shew, J-Y., Bookstein, R., Scully, P., and Lee, W-H. (1988) Inactivation of the retinoblastoma susceptibility gene in human breast carcinoma. *Science* **241**: 218-221.
- Lee, H., Trainer, A.H., Friedman, L.S., Thistlethwaite, F.C., Evans, M.J., Ponder, B.A., Venkitaraman, A.R. (1999) Mitotic checkpoint inactivation fosters transformation in cells lacking the breast cancer susceptibility gene, *Brsa2*. *Mol. Cell* **4**: 1-10.
- Lee, J.S., Collins, K.M., Brown, A.L., Lee, C.H., Chung, J.H. (2000) hCds1-mediated phosphorylation of BRCA1 regulates the DNA damage response. *Nature* **404**: 201-204.
- Leonhard, K., Herrmann, J.M., Stuart, R.A., Mannhaupt, G., Neupert, W., and Langer, T. (1996) AAA proteases with catalytic sites on opposite membrane surfaces comprise a proteolytic system for the ATP-dependent degradation of inner membrane proteins in mitochondria. *EMBO J.* **15**: 4218-4229.
- Leonhard, K., Stiegler, A., Neupert, W., and Langer, T. (1999) Chaperone-like activity of the AAA domain of the yeast Yme1 AAA protease. *Nature* **398**: 348-351.
- Levy, D.B., Smith, K.J., Beazer-Barclay, Y., Hamilton, S.R., Vogelstein, B., and Kinzler, K.W. (1994) Inactivation of both APC alleles in human and mouse tumors. *Cancer Res.* **54**: 5953-5958.
- Li, D.M., and Sun, H. (1997) TEP1, encoded by a candidate tumor suppressor locus, is a novel protein tyrosine phosphatase regulated by transforming growth factor- $\beta$ . *Cancer Res.* **57**: 2124-2129.
- Li, F.P. and Fraumeni, J.F., Jr. (1975) Familial breast cancer, soft-tissue sarcomas, and other neoplasms. *Ann. Intern. Med.* **83**: 833-834.
- Li, F.P. and Fraumeni, J.F., Jr. (1982) Prospective study of a family cancer syndrome. *JAMA* **247**: 2692-2694.
- Li, F.P., Fraumeni, J.F., Jr., Mulvihill, J.J., Blattner, W.A., Dreyfuss, M.G., et al. (1988) A cancer family syndrome in twenty-four kindreds. *Cancer Res.* **48**: 5358-5362.
- Li, J., Yen, C., Liaw, D., Podsypanina, K., Bose, S., Wang, S., et al. (1997) *PTEN*, a putative protein tyrosine phosphatase gene mutated in human brain, breast and prostate cancer. *Science* **275**: 1943-1947.
- Li, S., Chen, P.L., Subramanian, T., Chinnadurai, G., Tomlinson, G., Osborne, C.K., Sharp, I.D., Lee, W.H. (1999) Binding of CtIP to the BRCT repeats of BRCA1 involved in the transcription regulation of *p21* is disrupted upon DNA damage. *J. Biol. Chem.* **274**: 11334-11338.
- Li, S., MacLachlan, T.K., De Luca, A., Claudio, P.P., Condorelli, G., and Giordano, A. (1995) The cdc2-related kinase, PISSLRE, is essential for cell growth and acts in G2 phase of the cell cycle. *Cancer Res.* **55**: 3992-3995.

- Liaw, D., Marsh, D.J., Li, J., Dahia, P.L., Wang, S.I., Zheng, Z., Bose, S., Call, K.M., Tsou, H.C., Peacocke, M., Eng, C., and Parsons, R. (1997) Germline mutations of the *PTEN* gene in Cowden disease, an inherited breast and thyroid cancer syndrome. *Nature Genet.* **16**: 64-67.
- Liehr, J.G. (2000) Is estradiol a genotoxic mutagenic carcinogen? *Endocr. Rev.* **21**: 40-54.
- Lichy, J.H., Dalbague, F., Zavar, M., Washington, C., Tsai, M.M., Sheng, Z-M., and Taubenberger, J.K. (2000) genetic heterogeneity in Ductal Carcinoma of the Breast. *Lab. Invest.* **80**: 291-301.
- Liu, B., Nicolaides, N.C., Markowitz, S., Willson, J.K.V., Parsons, R.E., Jen, J., Papadopolous, N., Peltomaki, P., de la Chapelle, A., Hamilton, S.R., Kinzler, K.W., and Vogelstein, B. (1995) Mismatch repair gene defects in sporadic colorectal cancers with microsatellite instability. *Nature Genet.* **9**: 48-55.
- Lobanenkov, V.V., Nicolas, R.H., Adler, V.V., Paterson, H., Klenova, E.M., Polotskaja, A.V., and Goodwin, G.H. (1990) A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5' flanking sequence of the chicken *c-myc* gene. *Oncogene* **5**: 1743-1753.
- Loewith, R., Meijer, M., Lees-Miller, S.P., Riabowol, K., and Young, D. (2000) Three yeast proteins related to the human candidate tumor suppressor p33(ING1) are associated with histone acetyltransferase activities. *Mol. Cell. Biol.* **20**: 3807-3816.
- Lou, H., Gagel, R.F., and Berget, S.M. (1996) An intron enhancer recognized by splicing factors activates polyadenylation. *Genes Dev.* **1**: 208-219.
- Lou, H., Neugebauer, K.M., Gagel, R.F., and Berget, S.M. (1998) Regulation of alternative polyadenylation by U1 snRNPs and SRp20. *Mol. Cell. Biol.* **18**: 4977-4985.
- Lou, H., Yang, Y., Cote, G.J., Berget, S.M., and Gagel, R.F. (1995) An intron splicing enhancer containing a 5' splice site sequence in the human calcitonin/calcitonin gene-related peptide gene. *Mol. Cell. Biol.* **15**: 3979-3988.
- 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. U.S.A.* **88**: 9628-9632.
- Luongo, C., Moser, A.R., Gledhill, S., and Dove, W.F. (1994) Loss of APC<sup>+</sup> in intestinal adenomas from Min mice. *Cancer Res.* **54**: 5947-5952.
- Lu, P.J., Sundquist, K., Baekstrom, D., Poulson, R., Hanby, A., Meier-Ewert, S., Jones, T., Mitchell, M., Pitha-Rowe, P., Freemont, P., and Taylor-Papadimitriou, J. (1999) A novel gene (PLU-1) containing highly conserved putative DNA/chromatin binding motifs is specifically up-regulated in breast cancer. *J. Biol. Chem.* **274**: 15633-15645.
- Lu, S.L., Akiyama, Y., Nagasaki, H., Saitoh, K., and Yuasa, Y. (1995) Mutations of the transforming growth factor- $\beta$  type II receptor gene and genomic instability in hereditary nonpolyposis colorectal cancer. *Biochem Biophys. Res. Commun.* **216**: 452-457.

- Lynch, H.T., Mulcahy, G.M., Harris, R.E., Guirgis, H.A., and Lynch, J.F. (1978) Genetic and pathologic findings in a kindred with hereditary sarcoma, breast cancer, brain tumors, leukemia, lung, laryngeal, and adrenal cortical carcinoma. *Cancer* **41**: 2055-2064.
- MacGrogan, D., Pergram, M., Slamon, D., and Bookstein, R. (1997) Comparative mutational analysis of DPC4 (Smad4) in prostatic and colorectal carcinomas. *Oncogene* **15**: 1111-1114.
- Magdinier, F., Ribieras, S., Lenoir, G.M., Frappart, L., and Dante, R. (1998) Down-regulation of BRCA1 in human sporadic breast cancer: analysis of DNA methylation patterns of the putative promoter region. *Oncogene* **17**: 3169-3176.
- Malkin, D., Li, F.P., Strong, L.C., Fraumeni, J.F., Jr., Nelson, C.E., Kim, D.H., et al. (1990) Germline *p53* mutations in a familial syndrome of breast cancer, sarcomas and other neoplasms. *Science* **250**: 1233-1238.
- Mancini, D.N., Rodenhiser, D.I., Ainsworth, P.J., O'Malley, F.P., Singh, S.M., Xing, W., and Archer, T.K. (1998) CpG methylation within the 5' regulatory region of the *BRCA1* gene is tumor specific and includes a putative CREB binding site. *Oncogene* **16**: 1161-1169.
- Mann, B., Gelos, M., Siedow, A., Hanski, M.L., Gratchev, A., Ilyas, M., Bodmer, W.F., Moyer, M.P., Riecken, E.O., Buhr, H.J., and Hanski, C. (1999) Target genes of beta-catenin-T cell factor/Lymphoid-enhancer factor signalling in human colorectal carcinomas. *Proc. Natl. Acad. Sci. U.S.A.* **96**: 1603-1608.
- Mansouri, M., Spurr, N., Goodfellow, P.N., and Kemler, R. (1988) Characterization and chromosomal localization of the gene encoding the human cell adhesion molecule uvomorulin. *Differentiation* **38**: 67-71.
- Markowitz, S., and Roberts, A. (1996) Tumor suppressor activity of the TGF- $\beta$  pathway in human cancers. *Cytokine Growth Factor Rev.* **7**: 93-102.
- Markowitz, S., Wang, J., Myeroff, L., Parsons, R., Sun, L., Lutterbaugh, J., Fan, R., Zborowska, E., Kinzler, K., Vogelstein, B., Brattain, M., and Willson, J. (1995) Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. *Science* **268**: 1336-1338.
- Matise, T.C., Perlin, M., and Chakravarti, A. (1994) Automated construction of genetic linkage map using an expert system (MultiMap): a human genome linkage map. *Nature Genet.* **6**: 384-390.
- Matsuoka, S., Huang, M., and Elledge, S.J. (1998) Linkage of ATM to cell cycle regulation by the Chk2 protein kinase. *Science* **282**: 1893-1897.
- Maw, M.A., Grundy, P.E., Millow, L.J., Eccles, M.R., Dunn, R.S., Smith, P.J., Feinberg, A.P., Law, D.J., Paterson, M.C., Telzerow, P.E., Callen, D.F., Thompson, A.D., Richards, R.I., and Reeve, A.E. (1992) A third Wilms' tumor locus on chromosome 16q. *Cancer Res.* **52**: 3094-3098.

- McDermott, C.J., Dayaratne, R.K., Tomkins, J., Lusher, M.E., Lindsey, J.C., Johnson, M.A., Casari, G., Turnbull, D.M., Bushby, K., and Shaw, P.J. (2001) Paraplegin gene analysis in hereditary spastic paraparesis (HSP) pedigrees innortheast England. *Neurology* **56**: 467-471.
- McKusick, V.A. (1998) *Mendelian Inheritance in Man*. **12th ed.**, Baltimore: Johns Hopkins University Press.
- McLaughlin, S., and Aderem, A. (1995) The myristoyl-electrostatic switch: a modulator of reversible protein-membraneinteractions. *Trends. Biochem. Sci.* **20**: 272-276.
- Meloni, I., Muscettola, M., Raynaud, M., Longo, I., Bruttini, M., Moizard, M.P., Gomot, M., Chelly, J., des Portes, V., Fryns, J.P., Ropers, H.H., Magi, B., Bellan, C., Volpi, N., Yntema, H.G., Lewis, S.E., Schaffer, J.E., and Renieri, A. (2002) *FACL4*, encoding fatty acid-CoA ligase 4, is mutated in nonspecific X-linked mental retardation. *Nature Genet.* **30**: 436-440.
- Miki, Y., Swensen, J., Shattuck Eidens, D., Futreal, P.A., Harshman, K., et al. (1994) A strong candidate for the breast and ovarian cancer susceptibility gene *BRCA1*. *Science* **266**: 66-71.
- Miller, J.R., and Moon, R.T. (1996) Signal transduction through beta-catenin and specification of cell fate during embryogenesis. *Genes Dev.* **10**: 2527-2539.
- Minekura, H., Fujino, T., Kang, M.-J., Fujita, T., Endo, Y., and Yamamoto, T.T. (1997) Human acyl-coenzyme A synthetase 3 cDNA and localization of its gene (*ACS3*) to chromosome band 2q34-q35. *Genomics* **42**: 180-181.
- Minoshima, S., Fukuyama, R., Yamamoto, T., and Shimizu, N. (1991) Mapping of human long-chain acyl-CoA synthetase to chromosome 4. (Abstract) *Cytogenet. Cell Genet.* **58**: 1888.
- Mitelman, F., Martens, F., and Johansson, B. (1997) A breakpoint map of recurrent chromosomal rearrangements in human neoplasia. *Nature Genet.* **15**: 417-474.
- Miyaki, M., Iijima, T., Kimura, J., Yasuno, M., Mori, T., Hayashi, Y., Koike, M., Shitara, N., Iwama, T., and Kuroki, T. (1999a) Frequent mutation of beta-catenin and *APC* genes in primary colorectal tumors from patients with hereditary nonpolyposis colorectal cancer. *Cancer Res.* **59**: 4506-4509.
- Miyaki, M., Iijima, T., Konishi, M., Sakai, K., Ishii, A., Yasuno, M., Hishima, T., Koike, M., Shitara, N., Iwama, T., Utsunomiya, J., Kuroki, T., and Mori, T. (1999b) Higher frequency of *Smad4* gene mutation in human colorectal cancer with distant metastasis. *Oncogene* **18**: 3098-3103.
- Miyashita, T., and Reed, J.C. (1995) Tumor suppressor p53 is a direct transcriptional activator of the human *bax* gene. *Cell* **80**: 293-299.
- Miyoshi, Y., Nagase, H., Ando, H., Horii, A., Ichii, S., Nakatsuru, S., Aoki, T., Miki, Y., Mori, T., and Nakamura, Y. (1992) Somatic mutations of the *APC* gene in colorectal

- tumors: mutation cluster region in the *APC* gene. *Hum. Mol. Genet.* **1**: 229-233.
- Moerland, E., Breuning, M.H., Cornelisse, C.J., and Cleton-Jansen A-M., (1997) Exclusion of BBC1 and CMAR as candidate breast tumor-suppressor genes. *Br. J. Cancer* **76**: 1550-1553.
- Morgan, J.G., Dolganov, G.M., Robbins, S.E., Hinton, L.M., and Lovett, M. (1992) The selective isolation of novel cDNAs encoded by the regions surrounding the human interleukin 4 and 5 genes. *Nucleic Acids Res.* **20**: 5173-5179.
- Morin, P.J., Sparks, AB., Korinek, V., Barker, N., Clevers, H., Vogelstein, B., and Kinzler, K.W. (1997) Activation of  $\beta$ -catenin-Tcf signalling in colon cancer by mutation in  $\beta$ -catenin or APC. *Science* **275**: 1787-1790.
- Muller, W.J., Sinn, E., Pattengale, P.K., Wallace, R., and Leder, P. (1988) Single-step induction of mammary adenocarcinoma in transgenic mice bearing the activated *c-neu* oncogene. *Cell* **54**: 105-115.
- Murillo, F.M., Kobayashi, H., Pegoraro, E., Galluzzi, G., Creel, G., Mariani, C., Farina, E., Ricci, E., Alfonso, G., Pauli, R.M., and Hoffman, E.P. (1999) Genetic localization of a new locus for recessive familial spastic paraparesis to 15q13-15. *Neurology* **53**: 50-56.
- Musgrove, E.A., Hamilton, J.A., Lee, C.S., Sweeney, K.J., Watts, C.K., and Sutherland, R.L. (1993) Growth factor, steroid, and steroid antagonist regulation of cyclin gene expression associated with changes in T-47D human breast cancer cell cycle progression. *Mol. Cell Biol.* **13**: 3577-3587.
- Nagamine, K., Peterson, P., Scott, H.S., Kudoh, J., Minoshima, S., Heino, M., Krohn, K.J., Lalioti, M.D., Mullis, P.E., Antonarakis, S.E., Kawasaki, K., Asakawa, S., Ito, F., and Shimizu, N. (1997) Positional cloning of the APECED gene. *Nature Genet.* **17**: 393-398.
- Nagase, T., Ishikawa, K., Suyama, M., Kikuno, R., Hirosawa, M., Miyajima, N., Tanaka, A., Kotani, H., Nomura, N., and Ohara, O. (1998) Prediction of the coding sequences of unidentified human genes. XII. The complete sequences of 100 new cDNA clones from brain which code for large proteins *in vitro*. *DNA Res.* **5**: 355-364.
- Nakai, K., and Horton, P. (1999) PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem. Sci.* **24**: 34-36.
- Nakamura, T., Alder, H., Gu, Y., Prasad, R., Canaani, O., *et al.* (1993) Genes on chromosome 4, 9, and 19 involved in 11q23 abnormalities in acute leukemia share sequence homology and/or common motifs. *Proc. Natl. Acad. Sci. U.S.A.* **90**: 4631-4635.
- Narod, S., Feunteun, J., Lynch, H., *et al.* (1991) Familial breast-ovarian cancer locus on chromosome 17q12-23. *Lancet* **338**: 82-83.
- Nataraj, A.J., Olivos-Glander, I., Kusukawa, N., and Highsmith, W.E.J. (1999) Single strand conformation polymorphism and heteroduplex analysis for gel based mutation detection. *Electrophoresis* **20**: 1177-1185.

- Nelen, M.R., Padberg, G.W., Peeters, E.A.J., Lin, A.Y., van den Helm, B., Frants, R.R., et al. (1996) Localization of the gene for Cowden disease to 10q22-23. *Nature Genet.* **13**: 114-116.
- Neuhausen, S., Gilewski, T., Norton, L., Tran, T., McGuire, P., et al. (1996) Recurrent *BRCA2* 6174delT mutation in Ashkenazi women affected by breast cancer. *Nature Genet.* **13**: 126-128.
- Newman, B., Austin, M.A., Lee, M., King, M-C. (1988) Inheritance of human breast cancer: Evidence for autosomal dominant transmission in high-risk families. *Proc. Natl. Acad. Sci. U.S.A.* **85**: 3044-3048.
- NIH/CEPH Collaborative Mapping Group. (1992) A comprehensive genetic linkage map of the human chromosome. *Science* **258**: 67-86.
- Nimchuk, Z., Marois, E., Kjemtrup, S., Leister, R.T., Katagiri, F., and Dangl, J.L. (2000) Eukaryotic fatty acylation drives plasma membrane targeting and enhances function of several type III effector proteins from *Pseudomonas syringae*. *Cell* **101**: 353-363.
- Nishisho, I., Nakamura, Y., Miyoshi, Y., Miki, Y., Ando, H., Horii, A., et al. (1991) Mutation of chromosome 5q21 genes in FAP and colorectal cancer patients. *Science* **253**: 665-669.
- Nobile, C., Hinzmann, B., Scannapieco, P., Siebert, R., Zimbello, R., Perez-Tur, J., Sarafidou, T., Moschonas, N.K., French, L., Deloukas, P., Ciccodicola, A., Gesk, S., Poza, J.J., Lo Nigro, C., Seri, M., Schlegelberger, B., Rosenthal, A., Valle, G., Lopez de Munain, A., Tassinari, C.A., and Michelucci, R. (2002) Identification and characterization of a novel human brain-specific gene, homologous to *S. scrofa* tmp83.5, in the chromosome 10q24 critical region for temporal lobe epilepsy and spastic paraplegia. *Gene* **282**: 87-94.
- Novelli, M.R., Durbin, H., and Bodmer, W.F. (1995) in *Topics in Molecular Medicine*, eds. Seiss, W., Lorenz, R., and Weber, P.C. (Raven, New York), Vol. 1, pp. 179-186<!-- HIGHWIRE ID="94:26:14578:4" --><!-- /HIGHWIRE -->.
- Nowell, P.C., and Hungerford, D.A. (1960) A minute chromosome in human chronic granulocytic leukemia. *Science* **132**: 1497-1499.
- Offit, K., Gilewski, T., McGuire, P., Schluger, A., Hampel, H., et al. (1996) Germline *BRCA1* 185delAG mutations in Jewish women with breast cancer. *Lancet* **347**: 1643-1644.
- Oltvai, Z.N., Milliman, C.L., and Korsmeyer, S.J. (1993) Bcl-2 heterodimerizes *in vivo* with a conserved homolog, Bax, that accelerates programmed cell death. *Cell* **74**: 609-619.
- Ono, Y., Fujii, T., Igarashi, K., Kuno, T., Tanaka, C., Kikkawa, U., and Nishizuka, K. (1989) Phorbol ester binding to protein kinase C requires a cysteine-rich zinc-finger-like sequence. *Proc. Natl. Acad. Sci. U.S.A.* **86**: 4868-4871.
- Onyango, P., Lubyova, B., Gardellin, P., Kurzbauer, R., and Weith, A. (1998) Molecular cloning and expression analysis of five novel genes in chromosome 1p36. *Genomics* **50**: 187-198.



- Orth, M., and Schapira, A.H. (2001) Mitochondria and degenerative disorders. *Am. J. Med. Genet.* **106**: 27-36.
- Osman, I., Scher, H., Dalbagni, G., Reuter, V., Zhang, Z.F., and Cordon-Cardo, C. (1997) Chromosome 16 in primary prostate cancer: a microsatellite analysis. *Int. J. Cancer* **71**: 580-584.
- Pandey, A., Andersen, J.S., and Mann, M. (2000) Use of mass spectrometry to study signaling pathways. *Sci STKE*. **2000**(37): PL1.
- Pandey, A., and Mann, M. (2000) Proteomics to study genes and genomes. *Nature*. **405**: 837-846.
- Pandey, A., and Lewitter, F. (1999) Nucleotide sequence databases: a gold mine for biologists. *Trends Biochem. Sci.* **24**: 276-280.
- Papkoff, J., Rubinfeld, B., Schryver, B., and Polakis, P. (1996) Wnt-1 regulates free pools of catenins and stabilizes APC-catenin complexes. *Mol. Cell. Biol.* **16**: 2128-2134.
- 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. U.S.A.* **88**: 9623-9627.
- Parry, P., Djabali, M., Bower, M., Khristich, J., Waterman, M., *et al.* (1993) Structure and expression of the human trithorax-like gene 1 involved in acute leukemias. *Proc. Natl. Acad. Sci. U.S.A.* **90**: 4738-4742.
- Parsons, R., Li, G.M., Longley, M.J., Fang, W.H., Papadopoulos, N., Jen, J., de la Chapelle, A., Kinzler, K.W., Vogelstein, B., and Modrich, P. (1993) Hypermutability and mismatch repair deficiency in RER+ tumor cells. *Cell* **75**: 1227-1236.
- Parsons, R., Myeroff, L.L., Liu, B., Willson, J.K., Markowitz, S.D., Kinzler, K.W., and Vogelstein, B. (1995) Microsatellite instability and mutations of the transforming growth factor beta type II receptor gene in colorectal cancer. *Cancer Res.* **55**: 5548-5550.
- Patel, S., and Latterich, M. (1998) The AAA team: related ATPases with diverse functions. *Trends Cell Biol.* **8**: 65-71.
- Peifer, M. (1997)  $\beta$ -catenin as oncogene: The smoking gun. *Science* **275**: 1752-1753.
- Peinado, M.A., Malkhosyan, S., Velazquez, A., and Perucho, M. (1992) Isolation and characterization of allelic loss and gains in colorectal tumors arbitrarily primed polymerase chain reaction. *Proc. Natl. Acad. Sci. U.S.A.* **89**: 10065-10069.
- Peltomaki, P., and de la Chapelle, A. (1997) Mutations predisposing to hereditary nonpolyposis colorectal cancer. *Adv. Cancer Res.* **71**: 93-119.
- Peterson, A., Patli, N., Robbins, C., Wang, L., Cox, D.R., and Myers, R.M. (1994) A transcript map of the Down syndrome critical region on chromosome 21. *Hum. Mol. Genet.* **3**: 1735-1742.

- Pfanner N., Orci, L., Glick, B.S., Amherdt, M., Ardern, S.R., Malhotra, V., and Rothman, J.E. (1989) Fatty acyl-coenzyme A is required for budding of transport vesicles from Golgi cisternae. *Cell* **59**: 95-102.
- Pfeifer, S.L., Herzog, T.J., Tribune, D.J., Mutch, D.G., Gersell, D.J., and Goodfellow, P.J.(1995) Allelic loss of sequence from the long arm of chromosome 10 and replication errors in endometrial cancers. *Cancer Res.* **55**: 1922-1926.
- Piccini, M., Vitelli, F., Bruttini, M., Pober, B.R., Jonsson, J.J., Villanova, M., Zollo, M., Borsani, G., Ballabio, A., and Renieri, A. (1998) *FACLA*, a new gene encoding long chain acyl-CoA synthetase 4, is deleted in a family with Alport syndrome, elliptocytosis, and mental retardation. *Genomics* **47**: 350-358.
- Pierceall, W., Woodard, A., Morrow, J., Rimm, D., and Fearon, E. (1995) Frequent alterations in E-cadherin and  $\alpha$ - and  $\beta$ -catenin expression in human breast cancer cell lines. *Oncogene* **11**: 1319-1326.
- Plass, C. (2002) Cancer epigenomics. *Hum. Mol. Genet.* **11**: 2479-2488.
- Powell, S.M., Zilz, N., Beazer-Barclay, Y., Bryan, T.M., Hamilton, S.R., Thibodeau, S.N., Vogelstein, B., and Kinzler, K.W. (1992) *APC* mutations occur early during colorectal tumorigenesis. *Nature* **359**: 235-237.
- Pras, E., Aksentijevich, I., Gruberg, L., Balow, J. E., Jr., Prosen, L., Dean, M., Steinberg, A. D., Pras, M., and Kastner, D. L. (1992) Mapping of a gene causing familial Mediterranean fever to the short arm of chromosome 16. *New Eng. J. Med.* **326**: 1509-1513.
- Pronk, J. C., Gibson, R. A., Savoia, A., Wijker, M., Morgan, N. V., Melchionda, S., Ford, D., Temtamy, S., Ortega, J. J., Jansen, S., Havenga, C., Cohn, R. J., de Ravel, T. J., Roberts, I., Westerveld, A., Easton, D. F., Joenje, H., Mathew, C. G., and Arwert, F. (1995) Localisation of the Fanconi anaemia complementation group A gene to chromosome 16q24.3. *Nature Genet.* **11**: 338-340.
- Pullman, W.E., and Bodmer, W.F. (1992) Cloning and characterization of a gene that regulates cell adhesion. *Nature* **356**: 529-532.
- Rabbits, T.H. (1994) Chromosomal translocations in human cancer. *Nature* **372**: 143-149.
- Radford, D.M., Fair, K.L., Phillips, N.J., Ritter, J.H., Steinbrueck, T., Holt, M.S., and Donis-Keller, H. (1995) Allelotyping of ductal carcinoma *in situ* of the breast: deletion of loci on 18p, 13q, 16q, 17p and 17q. *Cancer Res.* **55**: 3399-3405.
- Raman, N., Black, P.N., and DiRusso, C.C. (1997) Characterization of the fatty acid-responsive transcription factor FadR. Biochemical and genetic analyses of the native conformation and functional domains. *J. Biol. Chem.* **272**: 30645-30650.
- Rampino, N., Yamamoto, H., Ionov, Y., Li, Y., Sawai, H., Reed, J.C., and Perucho, M. (1997) Somatic frameshift mutations in the BAX gene in colon cancers of the microsatellite

mutator phenotype. *Science* **275**: 967-969.

- Rasheed, B.K.A., McLendon, R.E., Friedman, H.S., Friedman, A.H., Fuchs, H.E., Bigner, D.D., and Bigner, S.H. (1995) Chromosome 10 deletion mapping in human glioma: a common deletion region in 10q25. *Oncogene* **10**: 2243-2246.
- Ray, P., Zhang, D.-H., Elias, J.A., and Ray, A. (1995) Cloning of a differentially expressed I-kappa-B-related protein. *J. Biol. Chem.* **270**: 10680-10685.
- Reid, E. (1999a) The hereditary spastic paraplegias. *J. Neurol.* **246**: 995-1003.
- Reid, E., Dearlove, A.M., Osborn, O., Rogers, M.T., and Rubinsztein, D.C. (2000) A locus for autosomal dominant "pure" hereditary spastic paraplegia maps to chromosome 19q13. *Am. J. Hum. Genet.* **66**: 728-732.
- Reid, E., Dearlove, A.M., Rhodes, M., and Rubinsztein, D.C. (1999b) A new locus for autosomal dominant "pure" hereditary spastic paraplegia mapping to chromosome 12q13, and evidence for further genetic heterogeneity. *Am. J. Hum. Genet.* **65**: 757-763.
- Rep, M., and Grivell, L.A. (1996) *MBA1* encodes a mitochondrial membrane-associated protein required for biogenesis of the respiratory chain. *FEBS Lett.* **388**: 185-8
- Rice, J.C., Massey-Brown, K.S., and Futscher, B.W. (1998) Aberrant methylation of the *BRCA1* CpG island promoter is associated with decreased *BRCA1* mRNA in sporadic breast cancer cells. *Oncogene* **17**: 1807-1812.
- Riggins, G.J., Kinzler, K.W., Vogelstein, B., and Thiagalingam, S. (1997) Frequency of Smad gene mutations in human cancers. *Cancer Res.* **57**: 2578-2580.
- Riggins, G.J., Thiagalingam, S., Rozenblum, E., Weinstein, C.L., Kern, S.E., Hamilton, S.R., Willson, J.K., Markowitz, S.D., Kinzler, K.W., and Vogelstein, B. (1996) Mad-related genes in the human. *Nature Genet.* **13**: 347-349.
- Rinderle, C., Christensen, H.M., Schweiger, S., Lehrach, H., and Yaspo, M.L. (1999) *AIRE* encodes a nuclear protein co-localizing with cytoskeletal filaments: altered sub-cellular distribution of mutants lacking the PHD zinc fingers. *Hum. Mol. Genet.* **8**: 277-290.
- Romer, L.H., Burridge, K., and Turner, C.E. (1992) Signaling between the extracellular matrix and the cytoskeleton: tyrosine phosphorylation and focal adhesion assembly. *Cold Spring Harb. Symp. Quant. Biol.* **57**: 193-202.
- Rommens, J.M., Jannuzzi, M.C., Kerem, B-S., Drumm, M.L., Melmer, G., Dean, M., Rozmahel, R., Cole, J.L., Kennedy, D., Hidaka, N., Zsiga, M., Buchwald, M., Riordan, J.R., Tsui, L-P., and Collins, F.S. (1989) Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science* **245**: 1059-1065.
- Rommens, J.M., Lin, B., Hutchinson, G.B., Andrew, S.E., Goldberg, Y.P., Graves, M.L., Graham, R., Lai, V., McArthur, J., Nasir, J., Theilmann, J., McDonald, H., Kalchman, M., Clarke, L.A., Schappert, K., and Hayden, M.R. (1993) A transcription map of the region containing the Huntington disease gene. *Hum. Mol. Genet.* **2**: 901-907.

- Rosenfeld, M.G., Amara, S.G., and Evans, R.M. (1984) Alternative RNA processing: determining neuronal phenotype. *Science* **225**: 1315-1320.
- Rowley, J.D. (1973) A new consistent chromosomal abnormality in chronic myelogenous leukemia identified by quinacrine fluorescence and Giemsa staining. *Nature* **243**: 290-293.
- Rubinfeld, B., Albert, I., Porfiri, E., Fiol, C., Munemitsu, S., and Polakis, P. (1996) Binding of GSK3 $\beta$  to the APC- $\beta$ -catenin complex and regulation of complex assembly. *Science* **272**: 1023-1026.
- Rubinfeld, B., Robbins, P., El-Gamil, M., Albert, I., Porfiri, E., and Polakis, P. (1997) Stabilization of beta-catenin by genetic defects in melanoma cell lines. *Science* **275**: 1790-1792.
- Rubinfeld, B., Souza, B., Albert, I., Muller, O., Chamberlain, S.H., Masiarz, F.R., Munemitsu, S., and Polakis, P. (1993) Association of the APC gene product with  $\beta$ -catenin. *Science* **262**: 1731-1734.
- Ruffner, H., Jiang, W., Craig, A.G., Hunter, T., and Verma, I.M. (1999) BRCA1 is phosphorylated at Serine 1497 in vivo at a cyclin-dependent kinase 2 phosphorylation site. *Mol. Cell. Biol.* **19**: 4843-4854.
- Rutqvist, L.E., and Wallgren, A. (1985) Long-term survival of 458 young breast cancer patients. *Cancer* **55**: 658-665.
- Ryan, K.M., and Birnie, G.D. (1996) *Myc* oncogenes: the enigmatic family. *Biochem. J.* **314**: 713-721.
- Saha, V., Chaplin, T., Gregorini, A., Ayton, P., and Young, B.D. (1995) The leukemia-associated-protein (LAP) domain, a cysteine-rich motif, is present in a wide range of proteins, including MLL, AF10, and MLLT6 proteins. *Proc. Natl. Acad. Sci. U.S.A.* **92**: 9737-9741.
- Sambrook, J., Frisch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
- Samowitz, W.S., Powers, M.D., Spirio, L.N., Nollet, F., Van Roy, F., and Slattery, M.L. (1999) Beta-catenin mutations are more frequent in small colorectal adenomas than in larger adenomas and invasive carcinomas. *Cancer Res.* **59**: 1442-1444.
- Sánchez-García, I. (1997) Consequences of chromosomal abnormalities in tumor development. *Ann. Rev. Genet.* **31**: 429-453.
- Sato, T., Akiyama, F., Sakamoto, G., Kasumi, F., and Nakamura, Y. (1991a) Accumulation of genetic alterations and progression of primary breast cancer. *Cancer Res.* **51**: 5794-5799.
- Sato, T., Saito, H., Morita, R., Koi, S., Lee, J.H., and Nakamura, Y. (1991b) Allelotype of human ovarian cancer. *Cancer Res.* **51**: 5118-5122.

- Saugier-veber, P., Munnich, A., Bonneau, D., Rozet, J.M., LeMerrer, M., Gil, R., and Boespflug-Tanguy, O. (1994) X-linked spastic paraplegia and Pelizaeus -Merzbacher disease are allelic disorders at the proteolipid protein locus. *Nature Genet.* **6**: 257-262.
- Saurin, A.J., Borden, K.L., Boddy, M.N., Freemont, P.S. (1996) Does this have a familiar RING? *Trends Biochem. Sci.* **21**: 208-214.
- Savino, M., d'Apolito, M., Centra, M., van Beerendonk, H.M., Cleton-Jansen, A-M., Whitmore, S.A., Crawford, J., Callen, D.F., Zelante, L., and Savoia, A. (1999) Characterization of Copine VII, a new member of the copine family, and its exclusion as a candidate in sporadic breast cancer with loss of heterozygosity at 16q24.3. *Genomics* **61**: 219-226.
- Savitsky, K., Bar-Shira, A., Gilad, S., Rotman, G., Ziv, Y., Vanagaite, L., et al. (1995) A single ataxia telangiectasia gene with a product similar to PI-3 kinase. *Science* **268**: 1749-1753
- Schildkraut, J.M., Rich, N., Thomson, W.D. (1989) Evaluating genetic association among ovarian, breast, and endometrial cancer: Evidence for a breast/ovarian cancer relationship. *Am. J. Hum. Genet.* **45**: 521-529.
- Schultz, J., Copley, R.R., Doerks, T., Ponting, C.P., and Bork, P (2000) SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Res.* **28**: 231-234.
- Schuuring, E., Verhoeven, E., van Tinteren, H., Peterse, J.L., Nunnik, B., Thunnissen, F.B.J.M., Devilee, P., Cornelisse, C.J., van de Vijver, M.J., Mooi, W.J., and Michalides, R.J.A.M. (1992) Amplification of genes within the chromosome 11q13 region is indicative of poor prognosis in patients with operable breast cancer. *Cancer Res.* **52**: 5229-5234.
- Scott, H.S., Heino, M., Peterson, P., Mittaz, L., Lalioti, M.D., Betterle, C., Cohen, A., Seri, M., Lerone, M., Romeo, G., Collin, P., Salo, M., Metcalfe, R., Weetman, A., Papanavvas, M.P., Rossier, C., Nagamine, K., Kudoh, J., Shimizu, N., Krohn, K.J., and Antonarakis, S.E. (1998) Common mutations in autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy patients of different origins. *Mol. Endocrinol.* **12**: 1112-1119.
- Scully, R., Ganesan, S., Brown, M., De Caprio, J.A., Cannistra, S.A., Feunteun, J., Schnitt, S., Livingston, D.M. (1996) Location of *BRCA1* in human breast and ovarian cancer cells. *Science* **272**: 123-126.
- Scully, R., Chen, J., Plug, A., Xiao, Y., Weaver, D., Feunteun, J., Ashley, T., and Livingston, D.M. (1997a) Association of *BRCA1* with Rad51 in mitotic and meiotic cells. *Cell* **88**: 265-275.
- Scully, R., Chen, J.J., Ochs, R.L., Keegan, K., Hoekstra, M., Feunteun, J., and Livingston, D.M. (1997b) Dynamic changes of *BRCA1* subnuclear location and phosphorylation state are initiated by DNA damage. *Cell* **90**: 425-435.
- Sedgwick, S.G., and Smerdon, S.J. (1999) The ankyrin repeat : a diversity of interactions on a

common structural framework. *Trends Biochem. Sci.* **24**:311-316.

- Seery, L.T., Knowlden, J.M., Gee, J.M., Robertson, J.F., Kenny, F.S., Ellis, I.O., and Nicholson, R.I. (1999) BRCA1 expression levels predict distant metastasis of sporadic breast cancers. *Int. J. Cancer* **84**: 258-262.
- Selleri, L., Eubanks, J.H., Giovannini, M., Hermanson, G.G., Romo, A., Djabali, M., Maurer, S., McElligott, D.L., Smith, M.W., and Evans, G.A. (1992) Detection and characterization of "chimeric" yeast artificial chromosome clones by fluorescent in situ suppression hybridization. *Genomics* **14**: 536-541.
- Seri, M., Cusano, R., Forabosco, P., Cinti, R., Caroli, F., Picco, P., Bini, R., *et al.* (1999) Genetic mapping to 10q23.3-q24.2, in a large Italian pedigree, of a new syndrome showing bilateral cataracts, gastroesophageal reflux, and spastic paraparesis with amyotrophy. *Am. J. Hum. Genet.* **64**: 586-593.
- Shafman, T., Khanna, K.K., Kedar, P., Spring, K., Kozlov, S., Yen, T., *et al.* (1997) Interaction between ATM protein and c-Abl in response to DNA damage. *Nature* **387**: 520-523.
- Shah, Z.H., Hakkaart, G.A.J., Arku, B., de Jong, L., van der Spek, H., Grivell, L.A. and Jacobs, H.T. (2000) The human homologue of the yeast mitochondrial AAA metalloprotease Yme1p complements a yeast yme1 disruptant. *FEBS Lett.* **478**: 267-270.
- Shapiro, M.B., and Senapathy, P. (1987) RNA splice junctions of different classes of eukaryotes: Sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* **15**: 7155-7174.
- Shattuck-Eidens, D., McClure, M., Simard, J., Labrie, F., Narod, S., Couch, F., *et al.* (1995) A collaborative survey of 80 mutations in the *BRCA1* breast and ovarian cancer susceptibility gene. Implications for presymptomatic testing and screening. *JAMA* **273**: 535-541.
- Shibata, D., Peinado, M.A., Ionov, Y., Malkhosyan, S., and Perucho, M. (1994) Genomic instability in repeated sequences is an early somatic event in colorectal tumorigenesis that persists after transformation. *Nature Genet.* **6**: 273-281.
- Shinohara, A., Ogawa, H., and Ogawa, T. (1992) Rad51 protein involved in repair and recombination in *S. cerevisiae* is a RecA-like protein. *Cell* **69**: 457-470.
- Shiu, R.P.C., Watson, P.H., and Dubik, D. (1993) *c-Myc* oncogene expression in estrogen-dependent and -independent breast cancer. *Clin. Chem.* **39**: 353-355.
- Shizuya, H., Birren, B., Kim, U.J., Mancino, V., Slepak, T., Tachiiri, Y., and Simon, M. (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci. U.S.A.* **89**: 8794-8797.
- Shrago, E., Woldegiorgis, G., Ruoho, A.E., and DiRusso, C.C. (1995) Fatty acyl CoA esters as regulators of cell metabolism. *Prostaglandins Leukotriens Essent. Fatty Acids* **52**: 163-

- Sicinski, P., Donaher, J.L., Parker, S.B., Li, T., Fazeli, A., Gardner, H., Haslam, S.Z., Bronson, R.T., Elledge, S.J., and Weinberg, R.A. (1995) Cyclin D1 provides a link between development and oncogenes in the retina and breast. *Cell* **82**: 621-630.
- Sidransky, D., Tokino, T., Helzlsouer, K., Zehnbauser, B., Rausch, G., et al. (1992) Inherited *p53* gene mutations in breast cancer. *Cancer Res.* **52**: 2984-2986.
- Skirnisdottir, S., Eiriksdottir, G., Baldursson, T., Barkardottir, R.B., Egilsson, V., and Ingvarsson, S. (1995) High frequency of allelic imbalance at chromosome region 16q22-23 in human breast cancer: correlation with high PgR and low S phase. *Int. J. Cancer (Pred. Oncol.)* **64**: 112-116.
- Slamon, D.J., Clark, G.M., Wong, S.G., Levin, W.S., Ullrich, A., and McGuire, W.L. (1987) Human breast cancer: Correlation of relapse and survival with amplification of the *HER-2/neu* oncogene. *Science* **235**: 177-182.
- Sparks, A.B., Morin, P.J., Vogelstein, B., and Kinzler, K.W. (1998) Mutational analysis of the APC/ $\beta$ -catenin/Tcf pathway in colorectal cancer. *Cancer Res.* **58**: 1130-1134.
- Srivastava, S., Zou, Z.Q., Pirolo, K., Blattner, W., and Chang, E.H. (1990) Germ-line transmission of mutated *p53* gene in a cancer-prone family with Li-Fraumeni syndrome. *Nature* **348**: 747-749.
- Stanczak, H., Stanczak, J.J., and Singh, I. (1992) Chromosomal localization of the human gene for palmitoyl-CoA ligase (FACL1). *Cytogenet. Cell Genet.* **59**: 17-19.
- Stec, I., Wright, T.J., van Ommen, G-J.B., de Boer, P.A.J., van Haeringen, A., Moorman, A.F.M., Altherr, M.R., and den Dunnen, J.T. (1998) *WHSC1*, a 90 kb SET domain-containing gene, expressed in early development and homologous to a *Drosophila* dysmorphism gene maps in the Wolf-Hirschhorn syndrome critical region and is fused to *IgH* in t(4;14) multiple myeloma. *Hum. Mol. Genet.* **7**: 1071-1082.
- Steck, P.A., Pershouse, M.A., Jasser, S.A., Yung, W.K.A., Lin, H., Ligon, A.H., et al. (1997) Identification of a candidate tumor suppressor gene, *MMAC1*, at chromosome 10q23.3 that is mutated in multiple advanced cancers. *Nature Genet.* **15**: 356-362
- Stratton, M.R., Ford, D., Neuhasen, S.L., Seal, S., Wooster, R., et al. (1994) Familial male breast cancer is not linked to the *BRCA1* locus on chromosome 17q. *Nature Genet.* **7**: 103-107.
- Suggs, S.V., Wallace, R.B., Hirose, T., Kawashima, E.H., and Itakura, K. (1981) Use of synthetic oligonucleotides as hybridization probes: isolation of cloned cDNA sequences for human beta 2-microglobulin. *Proc. Natl. Acad. Sci. U.S.A.* **78**: 6613-6617.
- Suzuki, H., Komiya, A., Emi, M., Kuramochi, H., Shiraishi, T., Yatani, R., and Shimazaki, J. (1996) Three distinct commonly deleted regions of chromosome arm 16q in human primary and metastatic prostate cancers. *Genes Chrom. Cancer* **17**: 225-233.

- Swift, M., Morell, D., Massey, R.E., and Chase, C.L. (1991) Incidence of cancer in 161 families affected by ataxia-telangiectasia. *N. Engl. J. Med.* **325**: 1831-1836.
- Szabo, C.I., and King, M.C. (1995) Inherited breast and ovarian cancer. *Hum. Mol. Genet.* **4**: 1811-1817.
- Takagaki, Y., Seipelt, R.L., Peterson, M.L., and Manley, J.L. (1996) The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B-cell differentiation. *Cell* **87**: 941-952.
- Takagi, Y., Kohmura, H., Futamura, M., Kida, H., Tanemura, H., Shimokawa, K., and Saji, S. (1996) Somatic alterations of the *DPC4* gene in human colorectal cancers in vivo. *Gastroenterology* **111**: 1369-1372.
- Takaku, K., Miyoshi, H., Matsunaga, A., Oshima, M., Sasaki, N., and Taketo, M.M. (1999) Gastric and duodenal polyp in *Smad4* (*Dpc4*) knockout mice. *Cancer Res.* **59**: 6113-6117.
- Takaku, K., Oshima, M., Miyoshi, H., Matsui, M., Seldin M.F., and Taketo, M.M. (1998) Intestinal tumorigenesis in compound mutant mice of both *Dpc4*(*Smad4*) and *Apc* genes. *Cell* **92**: 645-656.
- Takekawa, M., and Saito, H. (1998) A family of stress inducible GADD45-like proteins mediate activation of the stress-responsive MTK1/MEKK4/MAPKKK. *Cell* **95**: 521-530.
- Tandon, A., Clark, G., Chamness, G.C., Ullrich, A., and McGuire, W. (1989) *HER-2/neu* oncogene protein and prognostic in breast cancer. *J. Clin. Oncol.* **7**: 1120-1128.
- T'Ang, A., Varley, J.M., Chakraborty, S., Murphree, A.L., and Fung, T.K. (1988) Structure rearrangement of the retinoblastoma gene in human breast carcinoma. *Science* **242**: 263-266.
- Tavtigian, S.V., Simard, J., Rommens, J., Couch, F., Shattuck-Eidens, D., *et al.* (1996) The complete *BRCA2* gene and mutations in chromosome 13q-linked kindreds. *Nature Genet.* **12**: 333-337.
- Tetsu, O., and McCormick, F. (1999) Beta-catenin regulates expression of cyclin D1 in colon carcinoma cells. *Nature* **398**: 422-426.
- The Finnish-German APECED Consortium. (1997) An autoimmune disease, APECED< caused by mutations in a novel gene featuring two PHD-type zinc-finger domains. *Nature Genet.* **17**: 399-403.
- Thibodeau, S.N., Bren, G., and Schaid, D. (1993) Microsatellite instability in cancer of the proximal colon. *Science* **260**: 816-819.
- Thiagalingam, S., Lengauer, C., Leach, F.S., Schutte, M., Hahn, S.A., Overhauser, J., Willson, J.K., Markowitz, S., Hamilton, S.R., Kern, S.E., Kinzler, K.W., and Vogelstein, B. (1996) Evaluation of candidate tumour suppressor genes on chromosome 18 in colorectal cancers. *Nature Genet.* **13**: 343-346.



- Thomas, G.A., and Raffel, C. (1991) Loss of heterozygosity on 6q, 16q, and 17p in human central nervous system primitive neuroectodermal tumors. *Cancer Res.* **51**: 639-643.
- Thompson, F., Emerson, J., Dalton, W., Yang, J.-M., McGee, D., Villar, H., Knox, S., Massey, K., Weinstein, R., Bhattacharyya, A., and Trent, J. (1993) Clonal chromosome abnormalities in human breast carcinomas I. Twenty-eight cases with primary disease. *Genes Chrom. Cancer* **7**: 185-193.
- Thomson, G., and Esposito, M.S. (1999) The genetics of complex diseases. *Trends Genet.* **15**: M17-M20.
- Thorlacius, S., Olafsdottir, G., Tryggvadottir, L., Neuhausen, S., Jonasson, J.G., Tavtigian, S.V., Tulinius, H., Ogmundsdottir, H.M., and Eyfjord, J.E. (1996) A single *BRCA2* mutation in male and female breast cancer families from Iceland with varied cancer phenotypes. *Nature Genet.* **13**: 117-119.
- Tkachuck, D.C., Kohler, S., and Cleary, M.L. (1992) Involvement of a homolog of *Drosophila trithorax* by 11q23 chromosomal translocations in acute leukemias. *Cell* **71**: 691-700.
- Tribioli, C., Mancini, M., Plassart, E., Bione, S., Rivela, S., Sala, C., Torri, G., and Toniolo, D. (1994) Isolation of new genes in distal Xq28: transcriptional map and identification of a human homologue of the *ARD1* N-acetyl transferase of *Saccharomyces cerevisiae*. *Hum. Mol. Genet.* **3**: 1061-1067.
- Trybus, T.M., Burgess, A.C., Wojno, K.J., Glower, T.W., and Macoska, J.A. (1996) Distinct areas of allelic loss on chromosomal regions 10p and 10q in human prostate cancer. *Cancer Res.* **56**: 2263-2267.
- Tsuda, H., Callen, D.F., Fukutomi, T., Nakamura, Y., and Hirohashi, S. (1994) Allele loss on chromosome 16q24.2-qter occurs frequently in breast cancers irrespectively of differences in phenotype and extent of spread. *Cancer Res.* **54**: 513-517.
- Tsuda, H., Zhang, W., Shimosato, Y., Yokota, J., Terada, M., Sugimura, T., Miyamura, T., and Hirohashi, S. (1990) Allele loss on chromosome 16 associated with progression of human hepatocellular carcinoma. *Proc. Natl. Acad. Sci. U.S.A.* **87**: 6791-6794.
- Tsurugi, K., and Mitsui, K., (1991) Bilateral hydrophobic zipper as a hypothetical structure which binds acidic ribosomal protein family together on ribosomes in yeast *Saccharomyces cerevisiae*. *Biochem. Biophys. Res. Comm.* **174**: 1318-1323.
- Turner, C., and Schapira, A.H. (2001) Mitochondrial dysfunction in neurodegenerative disorders and ageing. *Adv. Exp. Med. Biol.* **487**: 229-251.
- Valverde, P., Healy, E. Jackson, I., Lees, J.L., and Thody, A.J. (1995) Variants of the melanocyte stimulating hormone receptor gene are associated with red hair and fair skin in humans. *Nature Genet.* **11**: 328-330.
- Van Blitterswijk, W.J., and Houssa, B. (2000) Properties and functions of diacylglycerol

kinases. *Cell Signal* **12**: 595-605.

- Van de Vijver, M.J., Peterse, J.L., Mooi, W.J., Wisman, P., Lomans, J., Dalesio, O., and Nusse, R. (1988) Neu-protein overexpression in breast cancer. Association with comedo-type ductal carcinoma in situ and limited prognostic value in stage II breast cancer. *N. Engl. J. Med.* **319**(19):1239-1245.
- Varley, J.M., Swallow, J.E., Brammar, W.J., Whittaker, J.L., and Walker, R.A. (1987) Alterations to either c-erbB-2 (neu) or c-myc proto-oncogenes in breast carcinomas correlate with poor short-term prognosis. *Oncogene* **1**: 423-430.
- Vazza, G., Zortea, M., Boaretto, F., Micaglio, G.F., Sartori, V., and Mostacciuolo, M.L. (2000) A new locus for autosomal recessive spastic paraplegia associated with mental retardation and distal motor neuropathy, *SPG14*, maps to chromosome 3q27-q28. *Am. J. Hum. Genet.* **67**: 504-509.
- Wang, J., Sun, L., Myeroff, L., Wang, X., Gentry, L.E., Yang, J., Liang, J., Zborowska, E., Markowitz, S., Willson, J.K., and Brattain, M.G. (1995) Demonstration that mutation of the type II transforming growth factor- $\beta$  receptor inactivates its tumor suppressor activity in replication error-positive colon carcinoma cells. *J. Biol. Chem.* **270**: 22044-22049.
- Wang, S.C., Lin, S.H., Su, L.K., and Hung, M.C. (1997) Changes in *BRCA2* expression during progression of the cell cycle. *Biochem. Biophys. Res. Commun.* **234**: 247-251.
- Wang, T.C., Cardiff R.D., Zukerberg, L., Lees, E., Arnold, A., and Schmidt, E.V. (1994) Mammary hyperplasia and carcinoma in MMTV-cyclin D1 transgenic mice. *Nature* **369**: 669-671.
- Webb, S., Flanagan, N., Callaghan, N., and Hutchinson, M. (1997) A family with hereditary spastic paraparesis and epilepsy. *Epilepsia.* **38**: 495-499.
- Weinberg, R.A. (1991) Tumor suppressor genes. *Science* **254**: 1138-1146.
- Weinberg, R.A. (1995) The retinoblastoma protein and cell cycle control. *Cell* **81**: 323-330.
- Weissenbach, J., Gyapay, G., Dib, C., Vignal, A., Morissette, J., Millasseau, P., Vaysseix, G., and Lathrop, M. (1992) A second-generation linkage map of the human genome. *Nature* **359**: 794-801.
- Welsh, P.L., Owens, K.N., and King, M.C. (2000) Insights into the functions of *BRCA1* and *BRCA2*. *Trends Genet.* **16**: 69-74.
- Wheeler, D.L., Church, D.M., Lash, A.E., Leipe, D.D., Madden, T.L., Pontius, J.U., Schuler, G.D., Schriml, L.M., Tatusova, T.A., Wagner, L., and Rapp, B.A. (2001) Database resources of the national center for biotechnology information. *Nucleic Acids Res.* **29**:11-16.
- White, E. (1996) Life, death, and the pursuit of apoptosis. *Genes Dev.* **10**: 1-15.
- Whitmore, S.A., Crawford, J., Apostolou, S., Eyre, H., Baker, E., Lower, K.M., Settasatian,

- C., Goldup, S., Seshadri, R., Gibson, R.A., Mathew, C.G., Cleton-Jansen, A.M., Savoia, A., Pronk, J.C., Auerbach, A.D., Doggett, N.A., Sutherland, G.R., and Callen, D.F. (1998a) Construction of a high-resolution physical and transcription map of chromosome 16q24.3: a region of frequent loss of heterozygosity in sporadic breast cancer. *Genomics* **15**:1-8
- Whitmore, S.A., Settasatian, C., Crawford, J., Lower, K.M., McCallum, B., Seshadri, R., Cornelisse, C.J., Moerland, E.W., Cleton-Jansen, A.M., Tipping, A.J., Mathew, C.G., Savnio, M., Savoia, A., Verlander, P., Auerbach, A.D., Van Berkel, C., Pronk, J.C., Doggett, N.A., and Callen, D.F. (1998b) Characterization and screening for mutations of the growth arrest-specific 11 (GAS11) and C16orf3 genes at 16q24.3 in breast cancer. *Genomics* **15**:325-331.
- Wieland, I., Ropke, A., Stumm, M., Sell, C., Weidle, U.H., and Wieacker, P.F. (2000) Molecular characterization of the DICE1 (DDX26) tumor suppressor gene in lung carcinoma cells. *Oncol. Res.* **12**: 491-500.
- Wilson, C.A., Ramos, L., Villasenor, M.R., Anders, K.H., Press, M.F., Clarke, K., Karlan, B., Chen, J.J., Scully, R., Livingston, D., Zuch, R.H., Kanter, M.H., Cohen, S., Calzone, F.J., and Slamon, D.J. (1999) Localization of human BRCA1 and its loss in high-grade, non-inherited breast carcinomas. *Nature Genet.* **21**: 236-240.
- Wong, A.K., Pero, R., Ormonde, P.A., Tavtigian, S.V., Bartel, P.A. (1997) RAD51 interacts with the evolutionarily conserved BRC motifs in the human breast cancer susceptibility gene *BRCA2*. *J. Biol. Chem.* **272**: 31941-31944
- Wooster, R., Bignell, G., Lancaster, J., Swift, S., Seal, S., et al. (1995) Identification of the breast cancer susceptibility gene *BRCA2*. *Nature* **378**: 789-791.
- Wooster, R., Ford, D., Mangion, J., Ponder, P.A., Peto, J., Easton, D.F., et al. (1993) Absence of linkage to the ataxia telangiectasia locus in familial breast cancer. *Hum. Genet.* **92**: 91-94.
- Wooster, R., Neuhasen, S.L., Mangion, J., Quirk, Y., Ford, D., et al. (1994) Localisation of breast cancer susceptibility gene, *BRCA2*, to chromosome 13q12-13. *Science* **265**: 2088-2090
- Wu, L.C., Wang, Z.W., Tsan, J.T., Spillman, M.A., Phung, A., et al. (1996) Identification of a RING protein that can interact in vivo with the *BRCA1* gene product. *Nature. Genet.* **14**: 430-440.
- Wyman, A.R., and White, R.. (1980) A highly polymorphic locus in human DNA. *Proc. Natl. Acad. Sci. U. S. A.* **77**: 6754-6758.
- Xu, J., Wiesch, D.G., and Meyers, D.A. (1998) Genetics of complex human diseases: genome screening, association studies and fine mapping. *Clin. Exp. Allergy* **28** Suppl 5: 1-5.
- Xu, X., Weaver, Z., Linke, S.P., Li, C., Gotay, J., Wang, X.W., Harris, C.C., Ried, T., and Deng, C.X. (1999) Centrosome amplification and a defective G2-M cell cycle checkpoint induce genetic instability in *BRCA1* exon 11 isoform-deficient cells. *Mol. Cell* **3**: 389-

- Yagasaki, F., Jinnai, I., Yoshida, S., Yokoyama, Y., Matsuda, A., Yagasaki, F., Jinnai, I., Yoshida, S., Yokoyama, Y., Matsuda, A., Kusumoto, S., Kobayashi, H., Terasaki, H., Ohyashiki, K., Asou, N., Murohashi, I., Bessho, M., and Hirashima, K. (1999) Fusion of TEL/ETV6 to a novel ACS2 in myelodysplastic syndrome and acute myelogenous leukemia with t(5;12)(q31;p13). *Genes Chromosomes Cancer* **26**: 192-202.
- Yamamoto, H., Sawai, H., and Perucho, M. (1997) Frameshift somatic mutations in gastrointestinal cancer of the microsatellite mutator phenotype. *Cancer Res.* **57**: 4420-4426.
- Yamamoto, H., Sawai, H., Weber, T.K., Rodriguez-Bigas, M.A., and Perucho, M. (1998) Somatic frameshift mutations in DNA mismatch repair and proapoptosis genes in hereditary nonpolyposis colorectal cancer. *Cancer Res.* **58**: 997-1003.
- Yamashita, A., Watanabe, M., Tonegawa, T., Sugiura, T., and Waku, K. (1995) Acyl-CoA binding and acylation of UDP-glucuronosyltransferase isoforms of rat liver: their effect on enzyme activity. *Biochem. J.* **312**: 301-308.
- Yamashita, Y., Kumabe, T., Cho, Y.-Y., Watanabe, M., Kawagishi, J., Yoshimoto, T., Fujino, T., Kang, M.-J., and Yamamoto, T.T. (2000) Fatty acid induced glioma cell growth is mediated by the acyl-CoA synthetase 5 gene located on chromosome 10q25.1-q25.2, a region frequently deleted in malignant gliomas. *Oncogene* **19**: 5919-5925.
- Ye, S.C., Foster, J.M., Li, W., Liang, J., Zborowska, E., Venkateswarlu, S., Gong, J., Brattain, M.G., and Willson, J.K. (1999) Contextual effects of transforming growth factor beta on the tumorigenicity of human colon carcinoma cells. *Cancer Res.* **59**: 4725-4731.
- Yih, J.S., Wang, S.J., Su, M.S., Tsai, S.C., Lin, R.H., Lin, K.N., and Liu, H.C. (1993) Hereditary spastic paraplegia associated with epilepsy, mental retardation and hearing impairment. *Paraplegia* **31**: 408-411.
- Yonish-Rouach, E. (1996) The p53 tumour suppressor gene: a mediator of a G1 growth arrest and of apoptosis. *Experientia* **52**: 1001-1007.
- Yoshikawa, K., Honda, K., Inamoto, T., Shinohara, H., Yamauchi, A., Suga, K., Okuyama, T., Shimada, T., Kodama, H., Noguchi, S., Gazdar, A.F., Yamaoka, Y., and Takahashi, R. (1999) Reduction of BRCA1 protein expression in Japanese sporadic breast carcinomas and its frequent loss in BRCA1-associated cases. *Clin. Cancer Res.* **5**: 1249-1261.
- Yuan, Y., Mendez, R., Sahin, A., and Dai, J.L. (2001) Hypermethylation lead to silencing of the SYK gene in human breast cancer. *Cancer Res.* **61**: 5558-5561.
- Zhang, H., Tomblin, G., Weber, B.L. (1998a) BRCA1, BRCA2, and DNA damage response: collision or collusion. *Cell* **92**: 433-436.
- Zhang, X., Morera, S., Bates, P.A., Whitehead, P.C., Coffey, A.I., Hainbucher, K., Nash, R.A., Sternberg, M.J., Lindahl, T., and Freemont, P.S. (1998b) Structure of an XRCC1 BRCT domain: a new protein-protein interaction module. *EMBO J.* **17**: 6404-6411.

# **Appendix**

## **Publications**

---

Settasatian, C., Whitmore, S.A., Crawford, J., Bilton, R.L., Cleton-Jansen, A.M., (et al.), (1999) Genomic structure and expression analysis of the spastic paraplegia gene, SPG7.  
*Human Genetics*, v. 105 (1-2), pp. 139-144.

NOTE:

This publication is included in the print copy  
of the thesis held in the University of Adelaide Library.

It is also available online to authorised users at:

<http://dx.doi.org/10.1007/s004399900087>

Crawford, J., Ianzano, L., Savino, M., Whitmore, S., Cleton-Jansen, A.M., Settasatian, C., (et al.), (1999) The PISSLRE gene: structure, exon skipping, and exclusion as tumor suppressor in breast cancer.  
*Genomics*, v. 56 (1), pp. 90-97.

NOTE:

This publication is included in the print copy  
of the thesis held in the University of Adelaide Library.

It is also available online to authorised users at:

<http://dx.doi.org/10.1006/geno.1998.5676>

Cleton-Jansen, A.M., Callen, D.F., Seshadri, R., Goldup, S., McCallum, B., (et al.),  
(2001) Loss of heterozygosity mapping at chromosome arm 16q in 712 breast tumors  
reveals factors that influence delineation of candidate regions.  
*Cancer Research*, v. 61 (3), pp. 1171-1177.

NOTE:

This publication is included in the print copy  
of the thesis held in the University of Adelaide Library.



### Amendments of the Thesis

Study of the disease associated genes on the long arm of chromosome 16, at the region frequently lost in breast cancer

Page, Figure, or Table	Description of the amendments
Page 47	<p><i>Line 11</i>, in the bracket, inserts an additional reference (Van de Wetering et al., 2001) for the mutations of CDH1 reported in breast cancer cell lines. In the “References” section adds the full reference as:</p> <p>Van de Wetering, M., Barker, N., Harkes, I.C., van der Heyden, M., Dijk, N.J., Hollestelle, A., Klijn, J.G., Clevers, H., and Schutte, M. (2001) Mutant E-cadherin breast cancer cells do not display constitutive Wnt signaling. <i>Cancer Res.</i> <b>61</b>: 278-284.</p>
Page 50	<p><i>Line 1</i>, in the bracket, changing from “Cleton-Jansen, personal communication” to “Cleton-Jansen <i>et al.</i>, 2001” (this publication can be found in “Appendix” section)</p>
Figure 1.1	<p>Mechanism (ii), the complete description is “Loss and followed by reduplication, as observed by Cavenee <i>et al.</i> this resulted in three copies of the <i>Rb</i> chromosome</p>
Figure 4.11	<p>The correct genotype of SPG7 in individual S26-7 is the same as that in individual S26-8</p>
Figure 5.6	<p>Additional description for the legend, ..... The bands seen in the control, “minus (-)” reaction of lanes 2, 3 and lanes 4, 5 were from contaminating genomic DNA. Bands in lanes 2(-) and 3(-) resulted from the amplification of part of the gene that includes intronic sequence, later found to be between exons 8 and 9 of <i>T13</i>. Bands in lanes 4(-) and 5(-) were amplified from genomic sequence of exon 9 of <i>T13</i>.</p>
Table 4.5	<p>The changes seen in exon 7 (IVS6-34G/T and IVS7+5A/G), exon 8 (1032C/T), exon 9 (IVS9+10C/T), exon 10 (IVS10+19A/G), and exon 12 (IVS12+13C/T) should be described as polymorphism</p>



## Abbreviations (additional)

AMP	adenosine monophosphate
APC	adenomatous polyposis coli
APRT	adenine phosphoribosyl transferase
ATM	ataxia-telangiectasia mutated
BARD-1	BRCA1-associated ring domain 1
BRCA	breast cancer associated
CDH	cadherin
CDK	cyclin-dependent kinase
CMAR	cellular matrix adhesion regulator
DAG	diacylglycerol
FAP	familial adenomatous polyposis
HNPCC	hereditary non-polyposis colorectal cancer
HSP	hereditary spastic paraplegia
MMR	mismatch repair
MSI	microsatellite instability
NF	neurofibromatosis
PE	phorbol ester
PHD	plant homeodomain
PTEN	phosphatase and tensin homolog deleted on chromosome 10
RB	retinoblastoma