Analyses Of Candidate Tumour Suppressor Genes Mapping To The 16q24.3 Breast Cancer Loss Of Heterozygosity Region.

Jason Anthony Powell B.Sc. (Hon)

Thesis submitted for the Degree of Doctor of Philosophy

Department of Laboratory Genetics,
Women's and Children's Hospital, North Adelaide, South Australia.

Faculty of Sciences, Schools of Molecular and Biomedical Science, Discipline of Genetics,
University of Adelaide, South Australia.

November 2003.

Table of Contents

Amendments to Thesis

1. pg IV: line 4: "expresses" > "expressed"
2. pg V: line 9: "a nucleotide change present in the" > "a nucleotide change present in the constitutive DNA"
3. pg V: line 24: "affect" > "effect"

4a. pg 3: line 4: reference "Devilee and Cornelisse, 1994" > "Tavassoli, FA, Devilee P. (Eds) World Health Organization Classification of Tumours, Pathology and genetics of tumours of the breast and female genital organs. IARC Press Lyon, 2003, pg 23."

4b. According to this text invasive lobular carcinoma represents 5-15% of invasive breast tumours.

5. pg 4: line 14: "Webber" > "Weber"
6. pg 4: line 19: "pathologist" > "pathologists"
7. pg 4: line 29: Insert reference "Venkitaraman AR. (2002). Cancer susceptibility and the functions of BRCA1 and BRCA2. *Cell.* **108**:171-182.
8. pg 5: line 10: "novel" > "novel (*ie.* Lack of mutational hotspots)"
9. pg 7: line 4: reference "Yuan *et al*, 2003" > "Schuuring E, van Damme H, Schuuring-Scholtes E, Verhoeven E, Michalides R, Geelen E, de Boer C, Brok H, van Buuren V, Kluin P. (1998). Characterization of the EMS1 gene and its product, human Cortactin. *Cell Adhes Commun.* **6**(2-3):185-209."
10. pg 7: line 29: reference "Schultz *et al*, 1996" > "Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. (1987). Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science.* **235**: 177-182."
11. pg 8: line 4: reference "Menard *et al*, 2002" > "van de Vijver MJ, Peterse JL, Mooi WJ, Wisman P, Lomans J, Dalesio O, 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):1239-45.
12. pg 8: line 15: "oncogene" > "human oncogene"
13. pg 10: line 2: reference "Schultz *et al*, 2002" > "Varley JM, Brammar WJ, Lane DP, Swallow JE, Dolan C, Walker RA. (1991). Loss of chromosome 17p13 sequences and mutation of p53 in human breast carcinomas. *Oncogene.* **6**(3):413-21.
14. pg 11: line 15: Insert subheading "1.4.4) CDH1 (see section 1.5.1)"
15. pg 12: line 31: Insert reference "Wernert N, Locherbach C, Wellmann A, Behrens P, Hugel A. (2001). Presence of genetic alterations in microdissected stroma of human colon and breast cancers *Anticancer Res.* **21**(4A):2259-64.
16. pg 13: line 1: Insert sentence "LOH of CDH1 and somatic mutations have been found in lobular carcinoma *in situ* as well (Vos CB, Cleton-Jansen AM, Berx G, de Leeuw WJ, ter Haar NT, van Roy F, Cornelisse CJ, Peterse JL, van de Vijver MJ. (1997). E-cadherin inactivation in lobular carcinoma in situ of the breast: an early event in tumorigenesis. *Br J Cancer.* **76**(9):1131-3.
17. pg 14: line 15: "were labour intensive" > "was labour intensive"
18. pg 73: line 27: "Clenton-Jansen" > "Cleton-Jansen"
19. pg 78: line 4: "Weak expression" > "Very weak expression"
20. pg 141: line 8. Insert sentence " Retroviral mediated gene transfer was monitored by fluorescent microscopy (Fig 6d)"
21. pg 146: Figure 3 legend, line 8: "3-testies, 4-ovary" > "3-ovary, 4-testies"

Summary

Breast cancer is the second largest cause of cancer related deaths and the most common cancer occurring in women of the western world. Loss of heterozygosity (LOH) of chromosome 16q24.3 is a common genetic alteration observed in invasive ductal and lobular breast carcinomas. Chromosomal regions of frequent LOH are believed to represent the location of candidate tumour suppressor genes. Refined LOH analysis by collaborating laboratories was successful in narrowing down the candidate region to approximately 3 Mb, demarked by the 16q telomere and the marker *D16S498*.

The main aim of this thesis was to isolate the gene(s) responsible for breast cancer using a positional cloning strategy. At the commencement of this study a cosmid contig of this LOH region had already been constructed and stretched 800 kb from the 16q telomere to the gene *CDH15*. This physical map was extended another 1.8 Mb to totally encompass the 16q24.3 LOH interval. The resulting physical map consisted of 63 PAC and 102 BAC clones, with an average of 4.5X coverage. The individual restriction fragments from each clone were ordered and orientated, enabling precise clone overlaps to be established. From this detailed physical map, a minimal tiling path of 26 clones was selected and used as template DNA for large scale sequencing analyses. Each clone was shotgun sequenced, assembled and finished. The sequence data generated was used to anchor the sequence available from the first draft of the human genome. Detailed *in silico* analysis of the genomic sequence data enabled the identification of 116 predicted genes. Several Unigene EST clusters were linked together into single genes and numerous singleton ESTs were rejected as database contaminants. These *in silico* predicted genes were tested for their transcription status by RT-PCR. 104 of the 116 predicted genes amplified from pooled tissue RNA and therefore represent real genes.

Seven of theses genes have been previously screened for mutation in breast tumour DNA: SPG7, BBC1, CPNE7, CDK10, FANCA, GAS11, and C16ORF3. None of these genes were found to harbor any mutations in sporadic breast cancer DNA samples. It was impractical to use this labor-intensive procedure to screen all the remaining candidates and furthermore, tumour DNA stocks were limited.

To identify a possible involvement of any 16q24.3 genes in breast cancer, quantitative mRNA expression analyses of all the corresponding transcripts were performed. It was hypothesized that the tumour suppressor genes associated with the 16q24.3 LOH may be aberrantly expressed due to the presence of mutations or promoter hypermethylation. Using the expression value of each gene in normal breast tissue as the baseline, the expression variability of each gene was examined in a panel of human breast cancer cell lines. Two known tumour suppressor genes, *SYK* and *CDKN2A*, exhibited high variability across the panel of breast cancer cell lines and consequently, validated the screen. Of the 104 genes studied, 3 exhibited considerable expression variability across the panel of breast cancer cell lines, *CBFA2T3, CYBA* and Hs.7970. The notion of these genes being breast cancer tumour suppressors is compatible with what is known about their respective function.

Of the remaining transcripts, several still presented as good candidates based on additional functional evidence. For example, microcell mediated chromosome transfer experiments identified a 16q24 YAC clone capable of inducing cellular senescence in breast cancer cell lines (Reddy *et al*, 2000). The YAC 792E1 was mapped to the 16q24.3 LOH interval and shown to contain two genes, Hs.118944 and Hs.7970. The genomic structure and expression profile of these uncharacterized Unigene EST clusters was established. These candidates were subsequently screened for mutation in breast tumour DNA and finally, the strongest candidate, Hs.7970, was subjected to a cell-based assay for assessment of potential tumour suppressor function. Neither gene exhibited any disease specific mutation and furthermore, Hs.7970 failed to suppress the formation of colonies in soft agar and on plastic, a hallmark of known tumour suppressors. These results suggest that these genes are not involved in breast cancer.

Additional genes presented as candidates based on their homologies to other known human or mouse proteins. For example, TSG16 exhibited significant homology to the ankyrin (ANK) repeat motif of the BRCA1-associated RING domain protein 1 (BARD1). As ankyrin repeat domains function as sites for protein-protein interactions, it was hypothesized that TSG16 shares common protein partners with BRCA1 and therefore may itself function in breast carcinogenesis. *In silico* analysis identified

matching ESTs with reported down regulation in nasopharyngeal carcinoma (NPC). Interestingly, NPC also exhibits 16q LOH further supporting a possible tumour suppressor function for TSG16. Comprehensive RT-PCR and expression analysis identified TSG16 as a 9307 bp transcript, expressed in all tissues examined. The TSG16 ORF of 7995 nucleotides encodes a 2,664 amino acid protein with a predicted molecular weight of 298 kDa. Mutational analysis in sporadic breast tumour DNA identified one tumour specific missense mutation. The biological significance of the mutations awaits functional complementation. Furthermore, the screening of non-BRCA1 and BRCA2 familial breast cancer patients uncovered a nucleotide change present in the blood of three affected individuals from the same family. However, DNA analysis of one tumour revealed the wild type pattern and therefore it is unlikely that this nucleotide change is related to the risk of breast cancer. Subsequent experiments examined this large novel gene in detail.

The ANK repeat domain of TSG16 was used as bait to screen a human breast cDNA library and this domain identified several members of the PIAS (potent inhibitors of STAT transcription) protein family. PIAS proteins perform an increasing repertoire of functions including the inhibition of the STAT signaling. TSG16, the ANK repeat domain, PIAS1 and PIAS3 were all amplified and cloned into mammalian expression vectors for transient transfection and subsequent immunoprecipitation experiments. All exogenously expressed proteins were stability expressed in 293T cells, but failed to immunoprecipitate with each other. A STAT3 transcriptional reporter assay was performed to access the effect of TSG16 on PIAS mediated inhibition of STAT transcription. PIAS3 inhibited LIF stimulated STAT3 transcription, confirming earlier reports. The affect of TSG16 could not be assessed as this gene was found to activate the *Renilla* luciferase reporter construct, originally included for transfection efficiency normalization. It is possible that TSG16 may act as a generalized transcriptional activator and future experiments should investigate this possibility.

In summary, 104 candidate genes have been identified in the 16q24.3 LOH interval. Quantitative mRNA expression analysis in breast cancer cell lines served as an

initial screen for potential tumour suppressors. Additional genes presented as candidates and subsequent experiments examined these genes in detail. The precise mapping and sequencing of the human genome has greatly assisted positional cloning strategies for candidate disease gene isolation.

Declaration

This work contains no material which has been accepted for the 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 my consent to this thesis, when deposited in the University library, being available for loan and photocopying.

Signed: Date: 21/10/03

List of Publications

1. Cleton-Jansen, A. M., Callen, D. F., Seshadri, R., Goldup, S., Mccallum, B., Crawford, J., **Powell J. A.**, Settasatian, C., van Beerendonk, H., Moerland, E. W., Smit, V. T., Harris, W. H., Millis, R., Morgan, N. V., Barnes, D., Mathew, C. G. and Cornelisse, C. J. (2001). Loss of heterozygosity mapping at chromosome arm 16q in 712 breast tumours reveals factors that influence delineation of candidate regions. *Cancer Res.* **61**: 1171-1177.

2. Kochetkova, M., McKenzie, O. L. D., Bais, A. J., Martin, J. M., Secker, G. A., Seshadri, R., **Powell, J. A.**, Hinze, S. J., Gardner, A. E., Spendlove, H. E., O'Callaghan, N. J., Cleton-Jansen, A., Cornelisse, C., Whitmore, S. A., Crawford, J., Kremmidiotis, G., Sutherland, G. R. and Callen, D. F. (2002). CBFA2T3 (MTG16) is a putative breast tumour suppressor from the breast cancer loss of heterozygosity region at 16q24.3. *Cancer Res.* **62**: 4599-4604.

3. **Powell, J. A.,** Gardner, A. J., Bais, A. J., Hinze, S. J., Baker, E., Whitmore, S., Crawford, J., Kochetkova, M., Spendlove, H. E., Doggett, N. A., Sutherland, G. R., Callen, D. F. and Kremmidiotis, G. (2002). Sequencing, transcript identification, and quantitative gene expression profiling in the breast cancer loss of heterozygosity region 16q24.3 reveal three potential tumor-suppressor genes. *Genomics.* **80**(3): 303-310.

4. **Powell, J. A.,** Settasatian, C., Lower, K. M., McKenzie, O. L. D., Zarqa, S., Kumar, R., Bais, A. J., Crawford, J. C., O'Callaghan, N. J and Callen, D. F. (2003). ANKRD11 and ANKRD12 are novel 9kb genes encoding nuclear proteins with ankyrin domains: screening of the ANKD11 gene for involvement in breast cancer. *Gene.* Submitted September 2003.

Abbreviations

- ADH — Atypical Ductal Hyperplasia
- ANK — Ankyrin Repeat Domain
- ATM — Ataxia-telangiectasia
- APRE — Acute Phase Response Element
- ATTC — American Type Culture Condition
- BAC — Bacterial Artificial Chromosome
- bp — Base Pair
- BLAST — Basic Local Alignment Search Tool
- BARD1 — BRCA1-associated RING Domain Protein 1 (BARD1
- BSA — Bovine Serum Albumin
- CDI — Cyclin Dependent Kinase Inhibitor
- CDK — Cyclin Dependent Kinase
- cDNA — Complementary Deoxyribonucleic Acid
- CGH — Comparative Genomic Hybridisation
- CH16LARS — Chromosome 16 Specific Repeats
- CMV — Cytomegalovirus Immediate-early Promoter
- dbEST — Database of Expressed Sequence Tags
- DMSO — Dimethyl Sulphoxide
- DCIS — Ductal Carcinoma *in situ*
- dNTP — Deoxynucleotide Triphosphate
- DNA — Deoxyribonucleic Acid
- DCIS — Ductal Carcinoma *in situ*
- EBV — Epstein-Barr Virus
- EGFR — Epidermal Growth Factor
- EST — Expressed Sequence Tag
- FCS — Fetal Calf Serum.
- FGFR1 — Fibroblast Growth Factor Receptor 1
- FISH — Fluorescent *in situ* Hybridization

- GAS — INFγ Activation Sequence
- GFP — Green Fluorescent Protein
- gss — Genomic Survey Sequence
- HSV-TK — Thymidine Kinase Basal Promoter of the Herpes Simplex Virus
- htgs — High Throughput Genomic Sequence
- IDC — Infiltrating Ductal Carcinoma
- IHGSC — International Human Genome Sequencing Consortium
- IPTG — Isopropyl -β-D-thiogalactopyraniside
- kb — Kilobase Pairs
- kConFab — Kathleen Cuningham Foundation CONsortium For Research into Familial Breast Cancer
- kDa — Kilodaltons
- LB — Luria-Bertani Medium
- LCIS — Lobular Carcinoma *in situ*
- LOH — Loss Of Heterozygosity
- Mb — Megabase Pairs
- min — Minutes
- Mw — Molecular Weight
- µg — Microgram
- µl — Microlitre
- mg — Milligram
- ml — Milliliter
- mRNA — Messenger Ribonucleic Acid
- NCBI — National Center for Biotechnology Information
- ng — Nanogram
- NLS — Nuclear Localization Signal
- NPC — Nasopharyngeal Carcinoma
- ORF — Open Reading Frame
- PAC — P1 Artificial Chromosome

- PBS — Phosphate Buffered Saline
- PCR — Polymerase Chain Reaction
- PEG — Polyethylene Glycol
- PFGE — Pulse Field Gel Electrophoresis
- RACE — Rapid Amplification of cDNA Clones
- RFVI — Relative Fold Variability Index
- RNA — Ribonucleic Acid
- ROS — Reactive Oxygen Species
- RT-PCR — Reverse Transcription PCR
- SAGE — Serial Analysis Of Gene Expression
- SRO — Smallest Region Of Overlap
- SSCA — Single Stranded Conformation Analysis
- SSCP — Single Stranded Conformation Polymorphism
- STS — Sequence Tagged Site
- *Taq* — *Thermus Aquaticus*
- UDH — Usual Ductal Hyperplasia
- UTR — Untranslated Region
- WCH — Women's and Children's Hospital (Adelaide)
- X-gal — 5-bromo-4-chloro-3-indoyl-β-galactopyranoside
- YAC — Yeast Artificial Chromosome

Acknowledgments

I would like to The Department of Laboratory Genetics, Centre for Medical Genetics, Women's and Children's Hospital, North Adelaide for allowing me to pursue my studies. I would like to thank Professor Grant Sutherland for his support and interest in this project, and in the transition from the Women's and Children's Hospital to Bionomics.

I am extremely grateful to Associate Professor David Callen for his daily and overall supervision, encouragement and support throughout my candidature. David has established a great laboratory and I thank him for his excellent guidance and tuition.

I would like to thank my work colleagues for their large contributions to the research presented in this thesis, who without would not have been possible. Gabriel Kremmidiotis (TSG16 functional studies, sequencing and advice), Marina Kochetkova (Functional analysis of candidate transcripts and advice), Hayley Spendlove (TSG16 functional studies) Anthony Bais (Real time PCR and SSCP analysis), Joanna Crawford (SSCP analysis, physical mapping and YAC analysis), Scott Whitmore (SSCP analysis, physical mapping and YAC analysis), Olivia McKenzie (SSCP analysis), Susan Hinze (Real time PCR and transcript analysis), Alison Gardner (Sequencing), Andrew Dunbar, Genevieve Secker, Nathan O' Callaghan, Julie Martin, Karen Lower, Chatri Settasatian and Bev Johns.

I would also like to thank the Department of Molecular Biosciences, Discipline of Genetics, University of Adelaide, Bionomics Ltd., Associate Professor Jeremy Timmis (Discipline of Genetics, University of Adelaide) and Dr Sandra Nicholson (Department of Cancer and Hematology, The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria). I would also like to thank my family for their love and support.

Finally I would like to dedicate this thesis to my love, and best friend Sally Leane. I am extremely grateful of her understanding, love, support and patience throughout this candidature.

Chapter 1

Chapter 1: Introduction.

1.1) Breast cancer.

Breast cancer is one of the most common neoplasms in women of the western world and is the second leading cause of cancer related deaths (Polyak, 2001). One in twelve Australian women develop breast cancer in their lifetime with incidences steadily increasing (Anti-Cancer Foundation, 2002). Breast cancer is also seen in men contributing to 1% of all breast cancers and accounting for 1% of all malignancies found in the male population (Demeter *et al*, 1990). Breast cancer is derived from the epithelial lining of terminal mammary ducts and lobuli. Hormonal influences such as those exerted by oestrogen are important as the incidence of breast cancer increases in post-menopausal women. However, the initial steps in breast cancer development probably occur before the onset of menopause.

The exact cause of breast cancer is unknown but it is believed to be a complex, heterogenous disease that progresses in a step-wise manner. The development of a malignant tumour results from an accumulation of genetic changes involving both positive regulators (oncogenes) and negative regulators (tumour suppressor genes-TSGs) of cell function.

Initially these genetic changes allow cells to escape programmed senescence or irreversible growth arrest, resulting in cell immortalisation, the first essential step in the transformation of a normal cell to one that is cancerous (Reddel, 2000). This process extends the cell's proliferative life span allowing the accumulation of additional mutations resulting in loss of contact inhibition and acquisition of invasive growth capacity. In addition, there needs to be changes at the tissue level, such as evasion of host immune responses and growth restraints imposed by surrounding cells, and the formation of blood supply for the growing tumour (angiogenesis) (Greene, 2002). Molecular analysis of colorectal carcinoma has provided evidence for the notion that the sequential accumulation of genetic lesions within the same stem cell results in malignancy (Fearon *et al,* 1990).

The initial events leading to mammary tumourigenesis are poorly defined however the transformation of the breast epithelial cell is believed to develop through a number of histological stages including usual ductal hyperplasia (UDH), atypical ductal hyperplasia (ADH), ductal carcinoma *in situ* (DCIS), lobular carcinoma *in situ* (LCIS), and infiltrating ductal carcinoma (IDC) (Polyak, 2001). Patients whose biopsy specimens contain UDH, ADH, or DCIS have shown an approximate 2-fold, 4-fold, and 10-fold respective increase in relative risk of developing the invasive form

of breast cancer, and consequently poor prognosis. (Aubele *et al*, 2000). The most common form of breast cancer occurring in women is IDC which has been reported in ~80% of breast cancer cases, followed by LCIS accounting for 10-15% of cases (Devilee and Cornelisse, 1994). Evidence strongly suggests that DCIS is the precursor of IDC therefore the analysis of this early stage malignancy is predicted to be the most informative in identifying critical events for the development of breast cancer. LCIS and DCIS affected individuals have almost 100% survival rate (Carrera and Payne, 1999).

1.2) Hereditary breast cancer.

Hereditary breast cancer accounts for 5 to 10% of primary breast cancer cases (Easton *et al,* 1993). Hereditary breast cancer is characterized by a number of features that differ from sporadic breast cancer. These include earlier age of onset, with as many as 30% of women developing breast cancer before the age 35; an excess of bilateral disease; an association with other malignancies including ovarian, colon, and prostate cancer; and transmission of the disease through successive generations in an autosomal dominant pattern.

Most of the characteristics that distinguish inherited breast cancer from sporadic may be explained by Knudson's 'two-hit' model of carcinogenesis. This model suggests that cancer develops through successive genetic events (or 'hits') that lead to the loss of function of both the maternal and paternal alleles of a cancer susceptibility gene. In inherited forms of cancer, one of these hits is transmitted though the germline ("carriers") and in sporadic cancer the two hits are acquired at the somatic level (Devilee *et al*, 2001). This explains why inherited cases usually occur bilaterally and at earlier stages than the sporadic form of breast cancer because "carriers" of the mutation are predisposed, and only require one additional mutational event. Subsequently cytogenetic evidence suggested that two hits affected the two alleles of a single gene, and the concept of recessive gene inactivation formed the basis for tumour predisposition. The "hits" can arise from a number of mechanisms including mitotic recombination, mitotic non-disjunction with loss of the wild-type allele or duplication of the mutant-allele, gene conversion, deletion or mutation and promoter hypermethylation. Epidemiological analysis of retinoblastoma and Wilm's

tumour have provided important evidence supporting Knudson's 'two-hit' hypothesis (Cavenee *et al*, 1983; Wanger *et al*, 2003; Rahman *et al,* 1998).

The two most common inherited breast cancer autosomal dominant susceptibility genes are *BRCA1* at chromosome band 17q21, and *BRCA2* at chromosome band 13q12-q13 (Hall *et al*, 1990; Wooster *et al*, 1994). The *BRCA1* and *BRCA2* genes were discovered through genetic linkage studies performed in families affected by early-onset breast and in some cases ovarian cancer. Mutations in *BRCA1* account for most of the hereditary breast and ovarian cancer families, and documented in up to 40-50% of families with hereditary breast cancer only (Chen, *et al*, 1999). *BRCA1* and *BRCA2* are believed to function as recessive genes and carriers of *BRCA1* and *BRCA2* mutations are at elevated risk of cancer of the ovary, prostate, and pancreas (Bertwistle and Ashworth, 1998). Approximately 80% of *BRCA1* tumour restricted mutations are frameshift or nonsense mutations disrupting normal protein function (Crouch and Webber, 1996). The complete loss of the wild-type allele (LOH) normally exposes this mutant allele (Osorio *et al,* 2002). In sporadic cases of breast cancer there is virtually no sign of somatic mutations in the *BRCA1* and *BRCA2* genes (Futreal *et al,* 1994; Miki *et al,* 1994), suggesting that additional genes contribute to the development of sporadic breast cancer.

The identification of *BRCA1* and *BRCA2* has allowed pathologists to categorize subsets of familial breast cancer according to the underling germline mutation and resulting histological phenotype (Armes and Venter, 2002). For example, the majority of *BRCA1* associated breast cancers exhibit a distinct phenotype including high grade, oestrogen receptor-negative tumours displaying medullary features. Tumours associated with *BRCA2* mutations fail to exhibit any distinguishable histological characteristics.

The BRCA1 and BRCA2 proteins both exhibit many different biological functions. Firstly, BRCA1 and BRCA2 maintain genomic stability through their involvement in homologous recombination and in transcription-coupled double stranded break repair (Narod, 2002; Scully *et al*, 1997). For example, BRCA1 associates with RAD51 to mediate DNA repair (Fleming *et al,* 2003). The RAD51 protein is homologous to the bacterial RecA protein, a protein known to function in mitotic and meiotic recombination to repair double stranded DNA breaks (Shinohara,

et al, 1992). BRCA1 also associates with RAD50, another protein involved in homologous recombination and DNA damage responses (Zhong *et al*, 1999).

Many experiments have suggested a role for *TP53* in BRCA tumour development. For example, $Rad^{-/-}$ and $Brca^{-/-}$ knock-out mice do not survive, but their phenotypes can be rescued in mice with p53 null backgrounds (Suzuki *et al*, 1997; Lim and Hasty, 1996; Hakem *et al*, 1998). Similarly, $Brca^{+/-}$ $p53^{+/-}$ mice are more susceptible to mammary tumour development following ionising radiation than are $Brca^{+/+}$ $p53^{+/-}$ mice (Cressman *et al*, 1999). *TP53* somatic mutations are frequently found in both BRCA induced breast and ovarian cancers. Most of these *TP53* mutations are novel as the mutant p53 protein retains the ability to transactivate, suppress growth and induce apoptosis, but fail to suppress transformation (Smith *et al*, 1999) (see section 1.4.2). In summary, BRAC1 and BRCA2 maintain genomic stability through their interaction with many different DNA repair proteins, thus linking genome integrity to tumour suppression.

BRCA1 and BRCA2 exhibit additional cellular functions including proliferative induction in both early embryogenesis as well as within the breast epithelial cells during puberty, pregnancy and lactation (Rajan *et al*, 1996). BRCA protein analysis has revealed the presence of a RING finger domain (Miki *et al*, 1994). RING fingers are zinc binding domains defined by a conserved pattern of cysteines and histidines that mediate protein-protein or protein-DNA interactions. These domains facilitate the transfer of ubiquion to proteins tagged for degradation. From these observations it can be hypothesized that loss of BRCA RING finger domains will induce elevated levels of regulatory proteins, in particular cell cycle stimulators.

In addition, BRCA1 and BRCA2 have been shown to function in transcriptional regulation. For example, BRCA1 complexes with RNA polymerase II and both BRCA1 and BRCA2 interact with many transcriptional regulators including ATF1 (Houvras *et al,* 2000), STAT1 (Ouchi *et al,* 2000), *C-MYC* and RNA helicase A (Chen *et al,* 1999). Finally, BRCA1 represses both estrogen-dependent promoters and histone deacetylases (Yarden *et al*, 1999).

Not all families with multiple cases of breast cancer harbour germ-line mutations in *BRCA1* or *BRCA2*. Other factors appear to contribute to hereditary breast cancer including a host of genetic and non-genetic causes such as lifestyle choices or

mutations in other genes that may modulate the risk of breast cancer in mutation carriers (Tonin, 2000). Therefore hereditary syndromes that feature breast cancer include Li-Fraumeni syndrome, Cowden disease, and ataxia telangiectasia, where affected individuals have been shown to harbour germ-line mutations in the *TP53, PTEN*, and *ATM* genes, respectively (Ingvarsson, 1999; Nathanson *et al.*, 2001).

The lack of *BRAC1* and *BRAC2* mutations in 50-60% of hereditary breast cancers suggests additional loci exist that confer a high risk to breast cancer susceptibility. The failure to identify additional breast cancer genes associated with a high risk of disease suggests that these remaining cancers are highly heterogenous and more than one tumour suppressor may exist. Consistent with this was the identification of at least two ataxia-telangiectasia (*ATM*) gene mutations that were associated with breast cancer in multi-case families (Chenevix-Trench *et al*, 2002). ATM is an autosomal recessive disorder characterised by progressive neuronal degradation, immunologic deficiency, radiosensitivity, and an increased risk of cancer, particularly lymphoid malignancies with a predisposition to cancer (Yuille and Coignet, 1998). The *ATM* gene encodes a putative protein kinase and mutants are hypersensitive to DNA damage and display defective in cell cycle checkpoints. A number of well characterised tumour suppressor genes including BRCA1, TP53 and CHK2, lie downstream of ATM that, after DNA damage is detected, trigger cell cycle arrest, DNA repair, or apoptosis (Khanna and Jackson, 2001; Gatei *et al*, 2001).

1.3) Oncogenes and breast cancer.

Proto-oncogenes are the genes whose action promotes the positive regulation of the cell cycle. Oncogenes are activated forms of proto-oncogenes that function in a genetically dominant manner to initiate and promote tumour progression. Mutations converting normal proto-oncogenes into oncogenes are gain-of-function mutations and include point mutations, gene amplification, and hypomethylation promoting the loss of association with their regulatory mechanisms responsible for maintaining the balance between normal and abnormal proliferation. These cells exhibit many cellular properties that are characteristic of tumour cells including enhanced cell cycle progression, loss of contact inhibition, reduced sensitivity to growth inhibitors, enhanced resistance to apoptosis, invasiveness and suppression of terminal differentiation. The *MYC, CCND1*, *HER2* and *RAS* genes are examples of important oncogenes involved in many cancers including breast cancer, and are briefly

discussed below. Other oncogenes are found to be amplified less frequently in tumours and include *FGFR1,* fibroblast growth factor receptor (DiMario *et al,* 2002); *MDM2* a protein that inactivates the tumour suppressor *TP53* (Chehab *et al,* 2000); *EMS1*, a cytoplasmic protein involved in cytoskeletal organisation (Yuan *et al,* 2003), and Ras-related GTP binding proteins (Downward, 2003).

1.3.1) MYC.

The *MYC* gene located on chromosome 8q24 is amplified in approximately 25% of breast cancers (Brenner and Aldaz, 1997). *MYC* is a transcription factor involved in the regulation of a variety of cellular activities including proliferation, differentiation, and apoptosis. Amplification of this gene has been detected in DCIS, indicating that this is an early event in mammary tumourigenesis. Other studies have suggested a role in tumour promotion with associations between *MYC* amplification and certain aggressive features such as large tumour size and high tumour grade (Bowcock, 1999). *C-MYC* functions as a heterodimer with a second transcription factor, *MAX*, and while it is clear that both cooperate to facilitate neoplastic transformation the exact sequence of events have yet to be defined.

1.3.2) Cyclin D1 (CCND1).

The *CCND1* gene is located on chromosome 11q13, a region with reported amplification in 15-20% of breast tumours (Lammie *et al,* 1991). This gene encodes the *CCND1* protein, a direct regulator of the cell cycle. *CCND1* protein overexpression has been suggested to be the boundary between pre-malignancy and cancer, with elevated mRNA levels reported in only 18% in atypical ductal hyperplasia (ADH), 76-87% of DCIS and 83% of IDC. These alterations have also been detected in pre-malignant breast lesions, confirming *CCND1* inactivation as an early event in breast tumour development (Brenner and Aldaz, 1997).

1.3.3) HER2.

The most studied oncogene in breast cancer is the *HER2* (*c-erbB2/neu*) gene. *HER2* is located on chromosome 17q21 a region that is frequently amplified in breast cancer (Schultz *et al,* 1996). The HER2 protein functions as a tyrosine kinase growth factor receptor, exhibiting high homology to the epidermal growth factor receptor (EGFR). Amplification of the *HER2* gene, and overexpression of its product, induces

cell transformation. Numerous studies have reported amplification in up to 40% of human breast tumours. In addition, elevated levels of HER2 have been reported in 40-60% of cases of DCIS, which suggests its involvement in the development of premalignant lesions (Menard *et al*, 2002).

Oncogenes are viewed as potential therapeutic targets with approaches directed at developing antagonists for these inappropriately expressed proteins. For example, HER2's extracellular accessibility and involvement in tumour aggressiveness makes this receptor an attractive target for tumour-specific therapies (Kirschbaum and Yarden, 2000). Positive results have been obtained from *in vitro*, preclinical and clinical trials using humanized monoclonal antibodies directed to the extracellular domain of this oncogenic protein. Herceptin and Iressa are novel mediating cancer growth inhibition drugs that have been developed to target HER2 and EGFR (HER1) respectively.

1.3.4) RAS.

RAS represents the first oncogene to be discovered and the most widely activated oncogene in human cancers, with incidences ranging from 90% in pancreatic tumours, to 30% in most solid tumours (Bos, 1989). *RAS* is one of a large family of proteins that can bind GTP and act as a signal transduction molecule. GTP-bound activated RAS interacts and regulates other downstream proteins. The mutated oncogenetic form of RAS does not release GTP and remains permanently activated. The identification of RAS mutations in lung precursor carcinogenic lesions (hyperplastic) suggests this mutational event occurs early in the carcinogenesis process (Westra *et al,* 1996). The *RAS* gene exhibits mutational "hot spots" that reside in areas involved in GTP binding and hydrolysis (Minamoto *et al,* 2000). *RAS* has been shown to interact with additional GTPases including p120GAP (Scheffzek *et al,* 1997), Rho (Prendergast *et al,* 1995) and RAC (Scita *et al,* 2000). Rho and its family members function to regulate kinase cascades, gene expression and actin cytoskeleton. The full ranges of molecular and cellular consequences of activated *RAS* are still poorly defined.

1.4) Tumour suppressor genes and breast cancer.

In contrast to oncogenes, tumour suppressor genes have a negative effect on cell growth and survival. The original evidence implicating tumour suppressor genes

in multi-stage carcinogenesis came from somatic cell hybridisation experiments where fusion of malignant and normal cells lead to a loss of tumourgenicity (Harris *et al* 1969; Stanbridge 1976). This suggests that loss of specific tumour suppressor genes results in the gain of uncontrolled cellular growth. Following prolonged culture of these somatic cell hybrids, reversion to tumourgenicity occurred in some cells and this was linked to the loss of specific chromosomes.

Retinoblastoma (*RB1*) is one of many tumour suppressor genes that conform to the classic definition of Knudson's two-hit hypothesis. *RB1* is discussed in more detail below followed by other examples of tumour suppressor genes involved in breast tumourigenesis. Functional inactivation of tumour suppressor genes can be caused by many epigenetic mechanisms besides mutation, including hypermethylation (Baylin *et al,* 1998), increased degradation (Storey *et al,* 1998) or mislocalization (Chen *et al,* 1995; Moll *et al,* 1992).

1.4.1) RB1.

The *RB1* tumour suppressor gene is located on chromosome 13q14 and was originally cloned as the cause of retinoblastoma, the most common childhood eye malignancy. Structural abnormalities including chromosomal loss and mutations have been reported in approximately 20-30% of breast cancers (Oesterreich and Fuqqua, 2002). The *RB1* gene encodes a nuclear protein pRB and when mitogen activated cyclin D-CDK4 complexes phosphorylate this nuclear protein it releases E2F, a transcription factor that activates genes responsible for DNA replication, thereby allowing cell cycle progression into S-phase (Weintraub *et al,* 1995). This altered conformation of pRb allows access of CDK2 during S-phase, which produces additional phosphorylation further inhibiting E2F binding (Harbour and Dean, 2000). Loss of the pRb function is thus believed to promote abnormal proliferation as a result of the disregulation of E2F transcription factors. The hypo-phosphorylated inactive Rb-E2F complex silences transcription by maintaining their core histones in a non-acetylated form, restricting access to transcription activators (Brehm *et al,* 1998).

The *RB1* gene exhibits many different mutations that result in truncated proteins that undergo premature termination or aberrant transcript splicing (Lohmann, 1999; Mancini *et al,* 1997). *RB1* promoter hypermethylation has also emerged as an additional mechanism for *RB1* down-regulation (Stirzaker *et al*, 1997).

1.4.2) TP53.

The *TP53* gene maps to chromosome 17p, a region associated with frequent LOH in breast cancer (Schultz *et al*, 1996). *TP53* encodes the p53 protein, a nuclear phosphoprotein that functions in cell cycle control to mediate differentiation, DNA repair, and apoptosis. *TP53* is a stress response gene that normally exists in an inactive form, but can be activated by various stimuli including DNA damage caused by γ-irradiation, UV radiation, or chemical agents. Other factors leading to p53 activation include hypoxia, genome instability and overexpression of oncogene products. p53 activation occurs by CDK phosphorylation and acetylation (Chehab *et al,* 2000), resulting in the disassociation of the MDM2 protein and *TP53* transcriptional activation (Chehab *et al,* 2000). Activated p53 stimulates the transcription of a number of genes including *CDKN1A* (p21/WAF1/CIP1), *MDM2*, *GADD45A*, *CCNG1* (cyclin G1) and *IGFBP3* (insulin-like growth factor binding protein 3). The p53 activated GADD45 proteins initiate DNA repair by binding and destabilising the histones surrounding the damaged DNA. This protein-DNA disassociation permits access of the proteins involved in the repair process (Carrier *et al,* 1999). Ultimately the activation of the p53 protein results in growth suppression through either cell cycle arrest or induction of apoptosis (Bowcock, 1999).

To date *TP53* is the most common somatically mutated gene in sporadic cancers with mutations in more than 50% of all cancers and approximately 15-35% of breast tumours (Oesterreich and Fuqua, 1999). *TP53* exhibits mutational "hot spots" that reside in areas responsible for DNA binding (Cho *et al,* 1994). These aberrations result in either loss of function or dominant-negative protein action, and therefore disrupt normal p53 function and contribute to tumourigenesis. (Parant and Lozano, 2003).

1.4.3) Cyclin-Dependent Kinase Inhibitors (CDIs).

Alterations in the G1-S cell cycle checkpoint have been documented in human breast cancers. These include RB1 inactivation (previously discussed), amplification of *CCND1* (previously discussed) and inactivation of CDK-inhibitors (CDKIs). CDKIs compete with cyclins for binding of CDKs to prevent G1/S progression of the cell cycle. Certain CDKIs have possible roles in tumour suppression.

CDKI molecules are categorized into two separate families, the CDK2-interacting protein (CIP)/CDK-inhibitory protein (KIP)-family comprising CDKN1A ($p21^{CIP1}$),

CDKN1B ($p27^{KIP1}$) and CDKN1C ($p57^{KIP2}$), and the inhibitor of CDK4 (INK4)-family comprising of CDKN2A ($p15^{INK4B}$), CDKN2B ($p16^{INK4A}$), CDKN2C ($p18^{INK4C}$) and CDKN2D ($p19^{INK4D}$). In breast cancer most of the studies have investigated *CDKN2B* located at 9p21-22. CDKN2B inhibits the cell cycle progression at G1/S by interfering with the action of the cyclin D/CDK4 complex. Chromosome 9p21-22 exhibits 47% LOH in breast carcinomas and analysis of these same tumours revealed a high incidence of *TP53* mutations (Gorgoulis *et al*, 1998).

Extending the family of cell cycle inhibitors is prohibitin, *PHB*, a gene that is specifically implicated in the process of cell immortalisation. Expression of the PHB protein inhibits cell cycle transition and DNA synthesis in normal cells. The 3'untranslated region of prohibitin gene has been shown to function as a trans-acting RNA (riboregulator) that is crucial to its anti-proliferative activity (Russo *et al*, 1998). Prohibitin is localized to chromosome 17q21 where mutations have been reported in sporadic breast tumours (Saito *et al*, 1993).

In summary, the inactivation of various tumour suppressor genes has been shown to be critical in the development and progression of breast cancer. In normal somatic cells these genes act to suppress proliferation by inducing cell cycle arrest and promoting apoptosis.

1.5) Loss of heterozygosity (LOH) and cancer.

Loss of Heterozygosity (LOH) is frequently observed in tumours and is the reduction of constitutional heterozygosity seen at a particular polymorphic locus to homozygosity in the tumour. The large genomic deletions associated with LOH have provided a mechanism for recessive TSG inactivation. Moreover, LOH mediated removal of the wild-type allele exposes the pre-existing mutant allele. This LOH event complies with Knudson's 'two-hit' hypothesis for TSG inactivation (Knudson 1971; Knudson and Strong 1972, section 1.2).

LOH is detected by analyses of polymorphic markers from both disease and normal DNA samples from an affected individual. Frequently observed tumour restricted LOH of specific chromosomal regions is considered to indicate the location of tumour suppressor genes (Devilee and Cornelisse, 1994). This has been successfully demonstrated for *CDKN2A* (9p21) in melanoma (Smith-Sorensen and Hovig, 1999), *DMBT1* (10q25-q26) in brain tumours (Pang *et al*, 2003), *CDH1*

(16q22) in lobular breast carcinoma and diffuse gastric cancer (Berx *et al,* 1995), and *SMAD4* (18q23) in pancreatic cancer (Schneider and Schmid, 2003). These studies have been facilitated by homologous deletions encompassing the causative gene, a rare occurrence for the majority of primary tumours.

The use of LOH analysis to define the location of a TSG has a number of limitations including sporadic LOH due to chromosome instability that is unrelated to tumour suppressor location, clonal and tumour heterogeneity, and technical difficulties defining LOH in tumours with contaminating normal DNA. These experimental and biological limitations make interpretation of LOH data difficult and probably account for the failure to identify the causative genes in cancers with high LOH frequencies. Examples include chromosome 3p21 in lung cancer (Wistuba *et al,* 2000) and 16q24.3 in breast cancer (Whitmore *et al,* 1998a).

LOH intervals have previously been defined as the smallest region of overlap (SRO), or the smallest and simplest genomic deletion common to several tumours (Cleton-Jansen *et al,* 2001). The possibility of multiple tumour suppressor genes mapping within these SROs, suggests this original interpretation of LOH data may have been misleading. Moreover, the same disease phenotype may arise from two independent mutational events and if these genes coincidently map in close proximity to each other and remain indistinguishable in the LOH analysis, SRO interpretation is incorrect (see Devilee *et al,* 2001, for review). Complex patterns of 16q loss have been reported in breast tumours (Cleton-Jansen *et al,* 2001) supporting the hypothesis that multiple tumour suppressors map to this disease locus.

Laser capture microdissection (LCM) of breast tumours is a technique that minimizes the problems associated with normal cell contamination of tumour biopsies. Using this technique, the highest incidence of breast 16q LOH of 89% was recorded (Chen *et al*, 1996; Shen *et al*, 2000). LCM enables dissection of tumours such that the neoplastic epithelium can be separated from the surrounding cells. A recent study included the microdissection of sporadic invasive breast adenocarcinomas (Kurose *et al*, 2001). They reported a higher LOH in the neoplastic epithelial compartment when compared with the surrounding stromal cells. They concluded that LOH in neoplastic epithelial cells might precede LOH in the surrounding stromal cells, however both events are tightly linked.

1.5.1) 16q LOH and breast cancer.

40-60% of primary breast tumours exhibit 16q LOH (Devilee *et al,* 1994) and irrespective of the stage or histological phenotype, LOH is most frequently observed at 16q24.3 (Tsuda *et al,* 1994). A detailed study of 714 primary breast tumours identified three regions of frequent allelic imbalance, 16q22.1, 16q23.2-q24.1 and 16q24.3 (Cleton-Jansen *et al,* 2001). The region at 16q22.1 was defined by restricted LOH found in 4 tumours, while 25 tumours defined the 16q24.3 LOH region. The polymorphic marker *D16S498* and the 16q telomere defined the 16q24.3 LOH region. A previously constructed cosmid physical map of 800 kb spans part of this LOH interval, flanked by the *CDH15* gene and the microsatellite marker *D16S303* (Whitmore *et al*, 1998a). *D16S498* was estimated to map 2 Mb proximal of *CDH15*. *D16S303* was shown to define the boundary of transition from euchromatin to the subtelomeric repeats, defining the telomeric limit of the cosmid contig. No tumour suppressor loci have yet been identified in this region.

Mutation analyses on seven genes mapping to this interval has previously been conducted: *SPG7* (Settastian *et al*, 1999), *RPL13* (*BBC1*) (Moreland *et al*, 1997), *CDK10* (Crawford *et al*, 1999), *CPNE7* (Savino *et al*, 1999), *FANCA* (Cleton-Jansen *et al*, 1999), *GAS8* (*GAS11*) and *C16orf3* (Whitmore *et al*, 1998b). None of these genes were found to harbour any mutations in sporadic breast cancer DNA samples. Currently the 16q24.3 tumour suppressor gene remains elusive.

Frequent LOH at 16q has been observed in several other tumour types including, prostate (Suzuki *et al*, 1996), lung (Sato *et al*, 1998) hepatocellular (Chou *et al*, 1998), gastric (Yakicier *et al*, 2001), head and neck squamous cell carcinoma (Wang *et al*, 1999) and rhabdomyosarcoma (Visser *et al*, 1997). Detailed analysis of prostate carcinomas has revealed a restricted region of 16q24.3 LOH that overlaps with the breast tumour 16q24.3 LOH region. Therefore it is possible that this tumour suppressor gene may be common to both diseases.

The epithelial cadherin gene (*CDH1*) has been suggested as the LOH target gene at 16q22.1 (Berx *et al,* 1995; Huiping *et al*, 1999). This gene encodes a transmembrane glycoprotein important for the maintenance of calcium dependent cell-cell adhesion to control cell polarity and morphogenesis (Takeichi, 1991). There is convincing evidence that *CDH1* acts as a tumour suppressor gene in invasive lobular breast cancers (ILC) For example, *CDH1* exhibits tumour specific mutations in ILC (Berx *et al,* 1995) and loss of function contributes to increased proliferation,

invasion and metastasis (Huiping *et al*, 1999). Interestingly *CDH1* germline mutations have also been identified diffuse gastric cancer (Guilford *et al,* 1999).

However, despite these promising findings, no mutations have been identified in this gene in the more common sub-type of breast cancer, invasive ductal carcinoma (IDC) (reviewed in Cleton-Jansen 2002). Therefore ILC and IDC probably represent different forms of breast cancer, despite sharing similar patterns of 16q22.1 LOH. The IDC specific 16q22.1 tumour suppressor gene currently remains elusive.

1.6) The human genome project.

Over the past few years different cloning strategies have evolved in an effort to map and sequence the human genome. The initial aim was to construct physical maps consisting of overlapping genomic clones (contigs). These physical maps enabled the selection of suitable clones for subsequent sequencing. The first physical maps were based on cosmid clones containing relatively small genomic inserts (30-40 kb), but suffered from two major limitations The construction of cosmid-based physical maps were labour intensive and subsequent sequencing efforts were expensive due to vector sequence redundancy. The Yeast Artificial Chromosome (YAC) cloning vectors have large genomic inserts (> 1Mb) but themselves are notoriously unstable and not suitable for sequencing templates. Bacterial based cloning is now the preferred methodology as it offers both stability and large genomic insert sizes of 100-250 kb.

A BAC/PAC clone-based approach was adopted by the International Human Genome Sequencing Consortium (IHGSC) to construct the first draft of the human genome project (Golden path, International Human Genome Sequencing Consortium, 2001: 7). Initially individual clones were "fingerprinted" and assembled into overlapping contigs. Genomic clones representing the minimal tiling path were selected for shotgun sequencing and subsequently assembled. This hierarchical mapping and sequencing strategy enabled verification of the sequence integrity. However as the sequencing effort commenced prior to the completion of the detailed physical maps, initial clone selection and sequencing was partly at random. The data presented in this thesis has contributed to the IHGSC, and currently provides a short-term resource for gap closure. Sequence finishing efforts are likely to continue until an accurate detailed version of the human genome becomes available combining all information from the many physical, genetic and transcriptional maps.

A private commercial version of the human genome sequence has been generated and assembled using a whole genome shot gun approach (Venter *et al,* 2001). The "Golden path" was used to anchor and orientate data from these randomly sequenced clones in addition to identifying several sequence mis-assemblies (Katsanis *et al,* 2001). Currently the genome sequence consists of 3 billion base pairs and estimates of gene number range from 30,000 (Ewing and Green, 2000; Roest Crollius *et al,* 2000) to 75,000 (Wright *et al,* 2001). These reports loosely agree on 15,000 known genes and 30,000 transcripts with significant homologies to proteins of known function. The differences arise from estimates of novel genes, with *in silico* gene prediction programmes failing to identify many *in vitro* confirmed genes (Wright *et al,* 2001). Nevertheless, classical positional cloning strategies can now directly access the relevant genomic sequence data and identify all candidate genes.

1.7) Aims.

The minimum region for breast cancer LOH at 16q24.3 has been resolved to an interval between the polymorphic marker *D16S498* and the 16q telomere (Cleton-Jansen *et al,* 2001). A 800 kb cosmid physical map within this region has been previously constructed and is defined by the *CDH15* gene and the 16q telomere (Whitmore *et al,* 1998b). The first aim of this thesis is to extend this physical to encompass the entire 3 Mb 16q24.3 LOH interval. This map will be based on overlapping BAC and PAC clones, restricted and accurately assembled. Selected clones representing the minimal tiling path will be sequenced and assembled to enable candidate gene isolation. Preliminary *in silico* analyses of the genomic sequence data will permit putative gene identification and subsequent *in vitro* analyses will confirm their transcription status. The expression profiles of all candidates will be examined in breast cancer cell lines as it is hypothesized that tumour suppressor genes associated with the 16q24.3 LOH may be aberrantly expressed due to the presence of mutations or promoter hypermethylation. This expression variability screen will enable prioritisation of candidate transcripts for subsequent investigations.

Candidate breast cancer TSGs will be assessed based on homologies or possible links with established genes that are associated with breast carcinogenesis. Assessment of selected candidates will involve the isolation of full-length transcripts, detailed *in silico* analysis and the search for breast tumour restricted mutations.

Preliminary functional analyses will be performed on selected candidates to confirm their predicted function and explore any possible link with cancer.

This project was a collaborative effort between a number of researchers to allow the sequencing and transcript analysis of the entire 16q24.3 LOH interval. The research not undertaken by the candidate is noted at relevant places in this thesis. During the period of candidature the HGP was rapidly progressing and this publicly available sequence data is integrated into the sequenced physical map presented.

Chapter 2

2.1 Introduction.

Frequent LOH at 16q has been observed in several tumour types including breast (Cleton-Jansen *et al*, 2001), prostate (Suzuki *et al*, 1996), lung (Sato *et al*, 1998) hepatocellular (Chou *et al*, 1998), gastric (Mori *et al*, 1999), head and neck squamous cell carcinoma (Wang *et al*, 1999) and rhabdomyosarcoma (Visser *et al*, 1997). These LOH intervals have been proposed to contain tumour suppressor loci. 16q LOH ranges between 28-89% in previnvasive ductal breast tumours (Chen *et al*, 1996) and 40-60% in primary breast tumours (Devilee and Cornelisse, 1994). A detailed study of 714 primary breast tumours identified three regions of frequent allelic imbalance, 16q22.1, 16q23.2-q24.1 and 16q24.3 (Cleton-Jansen *et al*, 2001). Irrespective of disease stage or histological phenotype, LOH is most frequently observed at 16q24.3, suggesting a key role for this region in breast tumourigenesis.

The minimum region of breast cancer LOH at 16q24.3 has been resolved to an interval between the polymorphic marker *D16S498* and the 16q telomere (Cleton-Jansen *et al*, 2001). A cosmid physical map of 800 kb has previously been constructed and spans part of this LOH interval, flanked by the *CDH15* gene and the microsatellite marker *D16S303* (Whitmore *et al*, 1998a). *D16S498* was estimated to map 2 to 3 Mb proximal of *CDH15*. *D16S303* was shown to define the boundary of transition from euchromatin to the subtelomeric repeats, defining the telomeric limit of the contig. No tumour suppressor loci have yet been identified in this region.

The development of BAC and PAC genomic libraries made it possible to rapidly extend the cosmid contig to encompass the total 16q24.3 LOH interval. Each of these libraries offered five to six times coverage of the entire human genome with average genomic insert sizes ranging between 160 and 235 kb (Osoegawa, *et al,* 1998). A detailed physical map was constructed, with individual clones and genes precisely mapped using *EagI* restriction fragments. A cloned based walking approach was used to extend the map until all contigs were linked together to form a 2.4 Mb physical map. This map enabled a minimal tiling path of 26 clones to be selected for large-scale sequence analysis. This sequence was used to orientate, order and assemble all publicly available sequence data from the first draft of the human genome project (Golden path, International Human

Genome Sequencing Consortium, 2001). *In silico* and *in vitro* analyses identified 104 candidate genes in this region.

We have previously carried out mutation analyses on seven genes mapping to this interval: *SPG7* (Settastian *et al*, 1999), *RPL13* (*BBC1)* (Moreland *et al*, 1997), *CDK10* (Crawford *et al*, 1999), *CPNE7* (Savino *et al*, 1999), *FANCA* (Cleton-Jansen *et al*, 1999), *GAS8* (*GAS11*) and *C16ORF3* (Whitmore *et al*, 1998b). None of these genes were found to harbor any mutations in sporadic breast cancer DNA samples. It was impractical to use this labor-intensive procedure to screen all the remaining candidates and furthermore, tumour DNA stocks were limited.

As an independent selection criterion for the identification of tumour suppressors, we determined the expression profile of each transcript in a panel of breast cancer cell lines. Mutations in TSGs may result in aberrant mRNA expression due to nonsense-mediated mRNA decay (Rajavel and Neufeld 2001), disruption of promoter function (Giedraitis *et al* 2001), or alteration of mRNA stability (Wieland *et al*, 1999). In addition, promoter hypermethylation has been recently demonstrated to have a central role in the down-regulation of tumour-suppressor genes (Yuan *et al*, 2001). Consequently, it was reasonable to hypothesize that tumour-suppressor genes associated with the 16q24.3 LOH may be aberrantly expressed due to the presence of mutations or promoter hypermethylation. Classical mutational analyses have failed to evaluate the methylation status of the candidate transcripts. The expression profile of all the 16q24.3 transcripts was assessed in a panel of non-tumour and breast cancer cell lines containing 16q24.3 LOH. Several of these genes were found to exhibit aberrant mRNA expression. The notion of these genes being breast cancer tumour suppressors is compatible with what is known about their respective function.

2.2 Materials and Methods

2.2.1 Physical mapping

2.2.1.1) Standard PCR amplification. PCR reactions were performed in 10 µl volumes containing 1X PCR buffer (Invitrogen), 2 mM dNTPs, 1.5 mM $MgCl_2$, 1 mM of each primer, 1 unit of Taq polymerase. Template DNA amounts ranged from 30-50 ng for genomic DNA, 10 ng for clone DNA and single bacterial colonies suspended in paraffin oil for colony PCR. PCRs were performed in Eppendorf thermal cyclers with the standard cycle conditions of 94°C for 2 minutes, followed by 35 cycles of 94°C for 30 sec, 60°C for 1 minutes, 72°C for 2 minutes, followed by a final elongation step of 72°C for 10 minutes.

2.2.1.2) Plasmid, BAC and PAC DNA preparation. All DNA preparations were based on the alkaline lysis procedure adapted from Sambrook *et al*, (1989). cDNA clones, PAC and BAC clones received as bacterial stabs were streaked for single colony on L-ampicillin or L-kanamycin or L-chloramphenicol plates respectively. 200 ml cultures (L-Broth) were inoculated and incubated (150 rpm) at 37°C for 16 hours, containing either ampicillin (100 µg/ml) or kanamycin (50 µg/ml) or chloramphenicol (34 µg/ml). Cultures were harvested by centrifugation at 5,000g for 10 minutes, and DNA extracted using Qaigen buffers and Tip-100 columns. Slight modifications were made to the manufacture's instructions, including an extra centrifugation step at 20,000g to remove cell debris and warming the elution buffer to 65°C for superior recovery. DNA pellets were resuspended in Tris/EDTA and quantitated by spectrophotometry. All plasmid DNA preparations were isolated with similar methodologies except the cultures were reduced to 10 ml and Qiagen Tip-20 columns were used.

2.2.1.3) Preparation of bacterial glycerol stocks. All bacterial clones were maintained as glycerol stocks that were prepared from a 10 ml L-Broth overnight culture, spun at 4,000g and resuspended in 1 ml of L-Broth containing 15% glycerol. These stocks were kept at –70°C until more DNA was required, whereby a 5 µl aliquot was used to streak for single colonies on plates containing the appropriate antibiotic.

2.2.1.4) Oligonucleotide primer design. Oligonucleotide primers for PCR amplification of STSs and sequencing were designed using the Lasergene Primer Select software (DNASTAR). In the case of ESTs, primers were designed to span exon-intron boundaries. All primer sequences used for transcript amplification are presented in Table 1.

2.2.1.5) Digestion of genomic, BAC and PAC DNA. 10 µg of total human DNA, mouse/human somatic cell hybrid DNA (containing chromosome 16 as the only human component) and the hybrid A9 DNA (mouse background), were digested with 20 units of *Pst I* (New England Biosciences) for 10 hours at 37°C. For standard BAC and PAC digests, 0.5µg of DNA was digested with 5 units of restriction enzyme for 3 hours at 37°C. Double digests of clones were performed in buffers compatible for both enzymes.

2.2.1.6) Agarose gel electrophoresis. Electrophoresis of genomic DNA was performed in 0.8% (w/v) agarose and run at 18 mA per gel overnight. BAC, PAC and plasmid DNA was resolved in 0.8% agarose at 100 volts for 1 h. Analyses of PCR products ranging in size from 80-500 bp were performed in 1.5% agarose (w/v). 1X agarose loading buffer was added to each sample and all gels were run in 1X TBE. The gels were stained with ethidium bromide (0.02%) and the DNA visualized under UV light.

For pulse-field mapping experiments, 1µg of PAC or BAC DNA was digested with *Eag I* for 3 h at 37°C. 100ng of this digested DNA was resolved on a 1% agarose gel in 0.5X TBE. Electrophoresis was performed with a BioRad chief pulse mapper, set with standard parameters including a resolution range of 5-50 kb.

2.2.1.7) Southern blot analysis. The Southern analysis of genomic DNA, including somatic cell hybrid DNA was performed in 10X SSC, essentially as described by Southern (1975). Digested PAC, BAC and plasmid DNA was blotted in 0.4 M NaOH. GeneScreen Plus nylon membrane was used for transfer and the DNA was cross-linked by heating for 1 h at 60°C. Membranes were pre-wet in 5XSSC, placed into glass bottles and pre-hybridized with 10 ml of Southern hybridization solution for 2 h at 42°C The DNA probe was labeled (2.2.1.8), denatured for 5 minutes at 95°C and added directly to

the filters. Probes were hybridized overnight at 42°C and the following day membranes were washed three times in 2XSSC and 1% SDS for 10 minutes at room temperature. A high stringency wash of 0.1% SSC and 1% SDS at 65°C for 30 minutes was used when direct filter assessment suggested high background signal.

Southern membranes were stripped of the radiolabeled probe to allow for additional hybridizations. Membranes were incubated in 0.4 M NaOH for 30 minutes at 42°C and neutralized in 0.1 % SDS, 0.1 % SSC, 0.2 M Tris-HCL (pH 7.5) for 15 minutes. Membranes were re-probed up to 10 times.

2.2.1.8) ^{32}P radiolabeling of probes. A standard labeling reaction contained 50 ng of double-stranded DNA made up to 34 µl with sterile water, and incubated at 94°C for 5 minutes. After a brief centrifugation step, 5 µl of 10X dNTP labeling solution, 5 µl of 10X labeling buffer, 5 µl of [α-^{32}P} dCTP (50 µCi), and 5 units of *E. coli* DNA polymerase I were added. This was incubated for 30 minutes at 37°C. The reaction was terminated with the addition of 1 µl of 0.5 M EDTA, and if required, un-incorporated ^{32}P was removed using QIAquick columns (Qiagen). DNA probes that required blocking of repetitive sequences were pre-reassociated with 500 µg of sonicated human placental DNA (50 µl of 5mg/ml) in a final concentration of 5XSSC. Subsequently the probe/pre-reassociation mix was denatured for 5 minutes at 94 °C, cooled on ice for 1 minute and incubated for 1 h at 65 °C. The probe was then added directly to the pre-hybridized membranes. Probes not requiring pre-blocking were heated at 94 °C for 5 minutes before addition to pre-hybridized filters.

2.2.1.9) Sub-cloning DNA fragments. All DNA ligations were performed in a volume of 10 µl at 14°C for 1-16 h. Ligations involved varying the molar ratios of vector to insert from 1:1 to 1:4, as essentially described by Sanbrook *et al*, (1989). Each reaction contained 50 ng of vector DNA and 5 units of T4 ligase (NEB), as recommended by the manufacture's instructions. pGEM-T (Promega) cloning of PCR products was facilitated by the addition of adenosine residues to the 3' end of all *Taq* generated PCR products. Unique restriction enzyme sites were engineered into the 3' end of several primer pairs for directional cloning. Following amplification, the PCR products were verified by

direct sequencing. Finally, the PCR products and the destination vectors were digested and ligated together.

2.2.1.10) Preparation of competent bacterial cells. The *E. coli* strains XL-1 blue and DH5α were streaked for single colonies on L-tetracycline plates, growth overnight at 37°C and then used to inoculate a 10 ml L-Broth plus tetracycline (15 μg/ml) overnight culture. The next day, 50 ml of L-Broth plus tetracycline was inoculated with 1 ml of the overnight culture and grown until an A_{600} of 0.3 was reached, approximately 2 h. The culture was spun at 2,000g for 5 minutes and resuspended in 5 ml of ice-cold TSS media. The cells were left on ice for 10 minutes before being either transformed (2.2.2.11) or aliquoted and frozen at –70 °C.

2.2.1.11) Bacterial transformation. For each transformation, 50 μl of competent cells was mixed with either 1μl of the ligation (approximately 5 ng of DNA) for directional cloning or 5 μl of ligation (approximately 25 ng of DNA) for blunt-ended cloning. The cells were incubated on ice for 20 minutes and then heat shocked at 42°C for 30 seconds. The cells were returned to ice for 2 minutes before the addition of 450 μl of SOC medium. Cells were incubated at 37°C for 30 minutes and a 100 μl aliquot was spread on to L-plates with the appropriate antibiotic for selection of successful transformants. For vectors with blue/white colour selection, 40 μl of X-Gal (20 mg/ml) and 20μl of IPTG were spread with the cells and plates were incubated overnight at 37°C.

2.2.1.12) DNA sequencing. Two different DNA sequencing approaches have been utilize in this study, a large scale sequencing effort to sequence the 16q24.3 BAC/PAC minimal tiling path (2.2.2.1) and the sequence isolation and verification of various PCR products and cDNA clones used in physical map construction and consolidation. This latter approach entailed adding 250 ng of template DNA, 1.6 pmol of sequencing primer, 4μl of BigDye (version 2) and 4μ1 halfTERM dye terminator sequencing reagent, in a total volume of 20 μl. The reactions were thermo-cycled (25 X [96°C for 30 sec, 50°C for 15 sec, 60°C for 4 minutes]) and then the products were cleaned with isopropanol. In brief, reactions were mixed with 80μl of 70% isopropanol for 30 minutes at room temperature.

The tubes were spun at 13,000 rpm for 30 minutes and the pellets washed in 1 ml of 70% ethanol. Pellets were air dried and sequenced at an external sequencing center.

2.2.1.13) Probes generated for initial PAC/BAC identification. A cosmid physical map of 16q24.3 has previously been constructed and spans approximately 800 kb, flanked by the *CDH15* gene and the telomere (Whitmore *et al*, 1998a). Initial markers chosen to extend this map were those known to be located between *CDH15* and the somatic cell hybrid breakpoint CY2/CY3 (Callen *et al,* 1995). These included two genes *APRT* and *CAV,* the microsatellite markers *D16S413, D16S2621*, and *D16S3121* and the Unigene cluster Hs.7970, originally obtained from Gene map 99 (1) using the radiation hybrid panel Genebridge 4(GB4) server (2). All marker locations were confirmed by Southern blot analysis on a panel of mouse/human somatic cell hybrids (Callen *et al*, 1995). 10 μg of total human DNA, somatic cell hybrid DNA (containing chromosome 16 as its only human component) and the hybrid A9 DNA (mouse background) were digested for 10 h with *Pst I* (New England Biosciences), resolved by agarose gel electrophoresis (2.2.1.6) and transferred to nylon membrane (2.2.1.7). Membranes were probed with ^{32}P labeled PCR products (2.2.1.10) generated from either conventional genomic amplification of appropriate microsatellites (2.1.1.1), or from RT-PCR amplification of transcribed sequences mapping to the region (2.1.1.11).

2.2.1.14) Human PAC and BAC library screening. In total four different sets of BAC and PAC libraries were utilized in the mapping effort. One PAC library and the Human Release II BAC library filters were obtained from Genome Systems (St. Louis, Missouri, USA) while the RPCI-4 PAC and RPCI-11 (segment 3) BAC library filters were obtained from BAC/PAC Recourses (Rosewell Park Cancer Institute, New York, USA). These genomic clones were spotted in high-density arrays, 8 filters each containing 18,432 clones. On the filters, clones were grouped into sub-sets of 16 clones, 8 each in duplicate. Genuine positives were indicated by the hybridization to both clones within the single sub-set. Single dots were excluded as false positives. Each library provided 4-6 times coverage of the male human genome.

Each filter within a library was hybridized and washed according to the recommendations of the manufacture. Initially, membranes were individually pre-hybridized in large glass bottles for at least 2 h at 65°C in 20 ml of solution A (6X SSC, 0.5% SDS, 5X Denhart's, 100µg/ml denatured salmon sperm DNA). ^{32}P labeled probes were added to 20 ml of solution A and hybridized overnight at 65 °C. Filters were washed in 2X SSC, 0.5% SDS for 5 minutes at room temperature, 2XSSC, 0.1% SDS for 15 minutes at room temperature and 0.1XSSC, 0.5% SDS for 1 h at 37°C if needed. Additional *in silico* screening of the high throughput genomic sequence (htgs) database at the National Center for Biotechnology Information (NCBI (3)) was performed. BLAST analyses of all known mapped sequences within this genomic interval were used to isolate additional overlapping PAC and BAC clones, deposited at NCBI by other sequencing centers.

2.2.1.15) Sorting of clones. The genomic libraries were screened with multiple probes to maximize the use of the library, and to accelerate map construction. Agar stabs for the positive clones were obtained from the supplier of the library and sorted by colony hybridization. This enabled the sorting of individual probes with their corresponding positive clones. Single colonies representing the various clones were spotted onto master plates containing the appropriate antibiotic and grown overnight at 37°C. The next day the plates were chilled to 4 °C and covered with nylon filters for 5 minutes to allow colony transfer. The filter was removed and soaked in denaturing solution for 5 minutes, twice in neutralization solution for 5 minutes each, and once in 5X SSC for 10 minutes. Filters were heated for 1 h at 60°C before they were hybridized. This same master plate was used for the generation of two additional replica filters.

All probes used on the BAC/PAC filters were known to map between the 16q telomere and the somatic cell hybrid breakpoint CY2/CY3. Occasionally two or more markers hybridized to the same genomic clone indicating that such markers were physically adjacent. As map construction progressed, all new potential probes and database entries that emerged from the human genome project were screened against the established BAC/PAC contig by colony hybridization. This enabled the linkage of small physically adjacent contigs, ultimately resulting in a contiguous 2.4 Mb physical map.

2.2.1.16) Isolation of BAC and PAC ends. PAC and BAC DNA was isolated (2.2.1.2) and directly sequenced from the T7 and SP6 priming sites within the BAC and PAC vector backbone. These priming sites are positioned approximately 50 bp upstream of the large genomic insert. BigDye sequencing was performed as described earlier (2.2.1.12) except the amount of template DNA was increased to 0.5-1.0 μg and the number of amplification cycles was increased from 25 to 80. High quality template DNA was imperative and mean read-lengths of 600 bp were achieved. Primer pairs were designed to these BAC and PAC ends (2.2.1.4), in regions of non-repetitive DNA, and amplified from either total human genomic DNA, or from the original parental BAC or PAC clone (2.2.1.1). Clone ends were radio-labeled (2.2.1.8) and hybridized to various pulse field blots (2.2.1.7). Several of these fragments were shown to represent the extreme ends of the individual contigs and consequently selected as probes in subsequent rounds of library screens.

2.2.1.17) Physical map construction consolidation. The data generated from PAC and BAC pulse field gel electrophoresis experiments and the accompanying PCR and Southern based mapping approaches, has allowed the construction of a physical map with ordered *Eag I* restriction fragments. Several of the *Eag I* restricted PAC and BAC DNA fragments (2.2.1.5) either didn't conform to the current restriction fragment order, or exhibited unique restriction sizes. These fragments were excised from the pulse field agarose gels with sterile blades, melted at 80°C, and radio-labeled (2.2.1.8) for hybridization back to the physical map blots (2.2.1.7). These fragments proved to be either clone-end fragments, or polymorphic restriction fragments.

2.2.2 Sequence analysis and transcript identification

2.2.2.1) Sequence analysis. 26 BAC, PAC and cosmid clones were selected as the minimal tiling path of the 16q24.3 LOH interval. Dr Gabriel Kremmidiotis and Ms Alison Gardner performed a large scale sequencing effort on these selected clones and their methodologies are summarized.

DNA was extracted from each clone in the minimal tiling path, and used to construct random puc-18 sub-libraries. Single colonies were isolated, grown for DNA extraction and sequenced on ABI377 sequencers. The genomic sequence data was assembled and analysed using PHRED, PHRAP (Ewing *et al*, 1998) and GAP4 (Staden *et al*, 2000) software on a SUN workstation. The sequence generated was used to anchor other draft sequence data deposited to GenBank by other sequencing centers.

2.2.2.2) In silico analysis. Myself, Dr Gabriel Kremmidiotis, Ms Alison Gardner, and Ms Suzan Hinze performed these analyses.

The genomic sequence data generated was used to identify matching and/or homologous nucleotide and protein sequences. This involved the use of the BLASTn and BLASTp (Altschul *et al,* 1997) algorithms to search databases at NCBI. The algorithm LALIGN was used for global alignments between two nucleotide sequences. LALIGN parameters were set at default values (4). The Clustal algorithm was used to align multiple protein sequences. The GENSCAN program was used to (5) predict potential exons and genes from the genomic sequence data. Due to the presence of repeats throughout the human genome, all large genomic sequences were masked for repetitive stretches, before subsequent BLAST analyses were performed. The assembled genomic sequence was analysed in 100 kb sections, ensuring the identification of all potential transcripts.

2.2.2.3) Nomenclature. Genes previously described in the literature or in the public databases are referred to using the symbols approved by the Human Gene Nomenclature committee (6). Novel genes are referred to using the Unigene database code number as found at NCBI. Singleton ESTs are referred to by their GenBank accession numbers.

2.2.2.4) Cell lines. The breast cancer cell lines: BT-549, CAMA-1, MDA-MB-134, MDA-MB-157, MDA-MB-231, MDA-MB-436, MDA-MB-468, SKBR-3, T47-D, ZR-75-1 and ZR-75-30 were purchased from the American Type Culture Condition (ATCC) and cultured in RPMI or OPTI-MEM medium (Invitrogen) supplemented with 10% fetal calf serum (FCS). Additional culture supplements were used as recommended by ATTC. These cell lines possess LOH involving chromosome 16 (Callen *et al*, 2002). The cell line MCF12A was derived from normal breast epithelium and was included in our studies as a normal control.

2.2.2.5) RT-PCR amplification of cDNA clones. RT-PCR analyses of the *in silico* identified gene signatures was carried out on total RNA from 21 human tissues mixed into six pools: pool 1, brain, heart, kidney, liver; pool 2, lung, trachea, colon, bone marrow; pool 3, spleen, thymus, prostate, skeletal muscle; pool 4, testis, uterus, fetal brain, fetal liver; pool 5, spinal cord, placenta, adrenal gland, salivary gland, fetal lung; pool 6, breast (BD Biosciences, Stratagene, Ambion). Reactions were performed with Superscript RNase H- reverse transcriptase (Invitrogen). In brief, 3 µg of total RNA or 100 ng of polyA$^+$ mRNA was added to either 50 pmol of random hexamers (Perkin Elmer) or 100 pmol of oligo dT. The RNA was heated to 65 ^{o}C for 5 minutes, returned to ice for 1 minute before the addition of 1X PCR buffer, 4 mM DTT and 0.2 mM dNTPs. This was incubated at 42^{o}C for 2 minutes before the addition of 200 units of Superscript RNase H- reverse transcriptase, followed by a final incubation at 42^{o}C for 30 minutes, and the reaction terminated with heating to 70^{o}C for 10 minutes. The secondary cDNA amplification was performed as earlier described (2.2.1.1) with 1-20 ng of cDNA template.

Breast cancer cell line RNA was required for quantitative mRNA expression analysis (2.1.2.6) of all the genes identified in the 16q24.3 LOH interval (2.2.2.5). The isolation of RNA from breast cancer cell lines was performed under RNAse-free conditions. All solutions were treated with 0.2% (v/v) DEPC and autoclaved, filtered pipette tips were used and gloves worn. RNA was isolated using a procedure based on Chomczynski and Sacchi (1987) using the TRIzol reagent (Invitrogen), according to the manufacture's instruction. In brief, 1X10^7 cells grown in 175 cm^2 flasks, washed twice with 10 ml of

PBS, removed with a sterile plastic scraper and spun at 500g for 5 minutes. The cell pellet was resuspended in 500 µl of TRIzol, incubated at room temperature for 5 minutes before an equal volume of chloroform was added, incubated on ice for 5 minutes and spun at 13,000 rpm for 15 minutes at 4°C. Supernatant containing total RNA was precipitated with 100% isopropanol for 2 hours at -20 °C. The RNA was spun at 13,000 rpm for 15 minutes at 4°C, washed in 70% ethanol and resuspended in DEPC treated water. A second precipitation step involving the addition of one-twentieth the volume of 4 M sodium acetate and 2 volumes of 100% ethanol was performed, followed by a 70% ethanol wash and re-suspension in DEPC treated water.

DNA contamination was removed from all RNA preparations using RNAse-free DNase (Ambion). The absence of contaminating DNA by using primers that only amplify from intronic DNA (primers from the CHRNA4 locus: 5' CTGGAGATGTTTGTGGCCTT and 5' TGCTTCACCCATACGTC). This primer set failed to amplify from all the cDNA samples used in our study. RNA concentrations were estimated using RiboGreen (Molecular Probes). All RNA stocks were stored at -70 °C.

2.2.2.6) Real time PCR and data normalization. Mr. Anthony Bais and Ms Suzan Hinze performed this analysis, and below I have summarized the methodologies.
Real-time PCR was performed on a Rotor-Gene 2000 (Corbett Research). Reactions (25 µl) included 12.5 µl SYBR Green 1 PCR Master Mix (PE Biosystems), 0.2 µM of each primer, and 30 ng cDNA template. Amplification cycling was performed as follows: 94°C for 10 minutes, followed by 45 cycles of 93°C for 20 sec, 60°C for 30 sec, and 70°C for 30 sec. Fluorescence data acquisition was at 510 nm during each 70°C extension phase. Melt curve analysis and agarose gel electrophoresis was performed following each real-time PCR run to assess product specificity. For real-time PCR quantification we used internal standard curves as described (Bustin, 2000). mRNA copies per nanogram of total RNA used were calculated by the cycle threshold value (CT) extrapolation to standard curves constructed with serial dilutions of known concentration templates. Copy numbers per cell were calculated based on a total-RNA/cell value of 4 pg. Expression levels of the housekeeping genes; *PPIA* (peptidylprolyl isomerase, cyclophilin A), *GAPD* (*GAPDH*, glyceraldehyde-3-phosphate dehydrogenase), *RPB70* (RNA polymerase subunit II), and

ATP5A (ATPase subunit II) were analysed in breast cancer cell lines to determine the most accurate endogenous control for data normalization. *PPIA* displayed the most uniform expression profile and was used to normalize the results obtained from all the genes studied. Fold changes in expression for each cell line were calculated using the expression levels of normal breast as a reference. The range of relative fold changes in expression between the panel of 12 cell lines was calculated as a single value for each transcript. The term relative fold variability index (RFVI) was chosen to describe this value.

2.3 Results

2.3.1 Physical map construction.

2.3.1.1 The chromosome 16q somatic cell hybrid panel. A mouse/human somatic cell hybrid panel has been previously constructed by fusing human cell lines with the mouse cell line A9 (Callen *et al,* 1995). The hybrids containing human chromosome 16q were selected with the *APRT* gene at 16q24.3. In general these hybrids were derived from translocations or interstitial deletions and contained regions of chromosome 16 from the breakpoint to the 16qter.

The proximal and distal limits of the 16q24.3 LOH interval are defined by the somatic cell hybrid breakpoints CY2/CY3 and CY18YA respectively. The CY18A somatic cell hybrid was derived from a complex interstitial deletion and contains two fragments of chromosome 16, one at 16q24.3 and the other at 16q11.2-q13. The hybrids CY2/CY3 were derived from reciprocal re-arrangements between 16q24.3 and Xq26. This hybrid panel was used to precisely map all the markers employed for physical map construction.

2.3.1.2 Probe generation for PAC and BAC genomic library screening. A total of six markers including the genes *APRT* and *CAV,* the microsatellite markers *D16S413*, *D16S2621*, and *D16S3121*, and the Unigene cluster Hs.7970, were all initially mapped between the somatic cell hybrid breakpoints CY2/CY3 and CY18A. These hybrid breakpoints demark the genomic interval proximal to the pre-existing 800 kb cosmid

physical map of 16q24.3 (Whitmore *et al*, 1998a). These markers were mapped to the somatic cell hybrid panel by either PCR or Southern blot analysis. For PCR analysis, markers were amplified from various templates including somatic cell hybrid DNA, human genomic DNA and A9 cell line DNA, the latter serving as a mouse background control. Primer pairs were designed to amplify products ranging in size from 100 to 1000 bp. Figure 1 shows amplification and mapping of the *CYBA* gene. This 244 bp PCR product maps between the somatic cell hybrid panel breakpoints CY2/CY3 and CY18A, and consequently was used to screen the BAC and PAC genomic libraries. Alternatively, markers were mapped by Southern hybridization to the somatic cell hybrid DNA panel. Figure 2 is a representative Southern blot showing the mapping of the *APRT* gene. Figure 1 and 2 represent the mapping procedures routinely performed on all markers generated throughout map construction. This continual assessment of individual markers ensured that all chromosomal walks were restricted to 16q24.3.

2.3.1.3 Identification of BAC and PAC clones. The BAC and PAC end fragments that were shown to represent the extreme ends of the individual contigs were used to screen the BAC and PAC genomic libraries. The availability of several different genomic libraries enabled the generation of a contiguous physical map. Moreover, probes that failed to identify any positive clones from one particular library were screened against a second, and sometimes a third library, to obtain the corresponding clone. Figure 3 is a representative autoradiograph of a high-density BAC filter probed with the *GALNS* gene. Each BAC clone is represented twice within each subset, and genuine positives are indicated by hybridization to both clones. Single dots were excluded as false positives. As a consequence of each library offering 4-6 times coverage of the male human genome, and the pooling of multiple probes for each hybridization, many (3-30) positive clones were recovered from each screen. Positive clones were sorted by colony hybridization experiments. In summary, the initial genomic screen established 2 small contigs, one anchored by the genes *GALNS* and *APRT*, the other by the *CYBA* gene. These small contigs served as nucleation points for the subsequent physical map construction.

In addition to this manual screening of the PAC and BAC genomic libraries, *in silico* screening was also performed. BLAST analysis of the High Throughput Genomic

Sequence (htgs) and the Genomic Survey Sequence (gss) databases at NCBI identified several clones that had been deposited by other sequencing centers. These clones were subsequently obtained and integrated into the physical map.

2.3.1.4 Restriction enzyme analysis and PAC and BAC walking. The degree of overlap between clones was determined by restriction enzyme analysis. The development of the BAC and PAC vectors has allowed for the stable and efficient cloning of large genomic inserts (100-250 kb), greatly reducing the number of clones that would be required to construct an equivalent sized cosmid map. These large genomic inserts generate complex restriction fragment patterns making subsequent map assembly challenging. *Eag I* was chosen for restriction analyses, as this enzyme's 8 base pair recognition sequence produced a range of restriction fragments allowing easy visualization and identification of BAC/PAC overlaps.

Pulse field gel electrophoresis of *Eag I* restricted PAC and BAC clones (Fig 4) permitted the resolution of restriction fragments ranging in size from 1 to 50 kb. Restriction fragments less than 1 kb were mapped by normal electrophoresis in 2.5% agarose. The resulting band profiles were used for accurate size determination and preliminary map construction.

The restriction fragments representing the extreme ends of PAC and BAC clones served to extend each contig through subsequent rounds of genomic library screening. The availability of the *Eag I* restriction map made it possible to select PACs and BACs believed to represent the extreme ends of the contig. These clones were end sequenced from both the T7 and SP6 priming sites. This was necessary as the individual genomic inserts were cloned in a non-directional manner. This sequence data permitted exact clone positioning within the physical map. In some cases the clone end sequence was repetitive and all primer combinations failed to amplify specific products. The majority of end fragments were amplified from the original BAC or PAC clone and hybridized back to blots containing the restricted contig. Each end probe generally hybridized to a unique restriction fragment from the original clone itself in addition to restriction fragments present in the adjacent overlapping clones, and thus contained within the established contig (Fig 5A). End probes that hybridized to the parental clone alone were

identified as the extremity of the contig and used in subsequent BAC and PAC library screens (Fig 5B). Occasionally the BAC end sequencing identified cDNA clones or STSs present in the databases that had not been previously mapped to this interval. These markers were then used in subsequent library screenings. This screening procedure was repeated until all the contigs were eventually linked together, forming a contiguous BAC and PAC physical map. The PAC and BAC minimal tiling path of this expanding physical map was spotted onto filters and used to screen all new emerging end sequences and other probes for possible linkage to the established contigs. This proved successful for the identification of preliminary contig overlaps, which were subsequently subjected to detailed pulse field mapping experiments.

Eag I restriction fragment analysis revealed that some clones exhibited fragment sizes that conflicted with the restriction pattern of overlapping PAC and BAC clones. These bands were excised from pulse-field agarose gels, the DNA isolated and hybridized back against these same contigs. The variations in band sizes were a result of *Eag I* restriction fragment polymorphisms. More than 20 *Eag I* polymorphic restriction fragments were identified throughout this 2.0 Mb interval (Fig 6: individual restriction fragment sizes are shown).

2.3.1.5 Physical map summary. The previously described physical map bounded by the 16q telomere and the *CDH15* gene (Whitmore *et al*, 1998a) was extended by 1.8 Mb to the marker *D16S3028* (Fig 6). This contiguous physical map contains 63 PAC and 102 BAC clones, isolated from 38 different library screens and using a total of 87 probes.

The *Eag I* restriction fragments from each PAC and BAC clone were ordered and assembled into contigs, and subsequently verified by Southern analysis (Fig 6: vertical dashed lines). Several restriction fragments that were not anchored by a marker were excised from the pulse field gel and hybridized back to the original contig. The majority of the fragments analysed conformed to the physical map, however some restriction fragment sizes were inconsistent with the overlapping clones. For example, BAC 654O13 was shown to map at 16q24.3 (PCR data not shown) and contained a 65 kb unique *Eag I* restriction fragment, that conflicted with the restriction fragment order of several overlapping clones. Subsequent fluorescence *in situ* hybridization (FISH) experiments

confirmed that this BAC clone was a chimera and contained segments originating from both chromosome 16 and chromosome 8 (Elizabeth Baker, personal communication). These rearrangements were probably a result of ligation artifacts, as the BAC and PAC cloning vectors are known for their stability.

The proximal limit of the physical map was defined by the genetic marker *D16S3074* contained within BACs 321A1 and 77K21 (Fig 6). PCR based mapping of *D16S3074* revealed that this marker mapped proximal to the somatic cell hybrid breakpoints CY2/CY3, that is CY3 positive and CY2 negative (data not shown). The closest neighboring marker, *D16S3028,* maps distal to this breakpoint, and thus defines the physical location of this hybrid breakpoint (Fig 6). The physical map was linked to the previously described cosmid contig (Whitmore *et al*, 1998a) (Fig 7). *D16S303* defines the telomeric limit of this cosmid contig, the boundary of transition from the euchromatin to the subtelomeric repeats of 16q24.3.

The BAC and PAC libraries used for map construction each offered four to six fold coverage of the human genome. The 16q24.3 physical map exhibits similar average clone coverage but with some areas of both under- and over-representation. In areas of under-representation, some clones exhibited only small regions of overlap and did not share any common restriction fragments with the surrounding clones. In these instances, clone end fragments were isolated and hybridization to the back to the contig, confirming this minimal overlap.

The comprehensive clone coverage of this genomic interval has allowed for the selection of a minimal tiling path of 26 clones demarked by the *D16S3074* and the 16q telomere. The minimal tiling path contains cosmid, BAC and PAC clones (Fig 6 and 7: blue clones), with an average overlap of 14 kb, and ranging form 490 bp to 49.7 kb. Small clone overlaps were highly desirable as this minimized sequence redundancy and permitted greater sequence coverage of the region.

The 26 clones representing the minimal tiling path were individually mapped by fluorescence *in situ* hybridization (FISH) to normal metaphase chromosomes (Elizabeth Baker, personal communication). All clones specifically hybridized to the 16q24.3 chromosomal interval (data not shown) and these experiments served as a final check of clone integrity before they were selected for sequencing. Two of the clones, 863P13 and

21A4, hybridized to 16q24.3 and 16p11.2, indicating the presence of a chromosome 16 specific duplication or repeat. The end fragments of both 863P13 and 21A4 were mapped by to this interval by PCR (data not shown), but proved to be inadequate for library screening as they predominantly identified clones from 16p11.2. This duplicon has previously been identified and shown to contain the gene BANP (Birot *et al*, 2000).

2.3.2 Large-scale sequence analysis of the minimal tiling path. Dr Gabriel Kremmidiotis and Ms Alison Gardner performed these experiments and below I have summarized their results. This data was included since it forms the basis of subsequent work in this thesis.

The 26 clones representing the minimal tiling path served as the template for large-scale sequence analysis. The individual contigs were eventually orientated, ordered and linked together, with average sequence coverage of 2-3X. This sequence was used to identify and integrate sequence data from 13 clones (Fig 7: green clones) available from the first draft of the human genome project (Golden path, International Human Genome Sequencing Consortium, 2001). Our sequence data helped to map and orientate clones from the Golden path, in addition to identifying several sequence mis-assemblies. The availability of the physical map with ordered Eag 1 restriction fragments made it possible to orientate contigs and simplify sequence assembly when large regions of repetitive sequences were present. The combined sequence of 2.4 Mb stretches from the marker *D16S3028* to the telomere. The sequence coverage consists of 5 contiguous clones, 13 clones in 5 or less contigs, and 10 clones in seven or more contigs. Only two clones between the genes *BANP* and *SLCA5* were sparsely sequenced. There are three sequencing gaps spanned by the clones 2546F10, 1001J9, and 843G6. These gaps correspond to estimated sizes of 23 kb, 10 kb and 7.5 kb respectively.

2.3.3 Identification of genes and gene signatures.

2.3.3.1 In silico analysis. The genes contained within the contig were identified by BLAST analysis of the NCBI non-redundant, Unigene, and EST databases using the genomic sequence of the region as the query (3). The genomic sequence was masked for repetitive elements and broken into 100 kb segments for detailed inspection. These analyses led to the identification of 43 genes, 141 Unigene EST clusters and 382 singleton ESTs. Many Unigene EST clusters and singleton ESTs failed to exhibit clear intron and exon boundaries with corresponding open reading frames. These ESTs may define genes that can only be partially constructed from the information available in public databases or they may represent genomic contamination clones that are known to exist in the public EST databases (Gonzalez and Sylvester 1997). The term "gene signatures" was used to describe these ESTs and detailed *in silico* analysis was used to predict their status as either possible genes or database contaminants. All gene signatures corresponding to Unigene EST clusters were considered as possible genes, except those containing 3' genomic poly(A) sequence stretches. These latter ESTs are a result of oligo-dT primer binding to contaminating DNA in reverse transcription experiments. Detailed inspection and assembly of physically adjacent Unigene clusters enabled the grouping of several gene signatures into single genes. Genes signatures represented by singleton ESTs were considered to be possible genes if they were found to fulfill any one of the following criteria indicative of real genes: exhibit intron/exon structure, possess mouse or human protein homologies or GENSCAN predicted exons. This *in silico* selection identified a total of 116 genes (Table 1). 82 Unigene EST clusters (Fig 7: rejected gene signatures, shaded) and the majority of singleton ESTs, were rejected as database contaminants or redundancies.

2.3.3.2 RT-PCR analysis. The *in silico* predicted genes were tested for their transcriptional status by RT-PCR analysis. Primer pairs were designed to each of these 116 genes (Table 1) and transcripts with null or poor amplification were studied with at least two different primer sets. The primer pairs used for amplification are presented in Table 1. RNA from 18 different adult and three different fetal tissues was pooled as template for RT-PCR amplification of each predicted gene. 104 of the 116 predicted

genes amplified from the pooled RNA and therefore were considered to define actual genes (Table 1 & Fig 7). 69 of the 116 identified genes had been previously positioned on the *Eag I* restriction map by database screening and Southern blot analysis. The remaining 47 genes were recovered from the sequencing effort. Twelve predicted genes gave non-reproducible RT-PCR products or failed to amplify. These predicted genes are either not genuine, or represent genes that are expressed in a tightly regulated manner. For example, expression may be restricted to certain stages of development or to tissues not represented the panel used in this study. These *in silico* predicted genes were not studied any further. The mRNA expression profiles of the genes giving clear RT-PCR evidence of transcription were analyzed in normal breast and a panel of breast cancer cell lines.

2.3.3.3 Quantitative mRNA expression profiling of the 16q24.3 genes in breast cancer cell lines. Mutation analyses was previously performed on seven genes mapping to this interval; *SPG7* (Settastian *et al*, 1999), *RPL13* (*BBC1*) (Moreland *et al*, 1997), *CDK10* (Crawford *et al*, 1999), *CPNE7* (Savino *et al*, 1999), *FANCA* (Cleton-Jansen *et al*, 1999), *GAS8* (*GAS11*) and *C16ORF3* (Whitmore *et al*, 1998b). None of these genes were found to harbor any mutations limited to sporadic breast cancer DNA samples. It was impractical to use this labor-intensive procedure to screen all the remaining candidates and furthermore, tumour DNA stocks were limited.

As an independent selection criterion for the identification of tumour suppressors, we analysed the expression profile of each transcript in a panel of breast cancer cell lines (Fig 8). The panel included eleven cell lines that exhibited 16q LOH and MCF12A, the latter derived from normal breast epithelium and included as a normal control. Using the expression value of each gene in normal breast tissue as the baseline, we calculated the relative fold-difference between the cell line exhibiting the highest expression and the cell line exhibiting the lowest expression. This value was termed the "relative fold variability index" (RFVI). Initially, we examined four housekeeping genes *PPIA* (peptidylprolyl isomerase, cyclophilin A), *GAPD* (*GAPDH*, glyceraldehyde-3-phosphate dehydrogenase), *RPB70* (RNA polymerase subunit II), and *ATP5A* (ATPase subunit II) to establish a RFVI baseline range of 1 to 42 (RFVI $_{PPIA}$ =1 and RFVI $_{ATP5A}$ =42). This

baseline range was considered to reflect mRNA expression differences that are due to variation in different genetic backgrounds, or experimental variability. Subsequently, we determined the RFVI for *SYK* and *CDKN2A*, two known tumour suppressor genes that have previously shown aberrant expression in breast cancer cells (Coopman *et al*, 2000 and Bisogna *et al*, 2001). The RFVI values for *CDKN2A* and *SYK* were 10 times greater than the upper limit of the baseline range (Fig 8). Of the 104 transcripts analysed, 66 displayed RFVI values within the baseline range and 9 exhibited higher values. The gene signatures corresponding to *DPEP1, CDH15*, Hs.17074, Hs.189419, *SLC7A5*, and AA994450 exhibited RFVI values that were only marginally higher than the baseline (Fig 8).

CBFA2T3 , *CYBA* and Hs.7970 displayed RFVI values considerably higher than the baseline. Figure 9 shows the individual expression profiles of these genes and *SYK* across the cell line panel. *CBFA2T3* and *CYBA* display variable expression, with some cell lines exhibiting marginal up-regulation and some exhibiting marked down-regulation. The Hs.7970 gene was downregulated in all cell lines, with the down regulation being more pronounced in MDA-MB-134 and SKBR3. 29 transcripts displayed complex melt curves and/or very late C_T values. Late C_T values represent no detectable amplification after 20 amplification cycles. The expression of these genes was considered to be too low to allow meaningful interpretation of real-time PCR results.

2.4 Discussion

2.4.1) Physical mapping and sequencing. Draft versions of the human genome have recently become available in the form of electronically assembled, annotated scaffolds (International Human Genome Sequencing Consortium, 2001 (7); Venter *et al*, 2001 (8)). Naturally occurring repeats, electronic chimeras and duplications have resulted in the presence of a substantial number of mapping discrepancies and misalignments in these sequence drafts (Katsanis *et al*, 2001). In contrast, the 16q24.3 gene map presented in this report is based on sequence data derived from an experimentally determined physical map. Southern hybridization, PCR based probe mapping and *in situ* hybridisation were used for the construction of this physical map, making it a more reliable source of gene order in this region of the genome. The high-throughput sequencing of this region has

yielded an ordered sequenced "backbone" that allowed the identification and assembly of additional partially or randomly sequenced clones, deposited at GeneBank by other sequencing centers. This ordered sequence provides a valuable resource for future gap closure and completion of the sequence in this region. In addition, it serves as a resource for gene discovery efforts until the sequence of this region is completed.

2.4.2) In silico and in vitro gene identification. The draft sequence of 16q24.3 was used to identify genes mapping in this region. The strategy for gene identification in genomic sequence data was integrated, combining computational predictions, human curation and experimental validation. Genes were identified by *in silico* means involving comparisons to expressed sequence databases, cross species sequence conservation, algorithm-based gene prediction and manual inspection. This analysis allowed the identification of 43 actual genes that have been previously published or can be *in silico* constructed based on information available at the public databases. In addition, 73 possible gene loci were identified which cannot be constructed from the information available in the public databases and are represented by a very small number of ESTs and, in some cases, even single ESTs. Due to their limited sequence representation it cannot be determined *in silico* whether these ESTs represent genes or are the product of genomic DNA contamination in cDNA libraries. Consequently, experimental analysis in the form of RT-PCR was used to provide evidence for the representation of these genes in the human transcriptome. This analysis indicated that 61 of these 73 possible genes are transcribed. In total, the detailed *in silico* and RT-PCR analyses allowed the identification of 104 genes in the region.

The gene density in this region is not uniform. There are three gene-dense regions: one between *D16S3121* and the telomere, and one around each of the genes *MAP1A* and *MVD*. The gene density in this region is one gene per 15 kb. These three regions are separated by areas with densities of one gene per 35 kb. On average, there is approximately one gene every 20kb. This is consistent with previous observations that telomeric, Giemsa light bands are gene rich (Saccone *et al*, 1992).

2.4.3) Candidate gene analyses. Several of the genes identified present good tumour suppressor candidates based on their function or based on homologies to other genes or proteins of known function. These genes can be selected and prioritized for mutation screening in breast tumour DNA samples. However, the majority of genes identified are of no known or predicted function. Consequently, the mRNA expression profiles of all these genes were studied as an independent selection criterion for the identification of potential tumour suppressor(s). The working hypothesis was that tumour suppressor gene expression would exhibit marked variability across a panel of different breast cancer cell lines. This was found to be true for *SYK* and *CDKN2A*, two genes previously shown to act as tumour suppressors. Based on comparisons with the expression profiles of housekeeping genes, six genes were identified exhibiting moderately elevated variability, and three genes exhibiting significant variability values. *DPEP1, CDH15*, Hs.17074, Hs.189419, *SLC7A5* and AA994450 were found to be marginally more variable than the baseline defined by housekeeping genes. The possible involvement in tumour suppressor events of these marginally variable genes should be viewed with caution and will require confirmation by independent methodology. However, the genes *CBFA2T3, CYBA* and Hs.7970 exhibited significant expression variability.

2.4.3.1) CBFA2T3. *CBFA2T3* displayed expression that was five times more variable than the baseline. Expression of this gene was reduced in the cell lines BT-549, MDA-MB-157 and MDA-MB-231. These cell lines are also downregulated for *SYK* and have been previously shown to display typical features of malignant phenotype including increased motility and invasion (Coopman *et al*, 2000). *CBFA2T3* expression was also reduced in SKBr3 and MDA-MB-468. These two cell lines are not downregulated for *SYK* and display pre-malignant phenotypes with very low motility and predominantly non-invasive cluster formation (Coopman *et al*, 2000). Furthermore, *CBFA2T3* is also downregulated in MCF12A, which is non-tumorigenic. These findings suggest that *CBFA2T3* down-regulation is an early event in breast cancer. This is consistent with the notion that chromosome 16 LOH occurs early in carcinogenesis (Gong *et al*, 2001). *CBFA2T3* is a member of the ETO family that also includes *CBFA2T1*, *CBFA2T2* and *Drosophila melanogaster nervy*. *CBFA2T1* and *CBFA2T3* were identified through their

involvement in chromosomal translocations observed in acute myeloid leukemia (AML) (Kozu *et al*, 1993; Gamou *et al*, 1998). Such translocation events result in the generation of AML1/CBFA2T1 or AML1/CBFA2T3 fusion proteins that repress genes normally activated by AML1. The CBFA2T1protein has been shown to interact with corepressor complexes involving N-CoR, mSin3A and histone deacetylases (Hildebrand *et al*, 2001). Based on the high sequence conservation between *CBFA2T1* and *CBFA2T3* it was reasonable to hypothesise that CBFA2T3 also exhibits transcription repressor function.

Subsequent investigations, lead by Dr Marina Kochetkova, established *CBFA2T3* as a putative breast tumour suppressive gene (Kochetkova *et al*, 2002). This reported showed that the expression of the *CBFA2T3* gene was significantly reduced in primary breast tumours when compared with normal breast tissue. Over-expression of CBFA2T3 in breast tumour cell lines with low or undetectable levels of endogenous *CBFA2T3* mRNA resulted in a significant inhibition of colony growth on plastic and in soft agar. Furthermore, *CBFA2T3* was shown to function as a transcriptional repressor when tethered to the GAL4 DNA-binding domain in gene reporter assays, and therefore has the potential to be a transcriptional repressor in normal breast epithelial cells. Taken together, *CBFA2T3* exhibits genetic and functional properties that are consistent with those of a tumour suppressor gene and suggest a potential role for this gene in breast cancer tumourigenesis.

2.4.3.2) *CYBA.* *CYBA* was found to be three times more variable than the housekeeping gene baseline. Expression of this gene was reduced in the cell lines MDA-MB-134, SKBR3 and T47D. The role of CYBA as an NADPH oxidase in the microbicidal function of phagocytes has been studied extensively. Mutations in the CYBA gene are causative of chronic granulomatous disease characterized by recurrent bacterial and fungal infections (Rae *et al*, 2000). More recently, CYBA has been implicated in oxidases involved in epithelial and muscle cell gene regulation and function with implications in atherosclerosis (Sorescu *et al*, 2001). CYBA associates with CYBB to form cytochrome b-558, which is the membrane component of NADPH oxidase and functions as the final electron transporter in the oxidation of NADPH, resulting in the generation of Reactive Oxygen Species (ROS) such as O_2^- and H_2O_2. Several studies have

shown the involvement of reactive oxygen species in carcinogenesis and tumour progression (Gupta *et al*, 2001; Brown and Bicknell, 2001). ROS concentration levels are critical in the regulation of a number of genes involved in diverse pathways delineating transcription, proliferation and apoptosis (Burdon *et al*, 1996; Arnold *et al*, 2001; Jacobson, 1996). ROS levels are under a tight regulatory control involving the interplay of NADPH oxidases and antioxidant ROS scavengers (Griendling and Ushio-Fukai, 2000). Any disruption to these control mechanisms is likely to result in aberrant cell behaviour as that seen in cancer.

2.4.3.3) Hs.7970. In addition to the structural evidence provided by LOH, functional evidence supporting the presence of a tumour suppressor gene at 16q24.3 has come from micro-cell mediated transfer experiments. The introduction of a 16q24.3 YAC clone into breast cancer cell lines induced cellular senescence (Reddy *et al,* 2000). The gene represented by the unigene cluster Hs.7970 maps in the region included in this YAC clone (Chapter 4 describes the mapping and characterisation of the genes located in this YAC clone). Hs.7970 exhibited expression variability four times higher than the baseline. However, this finding should be interpreted with caution since this gene exhibited low expression in all the cell lines studied. Consequently, the exponential nature of PCR amplification combined with the small number of target molecules may result in the exaggeration of trivial variations. The Hs.7970 gene encodes a protein that appears to contain an F-box domain. F-box containing proteins function as ubiquitin protein ligases and the ubiqitin-proteasome pathway is often the target of cancer-related deregulation (Spataro *et al*, 1998).

The highly variable expression patterns observed for the *CBFA2T3*, *CYBA* and Hs.7970 genes suggests a possible involvement in breast cancer and warrants further experimentation to elucidate the exact role these genes may play in breast cancer processes. The notion of these genes being breast cancer tumour suppressors is compatible with what is known about their respective functions.

Figures

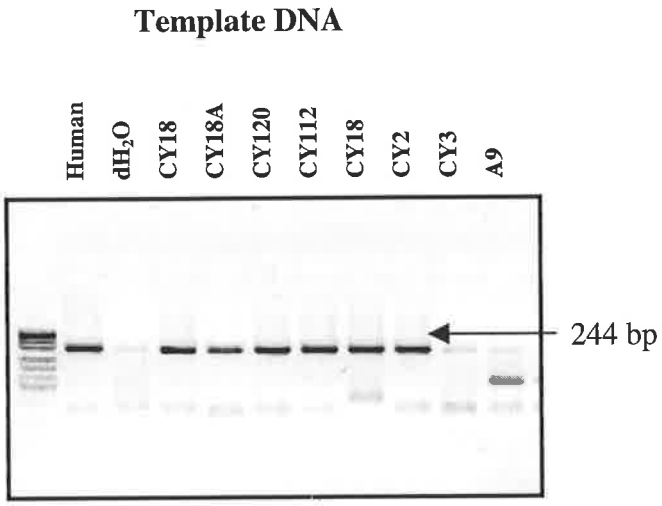


Figure 1: PCR based mapping of the *CYBA* gene. Primer pairs (5' GCGAGCGGCATCTACCTACTGG & 5' GGTTGCTGGGCGGCTGCTT GATGG) were used to PCR amplify *CYBA* from various templates including somatic cell hybrid DNA (**CY**), human genomic DNA and **A9** mouse cell line DNA. *CYBA* maps to the interval demarked by the somatic cell hybrid breakpoints CY2/CY3 and CY18(A).

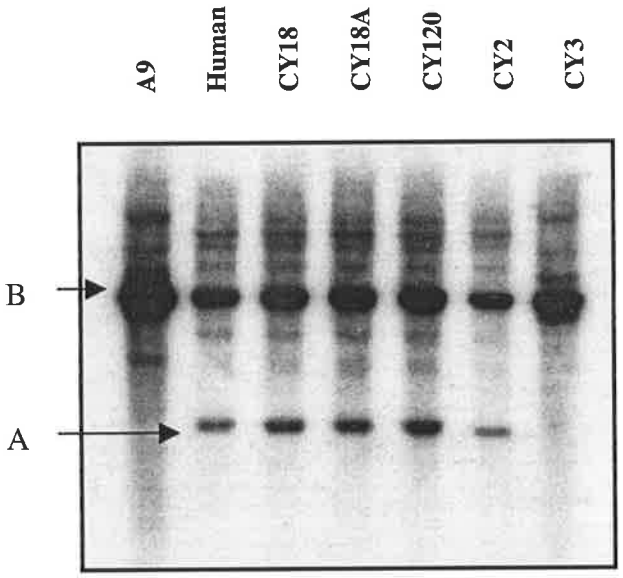


Figure 2: Southern blot analysis of *APRT*. Somatic cell hybrid DNA (5 μg) and total human genomic DNA (10 μg) was digested with Pst1, resolved on 0.8% agarose and blotted in 10X SSC to nylon membrane. The 316 bp *APRT* cDNA probe hybridized to a *Pst I* restriction fragment present in all lanes except CY3, and the mouse background control (A9) (A). This hybrid profile was indicative of all the genes shown to map in this candidate interval. A non-specific band was also detected in both mouse and human DNA (B)

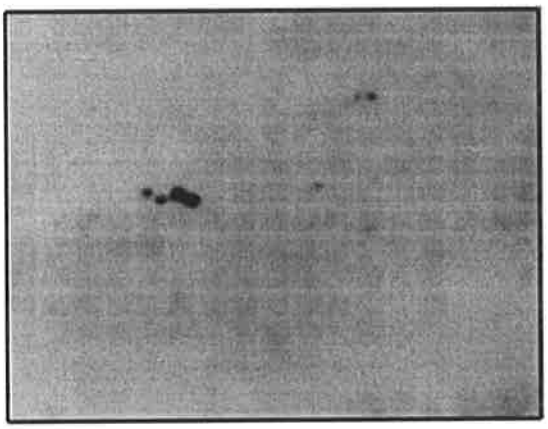


Figure 3: A Representative autoradiograph from a BAC genomic library screening (RPCI-11: segment 3). A 227 bp *GALNS* cDNA probe was hybridized to 6 filters within the library, and the autoradiograph shown represents 1/12 of a single filter. Each filter contains approximately 18,432 clones, arranged into subsets of 16, each represented twice within the subset to avoid false positive signals. Three positive clones were detected.

Figure 4: Pulse field gel electrophoresis of Eag1 restricted BAC and PAC clones. 150 ng of restricted DNA from different BAC and PAC clones was loaded in each lane (x axis). Clear resolution between 2 to 48 kb was achieved (y axis), enabling restriction map assembly.

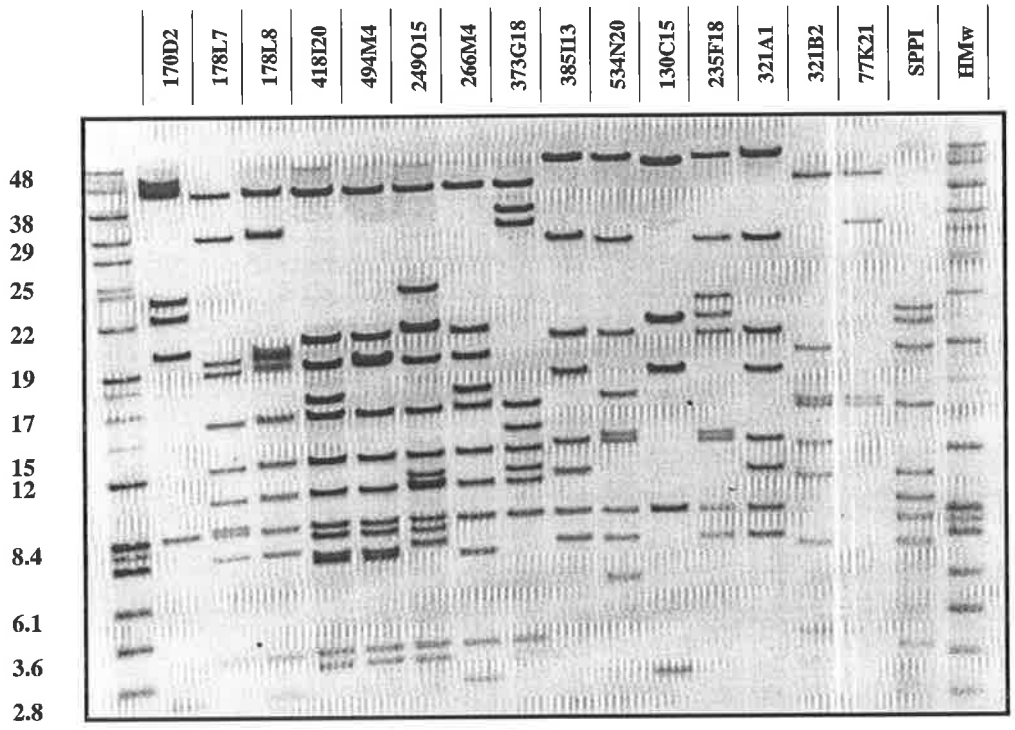
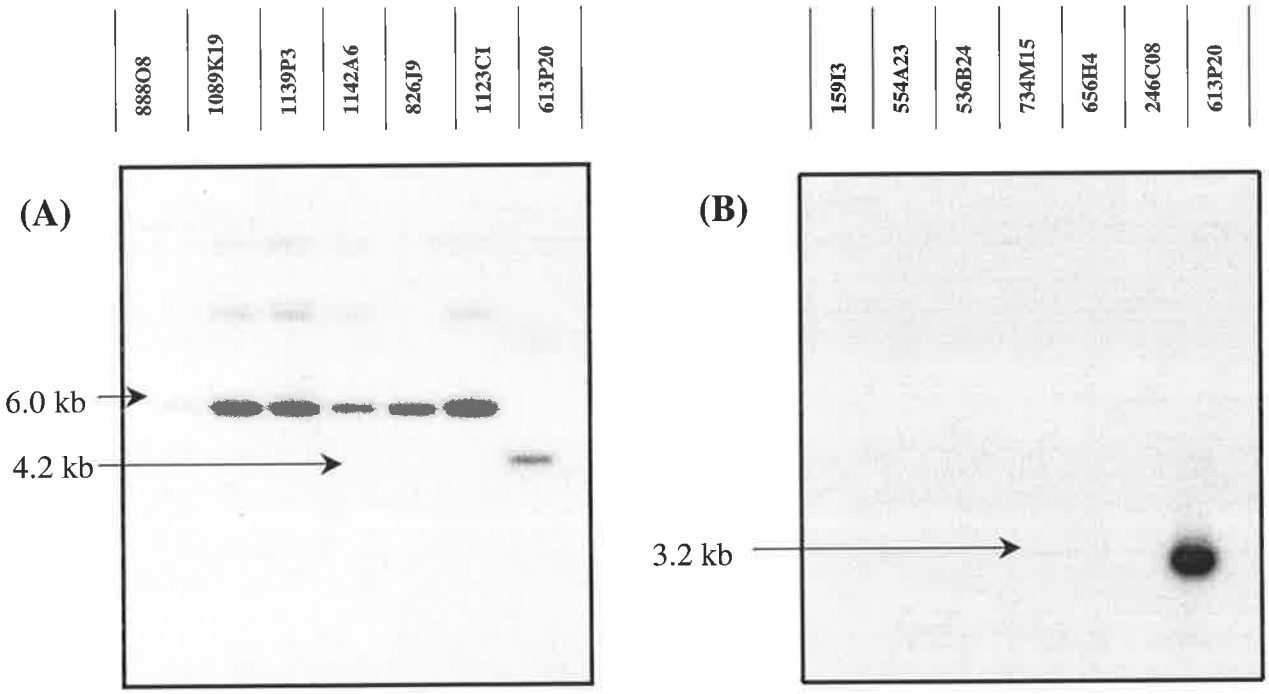


Figure 5: Mapping of BAC 613P20 and identification of contig ends. Probes generated to the SP6 (A) and T7 (B) ends of 613P20 were hybridized to pulse field blots containing the clone itself, plus several overlapping clones. The 613P20/SP6 end hybridized to a common 6.0 kb Eag1 restriction fragment present in many overlapping clones, in addition to the unique 4.2 kb SP6 end fragment of 613P20. The 613P20/T7 end hybridized to a unique 3.2 kb fragment and thus represents the extreme end of the contig, and was subsequently used in library re-screening.



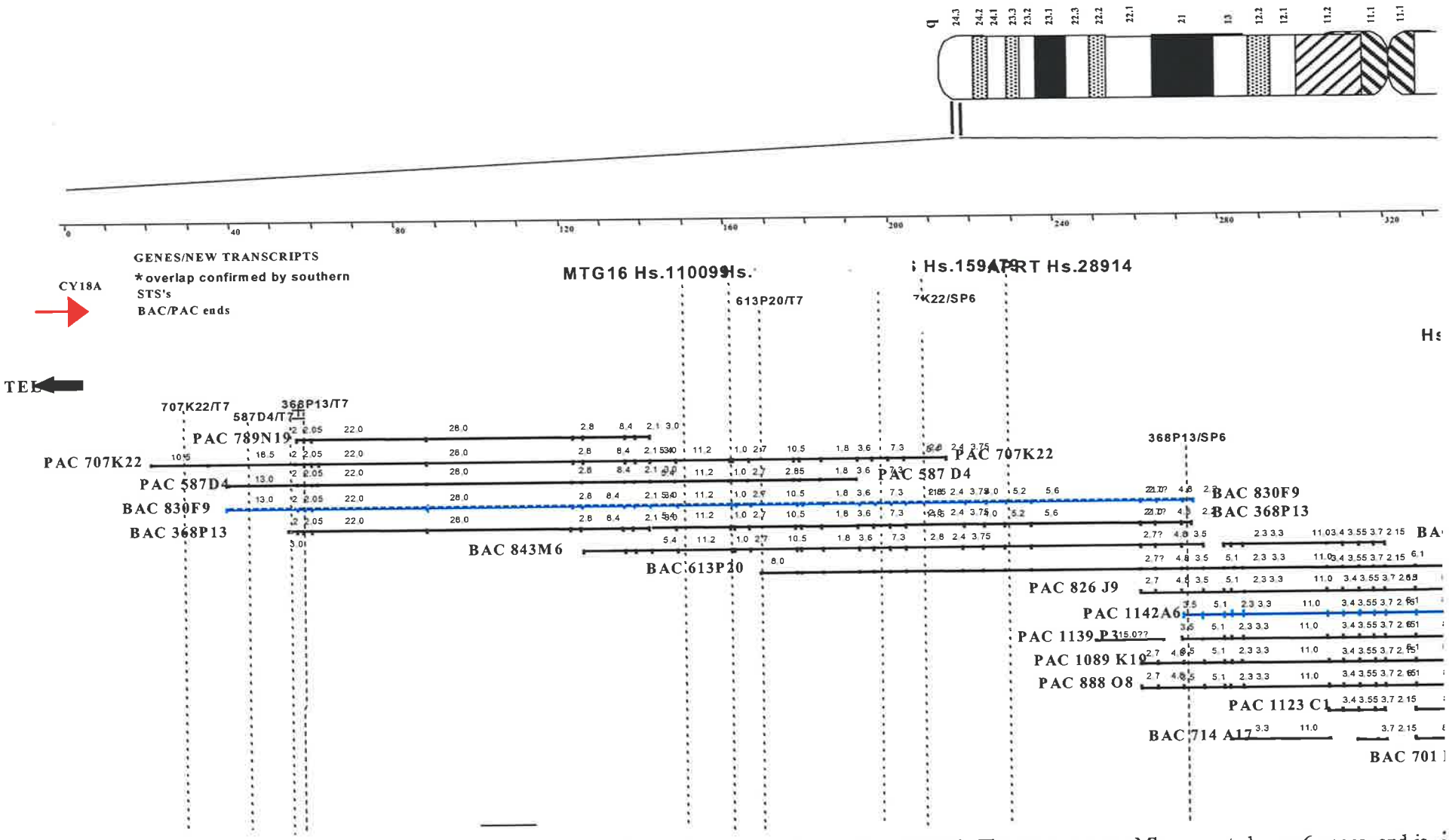
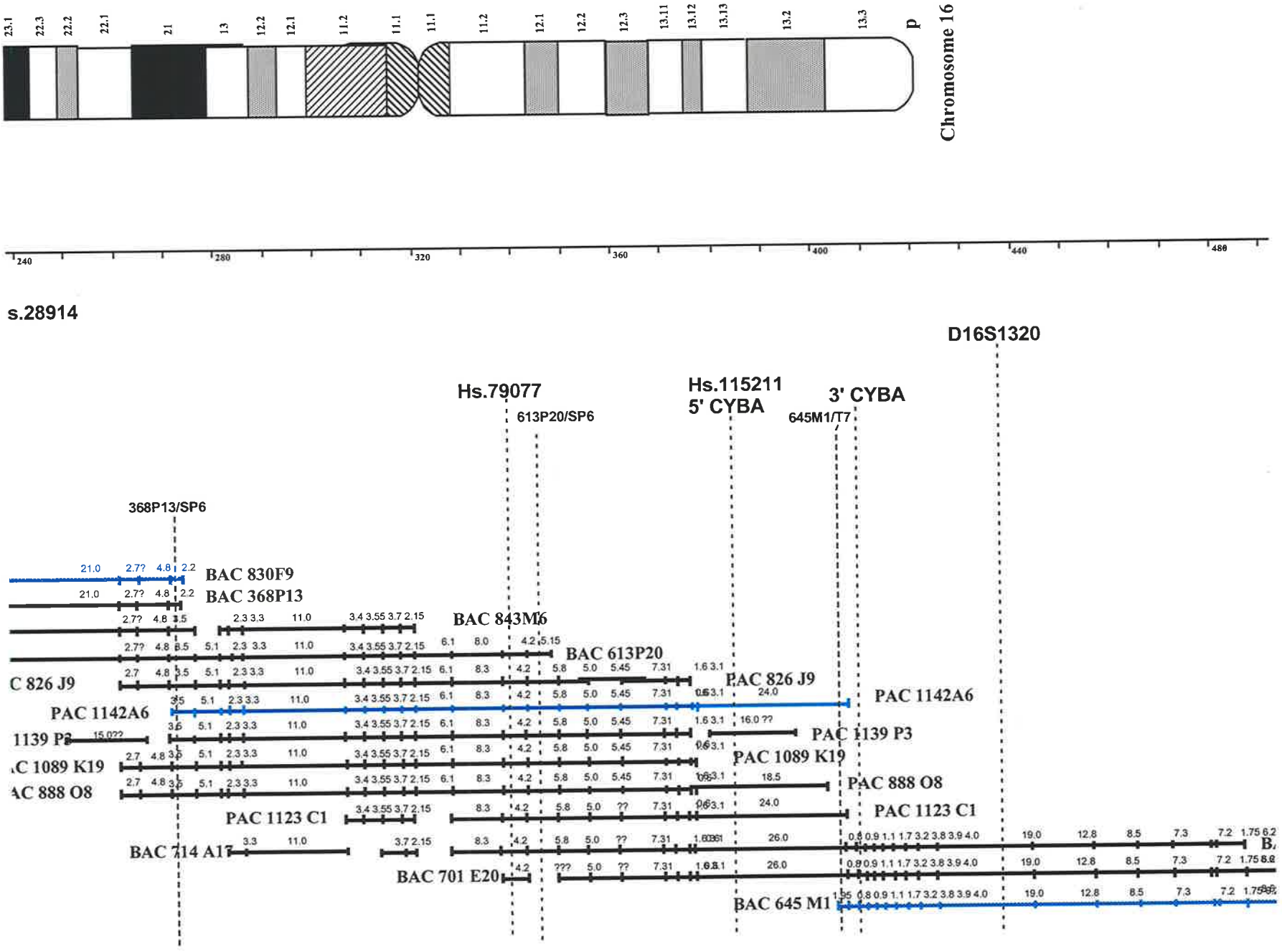
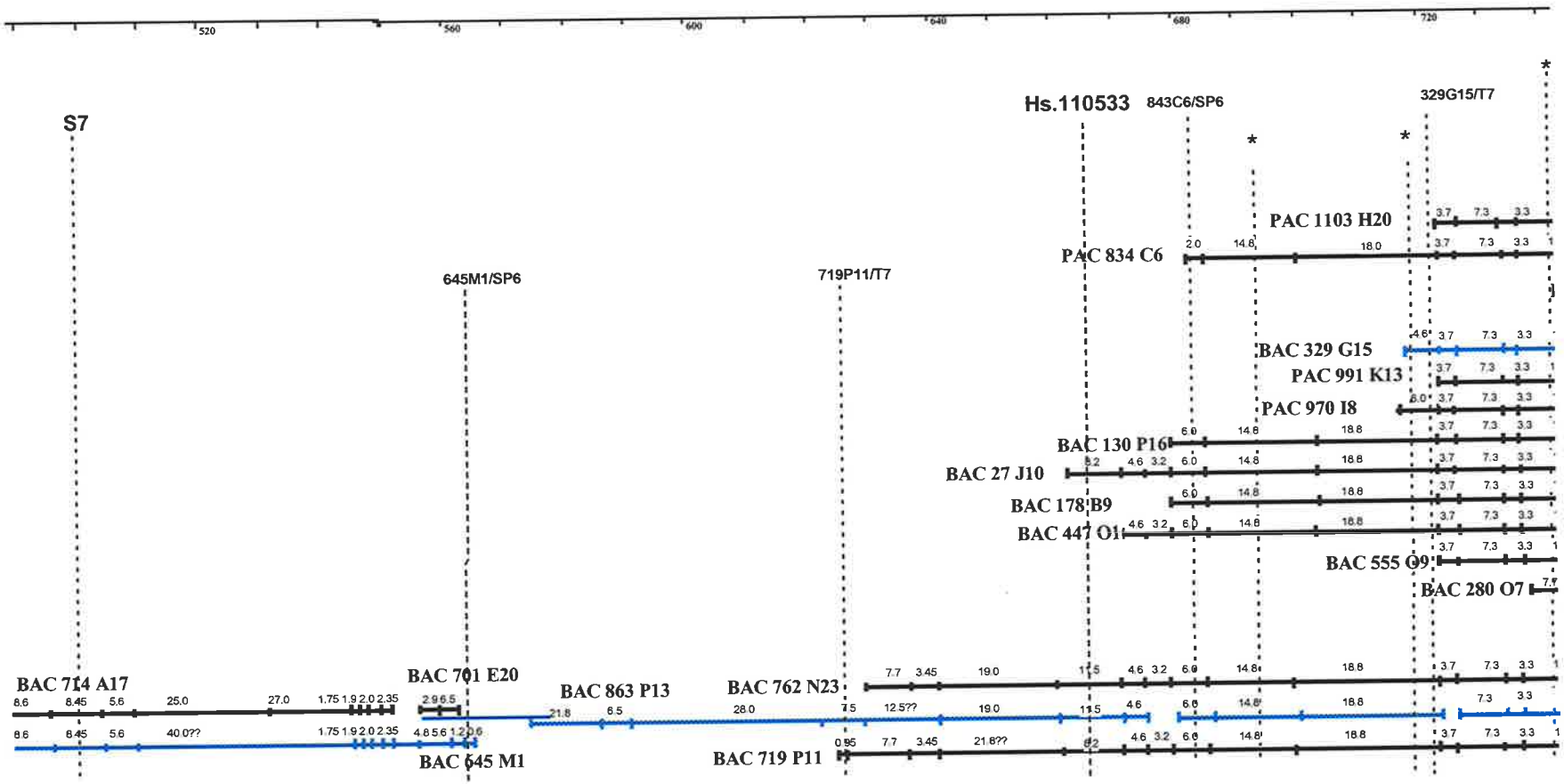
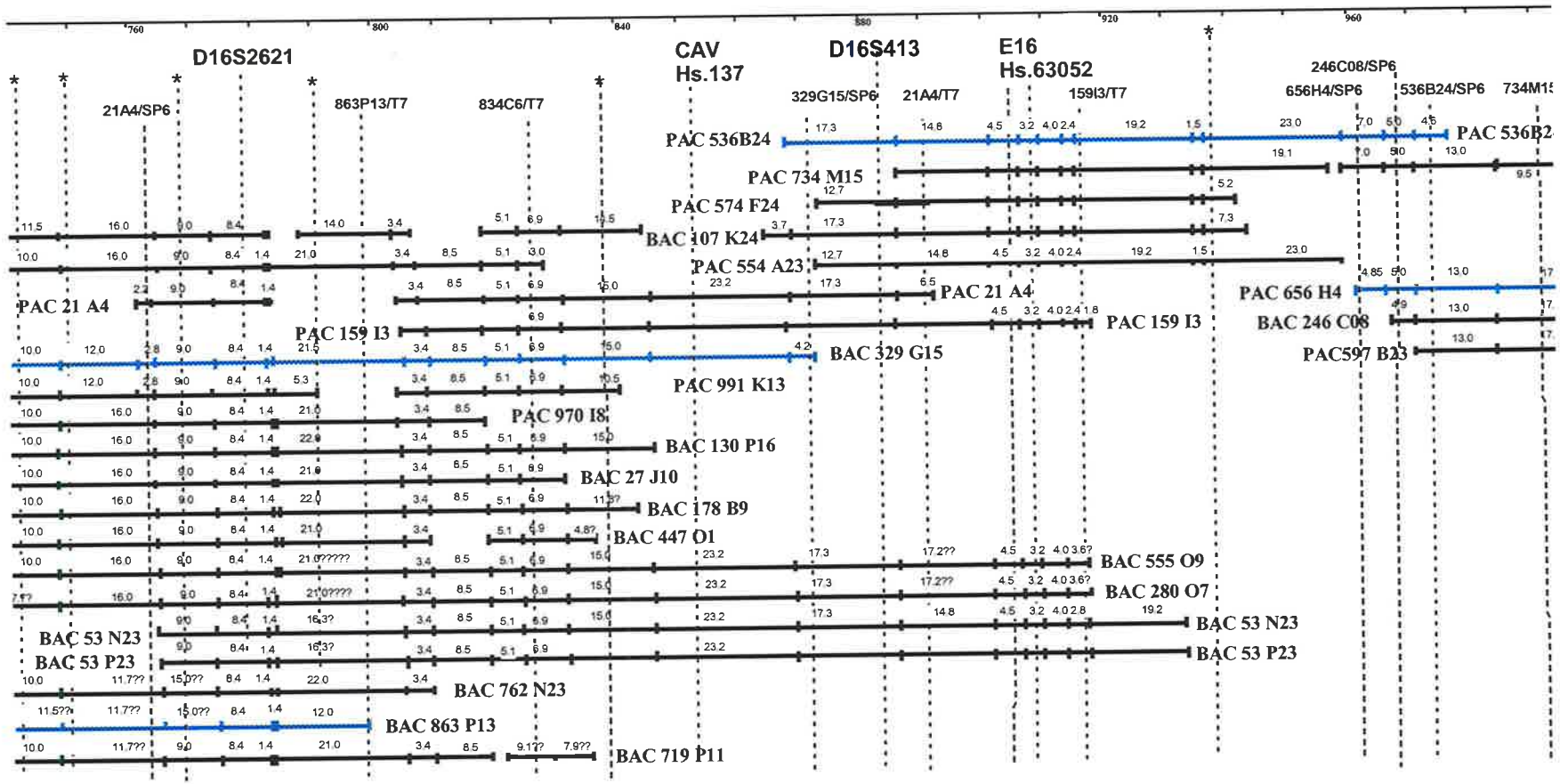
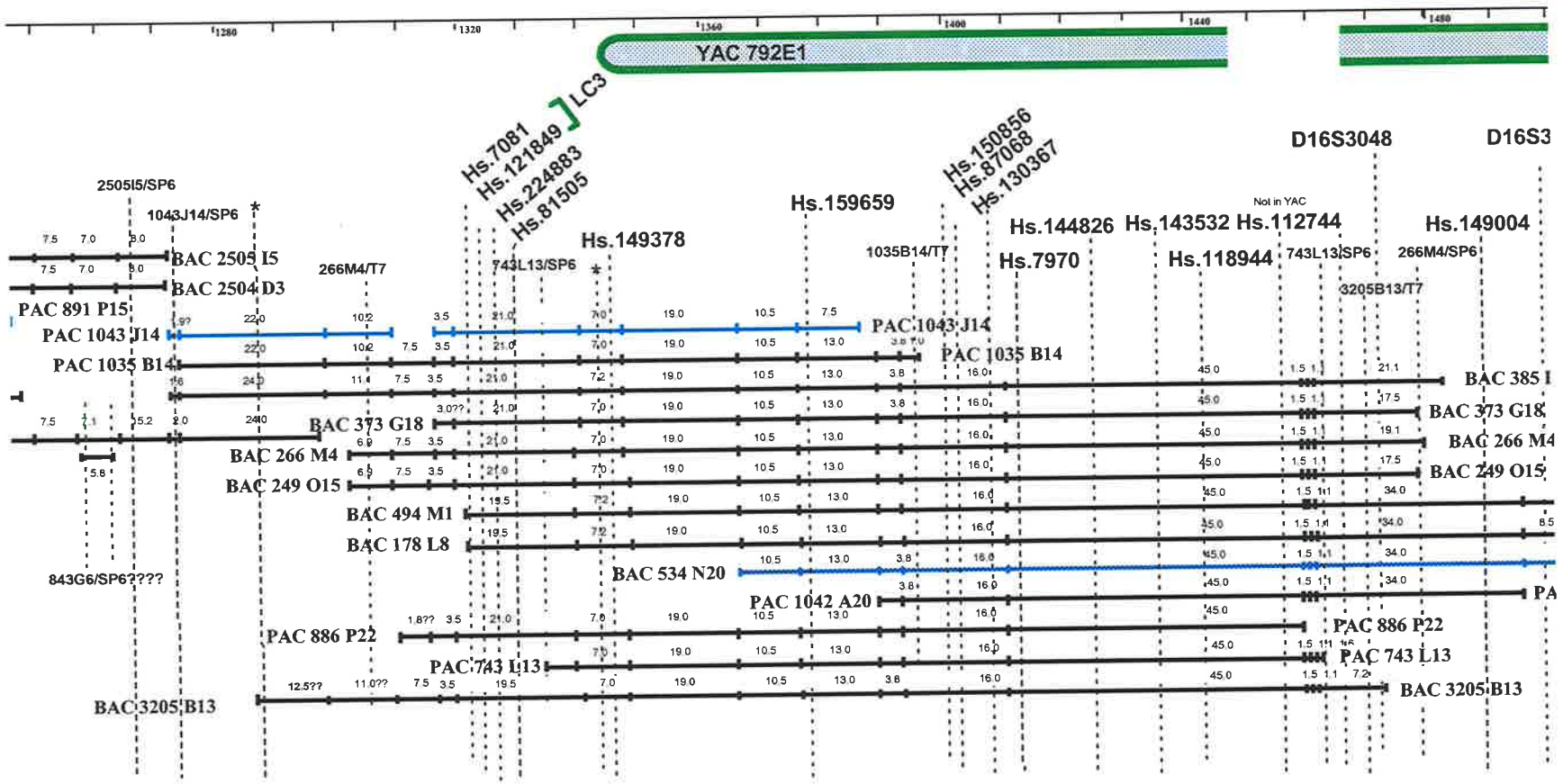


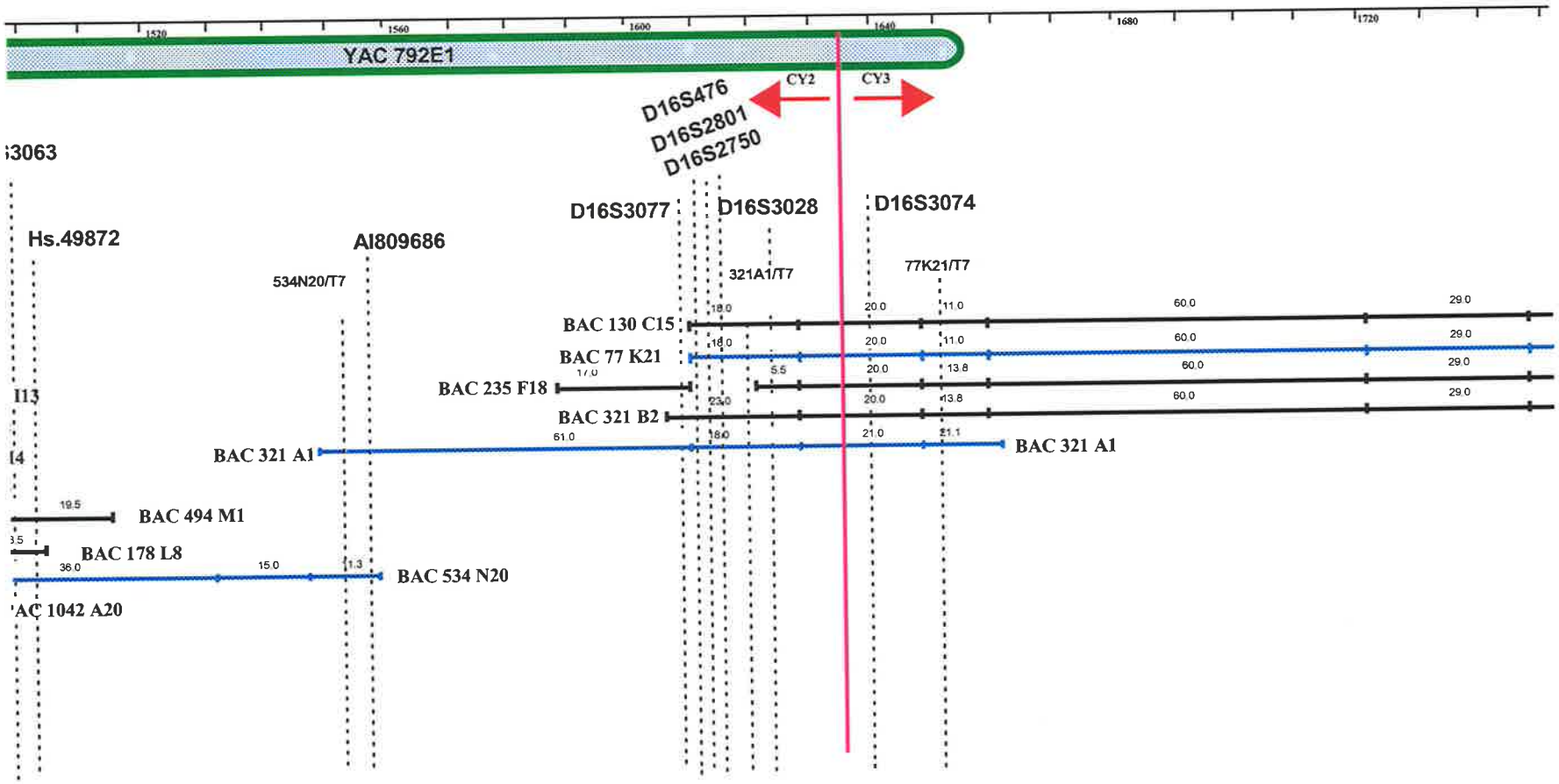
Figure 6: A detailed physical map of the restricted region of loss of heterozygosity (LOH) at 16q24.3. The maps spans 1.8Mb, presented over 6 pages, and is bounded by PAC 707K22/T7 (centromeric end) and the microsatellite *D16S3028* (telomeric end). At the top of each page is a scale bar indicating distance in kilobases. The horizontal black bars represent individual PAC and BAC clones with their name indicated. Each clone consists of ordered Eag1 restriction fragments, with their corresponding sizes indicated. The vertical dashed lines represent clone overlaps that have been confirmed by Southern hybridization. Probes used for Southern analyses included ESTs, STSs, clone end fragments and isolated restriction fragments, which are presented above their corresponding Eag1 restriction fragment. Clones representing the minimal tiling path are presented in blue.


23.1
22.3
22.2
22.1
21
13
12.2
12.1
11.2
11.1
11.1
11.2
12.1
12.2
12.3
13.11
13.12
13.13
13.2
13.3
p
Chromosome 16
240
280
320
360
400
440
480
s.28914
D16S1320
Hs.79077
613P20/SP6
Hs.115211
5' CYBA
3' CYBA
645M1/T7
368P13/SP6
BAC 830F9
BAC 368P13
BAC 843M6
BAC 613P20
C 826 J9
PAC 826 J9
PAC 1142A6
PAC 1142A6
1139 P3
PAC 1139 P3
C 1089 K19
PAC 1089 K19
AC 888 O8
PAC 888 O8
PAC 1123 C1
PAC 1123 C1
BAC 714 A17
BAC 701 E20
BAC 645 M1


520
560
600
640
680
720
S7
645M1/SP6
719P11/T7
Hs.110533
843C6/SP6
329G15/T7
PAC 1103 H20
PAC 834 C6
BAC 329 G15
PAC 991 K13
PAC 970 I8
BAC 130 P16
BAC 27 J10
BAC 178 B9
BAC 447 O1
BAC 555 O9
BAC 280 O7
BAC 714 A17
BAC 701 E20
BAC 863 P13
BAC 762 N23
BAC 645 M1
BAC 719 P11


D16S2621
CAV
Hs.137
D16S413
E16
Hs.63052
21A4/SP6
863P13/T7
834C6/T7
329G15/SP6
21A4/T7
159I3/T7
656H4/SP6
246C08/SP6
536B24/SP6
734M15
PAC 536B24
PAC 734 M15
PAC 574 F24
BAC 107 K24
PAC 554 A23
PAC 21 A4
PAC 159 I3
BAC 329 G15
PAC 991 K13
PAC 970 I8
BAC 130 P16
BAC 27 J10
BAC 178 B9
BAC 447 O1
BAC 555 O9
BAC 280 O7
BAC 53 N23
BAC 53 P23
BAC 762 N23
BAC 863 P13
BAC 719 P11
PAC 656 H4
BAC 246 C08
PAC597 B23


1280
1320
1360
1400
1440
1480
YAC 792E1
LC3
Hs.7081
Hs.121849
Hs.224883
Hs.81505
Hs.149378
Hs.159659
Hs.150856
Hs.87068
Hs.130367
Hs.144826
Hs.7970
Hs.143532
Hs.118944
Hs.112744
Not in YAC
D16S3048
Hs.149004
D16S3
2505I5/SP6
1043J14/SP6
266M4/T7
743L13/SP6
1035B14/T7
743L13/SP6
266M4/SP6
3205B13/T7
BAC 2505 I5
BAC 2504 D3
PAC 891 P15
PAC 1043 J14
PAC 1035 B14
BAC 373 G18
BAC 266 M4
BAC 249 O15
BAC 494 M1
BAC 178 L8
BAC 534 N20
PAC 1042 A20
PAC 886 P22
PAC 743 L13
BAC 3205 B13
843G6/SP6????
PAC 1043 J14
PAC 1035 B14
BAC 385 I
BAC 373 G18
BAC 266 M4
BAC 249 O15
PAC 886 P22
PAC 743 L13
BAC 3205 B13
PA


1520
1560
1600
1640
1680
1720
YAC 792E1
CY2
CY3
D16S476
D16S2801
D16S2750
3063
D16S3077
D16S3028
D16S3074
Hs.49872
AI809686
534N20/T7
321A1/T7
77K21/T7
BAC 130 C15
BAC 77 K21
BAC 235 F18
BAC 321 B2
BAC 321 A1
BAC 321 A1
I13
I4
BAC 494 M1
BAC 178 L8
BAC 534 N20
AC 1042 A20

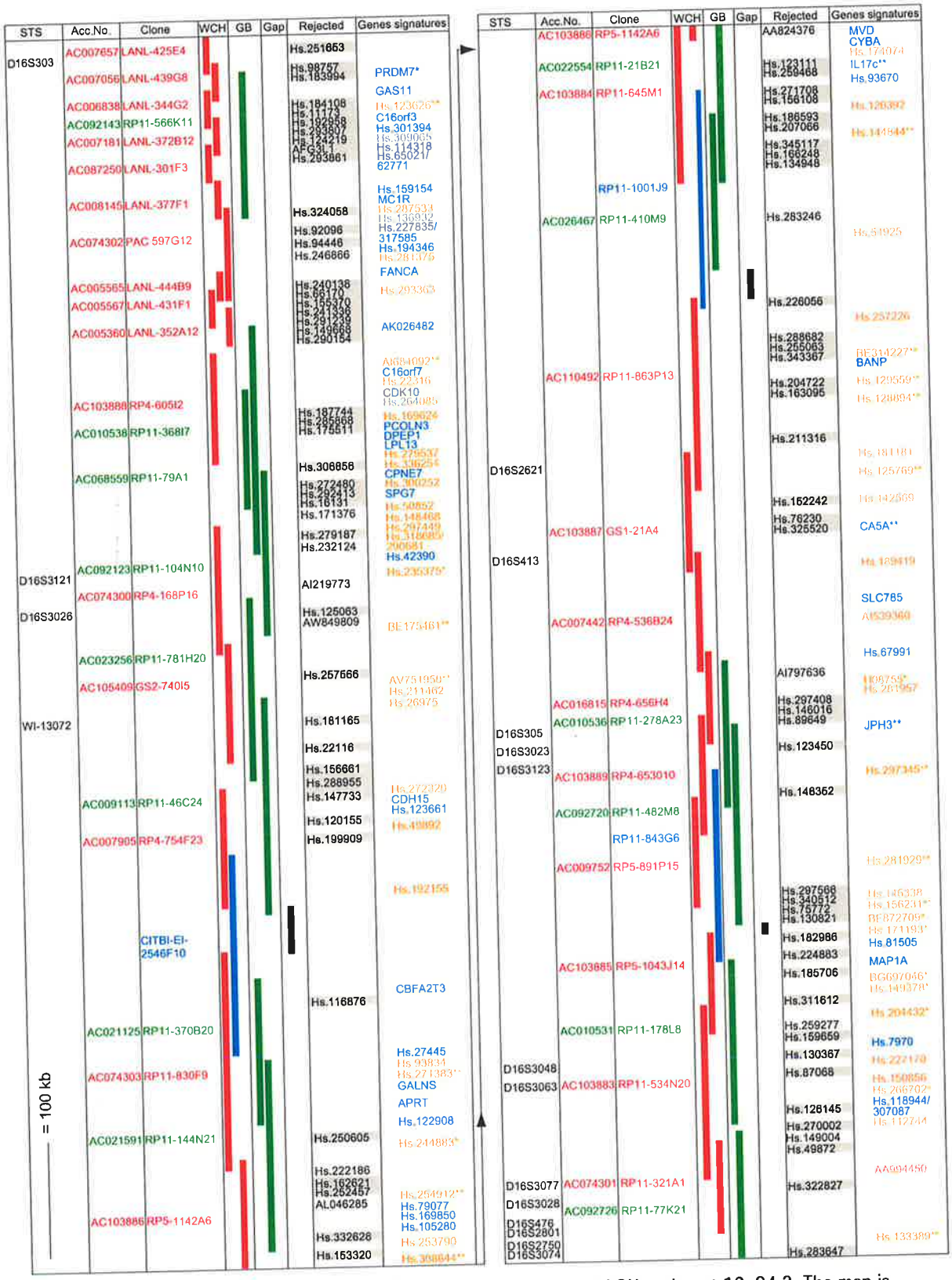


Figure 7: Integrated physical and transcript map of the sequenced LOH region at 16q24.3. The map is anchored at the telomere and extends just over 2.4 Mb centromeric, to the marker D16S3028. Clones sequenced by our group are shown in red (WCH); clones integrated from GenBank (GB) are shown in green. The physical map is contiguous and three sequencing gaps are indicated in black. BAC clones spanning sequence gaps are shown in (blue). Genes (blue), gene signatures (orange), and rejected gene signatures are ordered with their corresponding sequenced genomic clone adjacent. Gene signatures were

rejected on the basis of either in silico analysis (shaded) or RT-PCR (plain). Genes and gene signatures excluded from real-time analysis due to late C_T values (*) or unsatisfactory melt curves (**) are indicated.

Powell_Fig2

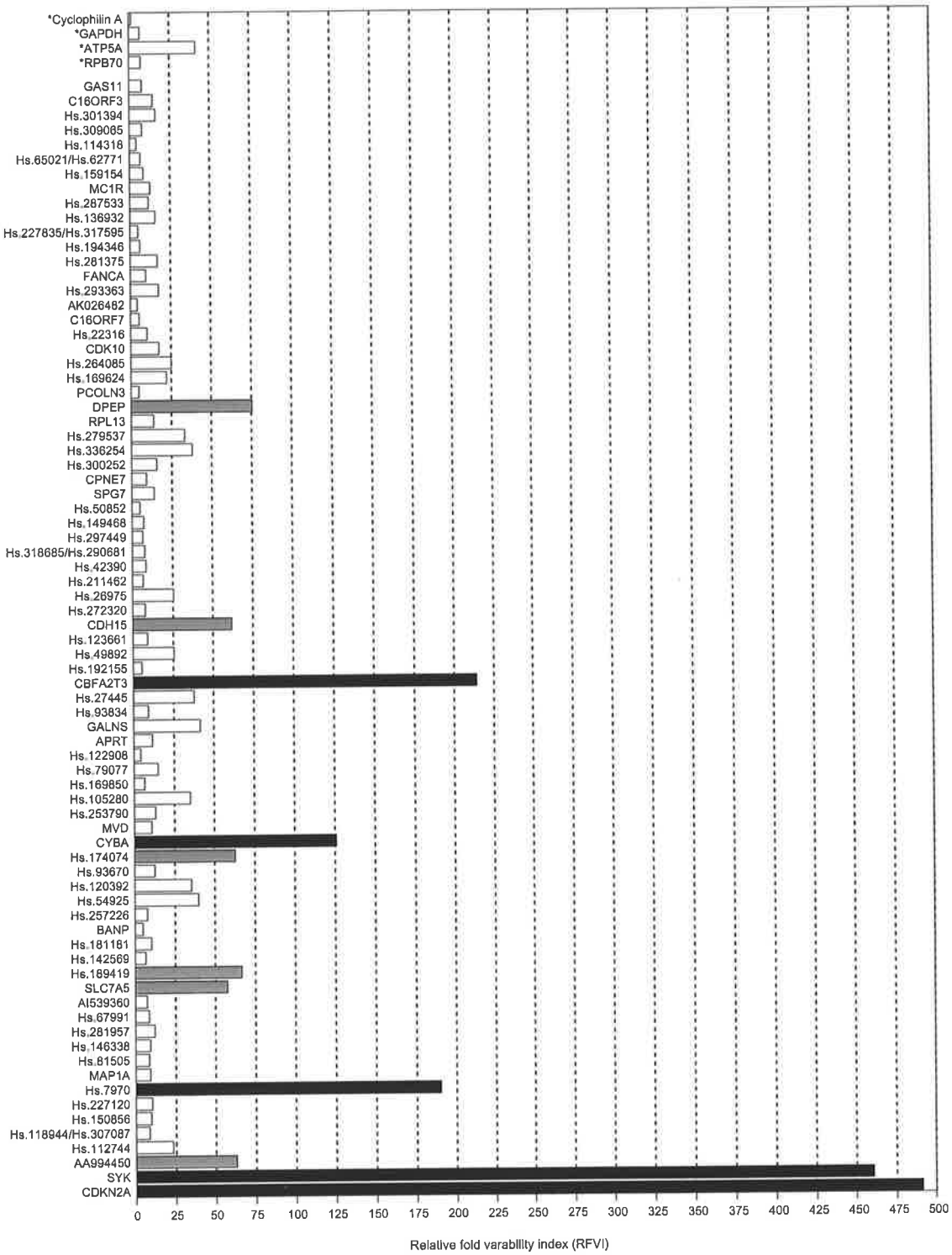


Figure 8: Gene expression variability in breast cancer cell lines exhibiting 16q LOH. The range of mRNA expression variability for each gene analyzed is expressed as RFVI. The RFVI values of four housekeeping genes (*) were used to establish a baseline range (RFVIbaseline =1-42). The tumour-suppressor genes SYK and CDKN2A display very high RFVI values. The RFVI values of 75 genes mapping at 16q24.3 are shown. Six genes display moderately elevated variability (gray bars) and three genes display significant elevated variability (black bars).

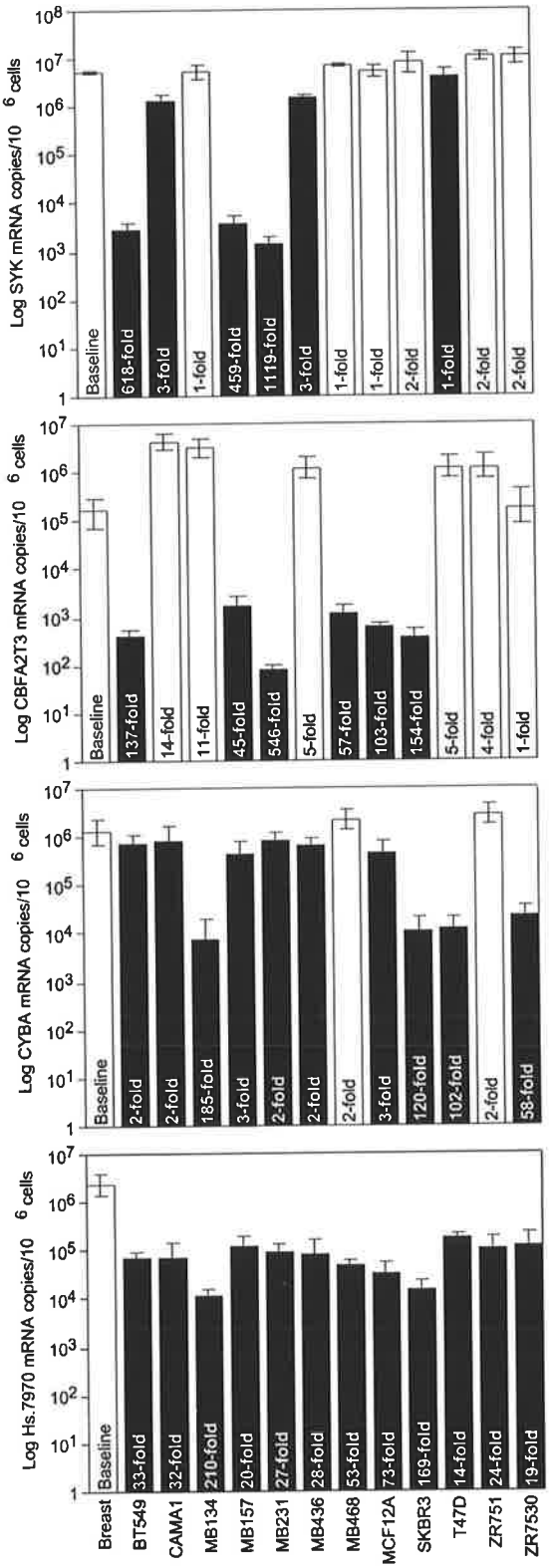


Figure 9: mRNA expression analysis of the tumour-suppressor gene SYK and the 16q24.3 transcripts CBFA2T3, CYBA and Hs.7970 by real-time RT-PCR. Data are expressed as log mRNA copy number/10^6 cells (± SD, n = 4). Fold changes shown were calculated relative to expression in the "baseline" normal breast tissue. Black and white bars represent fold up-regulation and down-regulation respective to the baseline.

Tables

Table 1: Summary of the *in silico* accepted genes and gene signatures. All predicted genes were RT-PCR amplified from pooled human tissue RNA. Genes and gene signatures were either accepted or rejected on the bases of late C_T values (very low expression) or unsatisfactory melt curves (multiple products) or no expression.

Genes & Unigene clusters	***Primer location***	***Size (bp)***	***Forward primer 5'- 3'***	***Reverse primer 5'- 3'***
Accepted:				
In vitro and in silico				
1. GAS11	I/E boundary	222	AGGATGCGGAATGATTTTGAGAGG	GGCCAGGTTGTTGAGGGTGATGTC
2. C16orf3	I/E boundary	223	ACCCCCATCCCTCCCCACCTG	TCCCTGTCCGCTTCCCATTCCTG
3. Hs.301394	I/E boundary	242	AGACCTGCTGGGGACTCG	CAGGGGACAGGGGACAATGG
4. Hs.309065	I/E boundary	259	ATGCAGCAGGATCACTACGAGAGC	TGGGCGACAGAGCGGAAGAC
5. Hs.114318	I/E boundary	168	ACCCTTGACTTGCATGAAACCTCT	GTTCACAAATAGCTACTCCTCTG
6. Hs.65021/Hs.62771	No gap	142	CTTTACTGTTTTCCCACCT	CCTGGTGAACTCGACATTG
7. Hs.159054	I/E boundary	233	GACGCCACGGCCGAGGAAGAG	GGGGGAGGACGAGGCCATAAATAC
8. MC1R	I/E boundary	206	GCATCGCCCGGCTCCACAAGA	GATGATGAGGGCGAGAAAGAGGTT
9. Hs.287533	No gap	141	CGCGCCCGGCCCTAAAATGTT	CCGGGTGGGAACGCTGTGTCTGT
10. Hs.136932	No gap	289	AACTTGGGGGTGGGGGTGTGG	TGACCTGCAGCTGAGCCTACTTCC
11. Hs.227835/Hs.317585	I/E boundary	291	TGCAGGCCACGTTCTTTCCCATCT	CCACTCCCCCTCCCCCTCAGC
12. Hs.194346	I/E boundary	202	ACCGCGCCAATCCAGAGAAGA	CTGCGTCGTGGCGTAAGTGTTG
13. Hs.281375	No gap	154	AAGGTCTTCCAGGCGGTAGG	TGGCTGTGTCCGGTTGTCGTGT
14. FANCA	I/E boundary	238	TGGCCGCACAGGAAATGAGGATA	AGACGAAGGCAGGCGGAGGAGGAT
15. Hs.293363	No gap	262	TCGAGGAGCCAGTGAAGG	TTCTTAACCATGCCGTATTGA
16. AK026482	I/E boundary	230	ACTGGGGGCTGGATGTGAT	GCGCCTGCTTGGTGTGG
17. C16orf7	I/E boundary	150	GCTGAAGACCGCGGTGGAGGAGAT	AGAGGCCACAGCGGGGAGAAAAAG
18. Hs.22316	I/E boundary	119	TGTCTCTGCGTACGTAGCCCATCT	AAGGAGTTCACCACCATCAAG
19. CDK10	I/E boundary	272	GCGCCTGCTGCACTTCCTGTTCAT	CCTGGCTCGCGGACCCTTCA
20. Hs.264085	I/E boundary	148	GGAGGCCACGGCGATTACCC	CAGTGAAGCCAGATTAGAGGACAT
21. Hs.169624	I/E boundary	164	CCTGCTGGACTTTTGCTTTAC	CTAAGGAGAAGTTGATGGAGAA
22. PCOLN3	I/E boundary	297	GCCTTCCCCGACCACACCCCAATG	AGCCGCCCACCACCTGCCTTTTCT

23. DPEP1	I/E boundary	172	GAGCCCCGGTCATCTTCAGC	TCGGCCACTTGGGACAGGTT
24. LPL13	I/E boundary	453	TGGCAGGAGGAAGTCACAGC	CGAGGAGGCGGAACAAGTC
25. Hs.279537	No gap	224	CCCGAAGCCCACAGTGAGGACAG	CACTTGGGGAATGGGACGCAGAAC
26. Hs.336254	No gap	136	GGCGCACCTGCACTTCCTCCAC	AGCGGCTCTGGGAACACCTTGATT
27. CPNE7	I/E boundary	171	CGTAGCCGTTGTTCCCCGAGATGT	CGACTGCTGGCGCCGTGAGGT
28. Hs.300252	No gap	193	GGAAGACGTGTGCGGATAGGA	CCCCTCAGCCTCCCAAAGTT
29. SPG7	I/E boundary	278	GCATCGGGCCCATCTCCTTCC	CCTCTGCGGTGCGATCATTTTCTT
30. Hs.50852	No gap	250	GCCAGCCGCGTAGAATGA	ACCCGCCTTACCAGTGAGCAG
31. Hs.148468	No gap	126	TCTGCCTTGTGGCTGTGTAACTGG	CGTAAATCTTCGGCTGCTTCC
32. Hs.297449	No gap	140	CTGCGCTTGGACTTGGTGTGACTT	TATGGGGCTAGAGCTGAGGGAACC
33. Hs.318685/Hs.290681	I/E boundary	168	CCCCCACGGCGACTCCTC	TGGTATTCACAGGTGCGGTCAGC
34. Hs.42390	I/E boundary	294	AGCCGGCTACCCCCTCTCC	GCCGTCCAGCCTGCGAAGTC
35. Hs.211462	No gap	227	GAAGCTTCCGGGGTCTGTCTG	TTCTTTACTGTTTTCCCACCTCAC
36. Hs.26975	I/E boundary	203	CCGACATTGCTTCCTTCCTTCCTG	GCTCAACGCTCGCACAACTCCA
37. Hs.272320	I/E boundary	191	CGCGGCCGAGGTCTTCC	GCCGGCCGTGCAGGGTGTCC
38. CDH15	I/E boundary	176	CCCCGGGGCGTCTTCTCTATC	CCAGGTCCAGGGCAAACGCTCTTA
39. Hs.123661	I/E boundary	286	AGCGCCCTGGAGGTGGAGTGG	TGCCGGGTCATGAGGGGTGGAAGT
40. Hs.49892	I/E boundary	118	CCAGCCATGCGGTCACA	TCAGGGCTGCAGGGGAGGTA
41. Hs.192155	No gap	254	AGACTCCCTCACCCTCCAACC	TCAGAGCGGTGGAACAAACAT
42. CBFA2T3	I/E boundary	80	GGGCCTGGTGAACTCGACATTGAC	ACGGCCGCAGAGGGAAGTTGGT
43. Hs.27445	I/E boundary	335	TCGCAGCACCCCTACGGAGAAC	TCAAAGGCCCTGCTGGACTGGAT
44. Hs.93834	No gap	253	CCCCGCCTCACCCCTCTGT	AGTAGCCGGGCCTTTCTCTGC
45. GALNS	I/E boundary	227	CACCCCCTGAAGCACGGATTTGAT	GGGTGGTGCCGTGCCTGTCTCTTA
46. APRT	I/E boundary	316	GACTGGGCTGCGTGCTCATCC	AGGCCCTGTGGTCACTCATACTGC
47. Hs.122908	I/E boundary	287	ATGAGCCCTGGGGACATGGAG	AGGGGCGCTGGGGACAGGAAG
48. Hs.79077	I/E boundary	285	ACTCGGCGGCCACAGACATCACG	GGTAGGCGGACAGGGCGAAGTCAC
49. Hs.169850	I/E boundary	136	GCCCCCGCTGCCTCCTCT	GGGGGTGTCCGGGTCTCAGTAA
50. Hs.105280	I/E boundary	346	AAAGCCACCGCAGCGACCCCAACC	GCTCGGCCCCGGCTCCCCTTCTG
51. Hs.253790	No gap	114	TGTGTTGGGATGAGGCGGAGTGC	CCATTTATTACCCCACGGCTGTTA
52. MVD	I/E boundary	169	CTGAATGGCCGGGAGGAGGATGTG	CCGCAGCCGTGGGGAAGTTGTT
53. CYBA	I/E boundary	115	GCGAGCGGCATCTACCTACTGG	GGTTGCTGGGCGGCTGCTTGATGG

54. Hs.174074	I/E boundary	180	GCTGGGGCGCGGCTCAAC	CCTCCTCCTCGCTGGGCTTCTT
55. Hs.93670	I/E boundary	205	TTCGCGGTCATCGTCCTACTCTGG	GGCCGGCTCTCCCTTGGTTCT
56. Hs.120392	I/E boundary	224	CTACCCCAGGCTCAAGTCAGA	GGAATGCGTGCCGAGTGGAGAT
57. Hs.54925	No gap	242	CCGGGGCGAGGGCTTCCACT	CCCCGAGGCGCCTTCCACAA
58. Hs.257226	No gap	128	CAGTCCTGAGTTGATCGCTAATAA	CTTCCGGCCGCTCTGTGG
59. BANP	I/E boundary	256	CCCCACGCGCCCTGCTCTCAC	GCGGCGGCGGCTGCTCTTTC
60. Hs.181181	No gap	238	CTGGATATCTGCGAGCTGCGTGTT	GCCCTGTTTGGAGTTGCCTGTTCT
61. Hs.142569	No gap	214	GCGGGGCTGAACGGCACTG	GATGGGGTTGCTGGCTGGTCCTGA
62. Hs.189419	No gap	172	GGTGGCCTCGGTTCGGTTCTTC	GATGGGCCGAGGGGATGGA
63. SLC7A5	I/E boundary	269	TCTTTGCCTATGGAGGATGG	ATTGACGGAGCCAAAGCAGG
64. AI539360	No gap	286	CCGGGTGGCCTGGTTTCA	GCTGGCACGGGGGAGGAG
65. Hs.67991	I/E boundary	181	GCGGCGCGGAGAAGACG	TGAACCGAGAGGGAGGCATTTA
66. Hs.281957	No gap	293	GGGGCACTCTATAAAACTCCTGA	CTCTCACACGCCCACACCAAT
67. Hs.146338	No gap	231	TTGCTGGGATGTTTATGTTGTTTG	GGGGGTGGCCTCGTCCTCAG
68. Hs.81505	No gap	241	TAGAATGAGGCCTTGTGATGTGAG	TTCCCCTAAATAGCTGTTGTAAT
69. MAP1A	I/E boundary	114	TGTTGTTACGGAAAGCAGCAGTGT	CCGGAAGGCAGAAGGGAGTGT
70. Hs.7970	I/E boundary	149	CCGGCGGGAGGCAGGAGGAGT	GCGGCGGTAGGTCAGGCAGTTGTC
71. Hs.227170	No gap	429	AGAACTAGCATCTGTTCACC	ATGGTGCTGCTTGTAGCAAG
72. Hs.150856	No gap	157	AACAACATGGGGTTTAAAGGG	GTCTCCACGTCAGCAACTG
73. Hs.118944/Hs.307087	I/E boundary	240	GCGGGGTGCGTCGGGCTCTG	AGGGGTGCTGCTTTCGGGGTCTCA
74. Hs.112744	I/E boundary	234	CCACGCGGGTCCATTAGAAGC	GGGAAGCCACCTCGGACTGA
75. AA994450	I/E boundary	255	GAAGCGAAAGGGAGGATGAGGTG	CCGCAGCAGGAGGAAAAC
Rejected: Late C_T values				
76. PRDM7	No gap	315	ATTCAACCCCAGAAAGACATAAAC	TGCATGCAGGCTTAACAC
77. Hs.235375	I/E boundary	268	CATCAGCTCTGCGTTGTTCAC	AGCCAGCTGTTCCTAGTCTTCAT
78. Hs.244883	No gap	264	TCAAGCAGCAAAAAGCAGAAAAAC	TGAGCGTGGCCCTGAGTGTAG
79. H08755	No gap	120	AAAAATTAAAGAGACCAAGAT	ATAAGAAAAATAACCATAAACAA
80. BF872709	No gap	110	GAGGGGATATTGCATTTGTCAGA	GAATGTTTTGCTCTGGTGGTCA
81. Hs.171193	No gap	122	TAGTTCAGATTTTGTTTTGTGTTC	AAGTCTGTAGTTCTATTGGATTCA
82. BG697046	I/E boundary	173	AATGATATGGTGCCTAAGAGAAAA	TTGGATGGGGAAACAGAAGTC

83. Hs.149378	I/E boundary	226	GCGCCGCGGCTTTTGTGGAGATAG	CTGCCGCGGCGCCTGGAACT
84. Hs.204432	I/E boundary	275	CCTTGCGCCCGTCTTGTTGTTTCT	CGCGCCGACTCTTCTCCATTCC
85. Hs.266702	No gap	274	AAGGGCAGTTTGGAGAATA	GAGGCAGACAGAGTTTCGTTAGTA
Rejected: Bad melt curves				
86. Hs.123626	No gap	255	GCCGGCAGCCCAGAAGC	GCGGGTGGGGTGAGTGAGG
87. AI684092	I/E boundary	212	TCGGGGCGGGATGAGAAAATG	ACCGGCTGTGTCCCCAAAGTCCT
88. BE175461	No gap	201	TGCTCCCTGCCACACTCA	CTACGGTCAAAAACTAAAAACACG
89. AV751950	I/E boundary	224	AGGGGATTTTCTTGGCTTTTTC	TGGATTTCACTCTGGGCTTTTC
90. Hs.271383	No gap	235	CACGCTGCGGCTGGTCTCC	GGGCTCCACGCTCATCACTCAC
91. Hs.254912	No gap	232	TCGGAGTGGGAGGCAGAGTGA	CCGGGGCAGGGAGTTCGT
92. Hs.308644	No gap	639	GGTGCCCTGCCTTTCCTCTG	CCGATGGGGTTTACTGG
93. IL17C	I/E boundary	262	CCACAGGGGGAGGCACGAGAGG	CCGGCGGCGCAGCACCAG
94. Hs.144844	I/E boundary	266	GGCGGCAGCGGTGGGGGTGAG	GCCGGAGACATGTCCAGGCGGAAA
	I/E boundary	151	ACCCCCGGCAGATCAAGCGTTCC	GGCGGCAGCGGTGGGGGTGAGTTA
95. BE314227	No gap	287	CTCAGCAGCAGCCTCAGAAT	ACTCCGCCCCTCCGTGTCC
96. Hs.129559	No gap	178	CCCGGGCCGCACATCTCA	GTTTTCCGACCGCCTCTGCTC
97. Hs.128894	No gap	171	CCCAGAGGATGGAGACTTTT	AGCCCACATGCCCATCCACAGGT
98. Hs.125769	No gap	258	CCTTCCCGCAGAGCCGCAGACA	GGCCCTCTGCACCCACCAACATC
99. CA5A	I/E boundary	150	TTAAAGCTCGGGGCCCATCATCAG	CGAGCCCGCGTAGGTCCAGTAATC
100. JPH3	I/E boundary	204	CCGGTTGCTGCGTTGGGACTT	TTTTGCTACCGCGACATCTCATCA
101. Hs.297345	No gap	205	CTCCAGGGCGGCAGACTCTC	CAGGCATGGGAGGCAGGTGT
102. Hs.281929	No gap	191	AGAGGCCCCCGGGAGTGAAGAAGA	AGAGGCTGTGTTCCCGAGAGGT
103. Hs.156231	No gap	291	ATTGCAAGGCGCTGGGGCTTCC	CACGGGCATTTAAGGCTCTGGGAG
104. Hs.133389	No gap	300	CACTGGGAGCCTCTGAAATACACC	AGAAGCGGACTCGGGAACTG
Rejected: In vitro				
105. Hs.246866	I/E boundary	162	ATGAAGATGGCCGGAGAAAAA	GCGTGGCCATGGGGAGGTC
	I/E boundary	150	ATAAATATGAGCCCCTGCCTGACA	CCCGTGCGGTATGGTGAAGA
106. Hs.171376	I/E boundary	300	GCACCTATGCTCTCCCTGTGTTG	TGAGTCTCATATGACAAGTTCCCT
	I/E boundary	132	AGTTTTTACTTTTGGTAGCGTTGA	TGCATAAAATCACAGATACCTTTG

107.	Hs.232124	I/E boundary	278	GATGGCGTCTCACTCTGTCACC	AGCTGGGCCACTGTCACTGT
108.	AI219773	I/E boundary	193	GGCCCATTTACAACTTGCTTAGAT	TGATGCCACGCTGAGATGC
		No gap	219	GCTGAGTTGATCGTTAAC	CTACAATGGGTAAACGATGG
109.	AW849809	I/E boundary	207	AGGGTCTCGGTGTCTCTGG	AGGGCCTTGTCAAACTCACC
		I/E boundary	106	CCCGGTGCTGGCTTGTGGA	TGAAACTGGGAGGGCTGGGATTAC
110.	Hs.222186	I/E boundary	295	CCCCACCCACCGTCGAAAGCACTA	GAAGGAGAGCCCACGGGTTGAAGC
111.	AL046285	No gap	98	GGAACTTACAGACGGAAACAT	CTCACAGGGCCCGATACG
112.	AA824376	I/E boundary	177	TGGGTCTTGGCAGGAATGAGGTG	CTGCGTTATGTCGTTTGGTTGATG
113.	Hs.283246	No gap	162	TAGAAAACACCCCACCCCCACCTC	ATCCCCTCCCTCCTCCAGTCACG
114.	AI797636	I/E boundary	231	ACGGTGGGGTGTCTGCTAAAAAG	GTGGGGAGGTGCAGACAGATTC
		I/E boundary	221	GAAGAGGCAGATCCACAGACACAG	GGACATGACCATAGCCTTACAGT
115.	Hs.259277	No gap	180	GGAAGCTTCATGTAGCTCAC	CGACACTCAAACTTCAGCATG
		I/E boundary	165	GCGGACCTCGTGGCTTTTTG	CAGTGCTTGTCCCCTTGTATTGAA
116.	Hs.149004	No gap	294	TTGAAGCTGAGATGAAAGCC	ACATGGTTTGCAGTGCTACC

House keeping genes

117.	Cyclophilin A	I/E boundary	193	AATGCTGGACCCAACACAAATG	CACAATATTCATGCCTTCTTTCACTTT
118.	GAPDH	I/E boundary	219	GCGGGGCTCTCCAGAACATCAT	CCAGCCCCAGCGTCAAAGGTG
119.	ATP5A	I/E boundary	207	TTGCGGAGGAACATTGGTG	TCTTCAGGCCTGGGTTTT
120.	RPB70	I/E boundary	106	CGAGGAGGCGGAACAAGTC	TGGCAGGAGGAAGTCACAGC

Chapter 3

Chapter 3: TSG16: Candidate gene analysis

3.1 Introduction.

Chapter 2 describes the identification and expression analysis of 75 candidate tumour suppressor genes mapping to the 16q24.3 LOH interval. The initial screen, based on expression variability in breast cancer cell lines, identified three possible candidates, *CBFA2T3 (MTG16), CYBA* and *FBXO31* (Hs.7970). It was hypothesized that tumour-suppressor genes associated with the 16q24.3 LOH may be aberrantly expressed due to the presence of mutations or promoter hypermethylation. This was found to be true for the known tumour suppressor genes *SYK* and *CDKN2A* (Chapter 2). Seven of the 75 candidates have been previously analysed for mutations in breast tumour DNA: SPG7 (Settasatian *et al*, 1999), BBC1 (Moerland *et al*, 1997), CDK10 (Crawford *et al*, 1999), CPNE7 (Savino *et al*, 1999), FANCA (Cleton-Jansen *et al*, 1999), GAS11 and C16ORF3 (Whitmore *et al*, 1998b). None of these genes were found to harbour any mutations in sporadic breast cancer DNA samples.

Of the remaining transcripts, several were considered as good candidates based on additional functional evidence. For example, microcell mediated chromosome transfer experiments identified a 16q24 YAC clone capable of inducing cellular senescence in breast cancer cell lines (Reddy *et al*, 2000). YAC 792E1 was mapped to the 16q24.3 LOH interval, and the availability of the sequenced physical map permitted the identification of all candidate genes within this YAC (Chapter 4).

3.1.1) Ankyrin domains and BARD1. In addition, other genes were considered candidates based on their homologies to other known human or mouse proteins. For example, TSG16 possessed an ankyrin (ANK) repeat domain that exhibited significant homology to the ankyrin repeat motif of the BRCA1-associated RING domain protein 1 (BARD1). ANK domains are found in a variety of proteins ranging from transcription factors to toxins, and are characterized by a 33 amino acid repeat with the consensus sequence xGxTPLHxAAxxGHxxxV/ AxxLLxxN/ Dxxxx, where x represents any amino acid (Lux *et al*, 1990). ANK domains function as sites for protein-protein interactions and X-ray crystallography has revealed an L-shaped structure consisting of α-helices and β–hairpins (Sedgwich and Smerdon, 1999).

BARD1 was first identified through a yeast two-hybrid screening with BRCA1. BARD1 consists of an amino terminal RING domain, an ANK domain consisting of 3 repeats, and a c-terminal BRCT domain (Wu *et al,* 1996). The function of BARD1 has not yet been fully elucidated but the identification of two somatic mutations in sporadic breast cancer and uterine cancer (Thai *et al,* 1998) suggests a possible role in cancer development.

The RING domain is a novel zinc-finger motif (Lovering *et al*, 1993) thought to be involved in mediation of protein-protein interactions, and in the ubiquitination of proteins for subsequent degradation (Borden, 2000, Freemont, 2000 and Lorick *et al*, 1999). BARD1 and BRCA1 interact via their RING domains (Wu *et al*, 1996) and co-localize with Rad51 in BRCA1 nuclear dots during S-phase of the cell cycle (Jin *et al*, 1997). The BRCA1-BARD1 heterodimeric RING finger complex exhibits ubiquitin ligase activity that can be disrupted by a breast cancer derived RING finger mutation in BRCA1 (Hashizume *et al*, 2001; Brzovic, *et al,* 2001).

BARD1 also interacts with the cleavage stimulation factor CstF-50 (Kleiman and Manley, 1999), which is involved in the polyadenylation of mRNA. In order for mRNA polyadenylation to occur, the polyadenylation factor and RNA polymerase II must associate at the site of polyadenylation. BARD1 interacts with RNA polymerase II, and the interaction of CstF-50 and BARD1 *in vivo* prevents the polyadenylation of mRNA (Kleiman and Manley, 1999). These findings suggest that CstF-50 and BARD1 may regulate RNA processing during the cell cycle, and during repair of damage DNA.

In addition to the proteins already mentioned, BARD1 also interacts with the Iκβs proteins Iκβ-α and Bcl3 (Dechend *et al*, 2000). Both of these proteins contain ANK domains and are involved in the inhibition of the NFκβ/Rel transcription activators (Verma *et al*, 1995). NFκβ/Rel activation involves IκB degradation, nuclear translocation of the NFκβ/Rel protein, and promotion of cell survival by inhibiting apoptosis (Sovak *et al*, 1997). The ANK domains of Iκβ proteins function as the nuclear import sequences (Sachdev *et al*, 1998) and are highly homologous to the ANK domain of TSG16. It has been shown that NFκβ/Rel activation is an early event in breast carcinogenesis, occurring prior to malignant transformation (Kim *et al*, 2000). Genes regulated by NFκβ/Rel include those involved in immune and inflammatory responses, cellular proliferation and

adhesion molecules (Reviewed in Baeuerle and Baltimore, 1996). The interaction of BARD1 with CstF-50, Iκβ-α and Bcl3, all occur via their respective ANK domains.

Since TSG16 maps to the 16q24.3 LOH interval, and ankyrin repeat domains function as sites for protein-protein interactions, it was hypothesized that TSG16 could interact indirectly with BRCA1 by sharing common protein partners, and therefore may itself function in breast carcinogenesis. *In silico* analysis identified matching ESTs with reported down regulation in nasopharyngeal carcinoma (NPC) (accession number NM_013275, GeneBank entry only). Interestingly, NPC also exhibits 16q LOH (Shao *et al*, 2001, Yan *et al,* 1999) and Li *et al*, 2001), further supporting a tumour suppressor function for TSG16. The NPC 16q LOH correlates with a high Epstein-Barr virus (EBV) titer, and both contribute to the aggressive etiology of this disease (Shao *et al*, 2001). Additional evidence for a role of TSG16 in cancer was from an unconfirmed report indicating that 16q24.3 may be the location of an inherited gene for a predisposition to breast cancer (Giles *et al,* 1992).

The genomic structure and expression profile of TSG16 was initially established. TSG16 was excluded as a candidate tumour suppressor as it failed to exhibit any breast cancer specific mutations. Subsequent functional analyses were performed to determine the biological function of TSG16. A homologous gene on chromosome 18, termed TSG18, was also identified and characterized, and together with TSG16, form a novel protein family.

3.2 Materials and Methods

3.2.1) TSG16 transcript characterization. TSG16 was initially identified from exons trapped from the 16q24.3 genomic DNA interval (Whitmore *et al*, 1998a). These exons are now recognized as the LZ16 mRNA sequence (NM_013275). Isolation of matching and/or homologous nucleotide and protein sequences were performed using the blastn and blastp algorithms respectively, at the National Center For Biotechnology Information (NCBI (3)). RT-PCR from normal breast RNA (BD Biosciences) was performed using Superscript RNase H- reverse transcriptase (Invitrogen) permitting the linkage of

physically adjacent Unigene clusters and ESTs to LZ16. RT-PCR was performed as earlier described (2.2.2.5). In brief, 3 μg of total RNA or 100 ng of polyA$^+$ mRNA was added to either 50 pmol of random hexamers (Perkin Elmer) or 100 pmol of oligo dT. The RNA was heated to 65 °C for 5 minutes, returned to ice for 1 minute before the addition of 1X PCR buffer, 4 mM DTT and 0.2 mM dNTPs. This was incubated at 42°C for 2 minutes before the addition of 200 units of Superscript RNase H- reverse transcriptase, followed by a final incubation at 42°C for 30 minutes, and the reaction terminated with heating to 70°C for 10 minutes. The secondary cDNA amplification was performed as earlier described (2.2.1.1) with 1-20 ng of cDNA template. The RT-PCR primers used to confirm all splice junctions are presented in Table 1. The transcript was further extended by 3'-Race experiments (Invitrogen) resulting in the generation of full length TSG16 transcript. Mr. Chatri Settasatian performed the RACE experiments.

The gene prediction program, GeneFINDER (9) was used for transcript extension and characterization. The protein prediction programs ProDom (10), Pfam (11), PESTfind (12) and PfScan (13) were used for preliminary functional analysis.

3.2.2) Northern blot analysis. A human multiple tissue northern blot membrane (BD Biosciences) was probed according to the manufacture's instructions. RT-PCR products were generated with the primer combination 5' GTGGCCCTTCTCATGCAGAT and 5' TGTCACCGTGGGGACAG for TSG16 and 5' ACTTCGTCACGGTGGAATC and 5' TGACACATGCTTTTCCTTGG for the homologous gene, TSG18. Membranes were pre-soaked in 5XSSC and pre-hybridized at 65°C for 2 h in 10 ml of ExpressHyb solution containing denatured salmon sperm DNA (100 μg/ml). After the cDNA probe had been labelled, QIAquick cleaned and pre-reassociated (2.2.1.8), the hybridization solution was removed and replaced with 10 ml of fresh ExpressHyb solution containing the probe and denatured salmon sperm DNA (100 μg/ml). Probes were hybridized overnight at 65°C and the following day membranes were washed three times in 2XSSC, 1% SDS for 10 min at room temperature. A high stringency wash of 0.1% SSC and 1% SDS at 65°C for 30 min, was used when direct monitoring of the filter showed high background signal.

3.2.3) Single-stranded conformation analysis and sequencing. SSCA was performed on the entire open reading frame and short flanking intronic regions of TSG16. Hex-labelled primers were designed in intronic regions flanking each exon, except for exon 9, which was divided into 31 overlapping amplicons (Table 2). Mutational analysis was performed on 46 paired breast tumour DNA samples and normal DNA samples from peripheral blood lymphocytes of the same patient. In addition, 100 peripheral blood lymphocyte DNA samples from non-BRCA1/BRCA2 familial breast cancer patients were screened. These samples were provided by kConFab (Kathleen Cuningham Foundation CONsortium for research into Familial Breast cancer). PCR reactions were performed in 10 µl reactions containing 1X PCR buffer, 2 mM dNTPs, 1.5 mM $MgCl_2$, 1 mM of each primer, 1 unit of Taq polymerase and 30 ng of patient DNA. Amplification involved 10 cycles of 94°C for 1 minute; 55°C for 2 minutes; 72°C for 2 minutes, followed by a further 25 cycles of 94°C for 1 minute; 60°C for 2 minutes; 72°C for 2 minutes, with a final extension of 72°C for 7 minutes. Following the PCR reaction, 10 µl of formamide loading buffer was added to each sample, mixed, incubated at 100°C for 5 minutes, placed on ice and resolved on 4% acrylamide gels using the GelScan 2000 (Corbett Research) and permitting the rapid real time acquisition of the data. PCR products showing conformational change were re-amplified with unlabelled primers and sequenced using BigDye terminator chemistry (Perkin-Elmer 2.2.1.12). Population screening of all polymorphisms identified in the patient samples was performed with 30 ng of genomic DNA isolated from the blood lymphocytes of normal individuals. Olivia McKenzie, Anthony Bais, Joanna Crawford and myself all contributed to the screening effort.

3.2.4) Plasmid constructs. All constructs were amplified from normal breast RNA (BD Biosciences) using Superscript RNase H- reverse transcriptase (Invitrogen) according to the manufacture's protocol. A total of 7 expression constructs were cloned and are described in detail in Table 3. TSG16 was amplified in 3 overlapping fragments, each cloned into either pSP72 or pBluescript (Invitrogen) (Table 4) and sequentially ligated together using unique restriction sites. The resulting full-length transcript was subsequently sub-cloned into the *Hind III* and *Cla I* sites of the mammalian expression

vector pLNCX2 (BD Biosciences). The 7995 bp open reading of TSG16 is defined by the forward primer 5' AGGAGCAGGACGATGCCCCG and the reverse primer 5' ACAAAGTCGTCGTTGACGTC. A C-terminal myc-tag was introduced for detection of the recombinant protein.

A 1380 bp fragment containing the ANK domain of TSG16 was amplified with the forward primer 5' GTGGCCCTTCTCATGCAGAT and the reverse primer 5' TGCCGTCGACTGAACTGGAAGGTG and subsequently cloned into pTarget (Promega). A N-terminal Flag-tag was introduced to produce the pTarget/TSG16-Ankyrin construct. PIAS1 was amplified with the primer combination 5' ATGGCGGACAGTGCGGAACT and 5' CAGTCCAATGAAATAATGTCTGG and cloned into pTarget (Table 3). The PIAS3 full length coding region was cloned using Gateway Technology (Invitrogen) resulting in the N-terminal GST fusion constructs PIAS3/pDEST27, or the non-tagged mammalian expression constructs PIAS3/pDEST12.2. The human STAT1 and STAT3 open reading frames were amplified with primer pairs containing unique restriction sites for directional cloning into pcDNA3.1. STAT1 was cloned into the EcoRI and NotI sites, and STAT3 was cloned into the HindIII and NotI sites. The STAT3 reporter construct pAPRE-luc, contains four copies of the APRE (Acute Phase Response Element) STAT3 binding site linked to the luciferase reporter gene (Nakajima *et al*, 1996). The STAT1 reporter construct pGAS-TA-Luc, contained the GAS enhancer element derived from the IFNγ activation sequence, which was specific for STAT1 homodimers (BD Biosciences). The pRL-CMV construct expressing *Renilla* luciferase was included as a transfection control (Promega). All primer sequences, cloning sites and vector backbones for all constructs are summarized in Table 3.

3.2.5) Cell culture and transient transfections. HEK 293T cells were obtained from the American Type Culture Collection (ATCC) and maintained in DMEM (Invitrogen) supplemented with 10% fetal calf serum, 2mM L-glutamine and 10mg/liter penicillin and lentamycin. Cells were grown at 37°C in 5% CO_2. Transfections were performed in Opti-MEM with lipofectAMINE 2000 according to the manufacture's instructions (Invitrogen). For 6-well dishes, $5X10^5$ cells were plated and grown overnight in complete

media. The following day the cells reached approximately 60% confluency and the medium was replaced with 2 ml of Opti-MEM. 2 μg of DNA was mixed with 500 μl of Opti-MEM, and 5 μl lipofectAMINE 2000 was mixed with a second 500 μl aliquot of Opti-MEM. The aliquots were incubated for 5 min at room temperature before they were mixed and incubated for an additional 20 min at room temperature. Complexes were added to the cells and incubated overnight at 37°C. The following day the medium was replaced with 2 ml of Opti-MEM and recombinant proteins were analyzed 48 h post-transfection. This transfection protocol was used in all experiments and the parameters were scaled up when the transfections were carried out in larger flasks.

3.2.6 Retroviral infection. The breast cancer cell lines MCF-7 and SKBR-3, HEK 293T cells, and the immortalized breast epithelial cell line MCF-12A, were maintained as previously described (2.2.2.4 and 3.2.5). Transfections of the 293T packaging cell line were performed with 10 μg of pLNCX2 retroviral expression vectors, 8 μg of pVPack-VSV-G and 8 μg of pVPack-GP (Stratagene) in 1.5 ml of Opti-MEM. 60 μl of LipofectAMINE 2000 reagent (Invitrogen) was added to 1.5 ml of Opti-MEM, incubated for 5 min, mixed with the DNA aliquot, and added to cells grown in 100 mm plates with Opti-MEM medium without FCS. The medium was replaced 12 h later and the virus-containing supernatants were harvested 48 h post-transfection. Supernatants were filtered through 0.45 μm syringe filters and polybrene (Sigma) was added to a final concentration of 8 μg /ml. Cells to be infected were plated in 6-well plates at 40% confluency and 24 h later the cell medium was removed and replaced with 1 ml of virus containing supernatant. 3 h later an additional 2 ml of complete medium was added. Cell assays were aborted as the size of TSG16 exceeded the limits for efficient retroviral packaging.

3.2.7) Immunofluoresence and STAT reporter assay. For immunofluoresence HEK 293T cells were seeded at $5X10^5$ cells/plate in 35mm dishes on glass cover slips and transfected 24 h later with 1-4 μg of the respective plasmid DNA. 24 h post transfection cells were stimulated with 10 ng/ml of LIF (Sigma) for STAT3 activation, or 10 ng/ml IFNγ for STAT1 activation, fixed 20 min later in PBS containing 3.7% formaldehyde for 15 min at room temperature, and permeabilized with 0.5% Triton X-100 for 5 min at 4°C.

Cells were incubated with either myc monoclonal antibodies or PIAS3 polyclonal antibodies (Santa Cruz Biotechnology Inc.) or Flag monoclonal antibodies (Sigma) followed by incubation with a either CY3-conjugated goat anti-mouse IgG (Jackson ImmunoResearch Laboratories, Inc.) or FITC-conjugated goat anti-rabbit IgG (Silenus, Australia). Cell nuclei were stained with DAPI and images obtained using CytoVision (Applied Imaging).

For gene reporter assays HEK 293T cells were seeded at $1X10^5$ cells/well in 24 well plates and transfected 24 h later with 0.7μg of total plasmid DNA. The DNA mixture contained 20 ng of the *firefly* luciferase reporter pRL-CMV, 180 ng of the *Renilla* luciferase reporter pAPRE-luc, and a 500 ng cocktail of expression constructs. 24 hours post-transfection cells were stimulated with 10 ng/ml LIF for STAT3 activation, or 10 ng/ml IFNγ for STAT1, and lysed 12 h later in 1X Reporter Lysis Buffer (Promega) containing 20 mM Na_2VO_3, 100mM NaF and 1X protease inhibitors. A dual luciferase reporter assay was performed in each lysate, as essentially described by the suppliers (Promega). Transfections were performed in triplicate and samples were assayed in duplicate using a luminometer. Data presented as the mean of two independent experiments ± standard error.

3.2.8) Immunoprecipitation. HEK 293T cells ($3X10^6$ cells/plate) were seeded in 100mm dishes transfected 24 h later with 15 μg of the respective plasmid DNA. 24 h post-transfection cells were stimulated with 10 ng/ml LIF for STAT3 activation, or 10 ng/ml IFNγ for STAT1, and lysed in 1% Brij-35, 10% glycerol, 50mM Tris-HCl, 150mM NaCl, 100mM NaF, 2mM PMSF, 20mM Na_2VO_3, 1mM DTT, 2mM EDTA, 2mM EGTA and 1X protease inhibitors (Roche). Proteins were immunoprecipitated for 2 h with either glutathione-sepharose (Amersham Pharmacia Biotech) or one of anti-PIAS3, anti-PIAS1, anti-flag, or anti-Myc antibodies coupled to protein-G sepharose (Amersham Pharmacia Biotech). Immunoprecipitates were resolved by NuPAGE (Invitrogen) and transferred to nitrocellulose (Amersham Pharmacia Biotech). Blots were probed with various primary antibodies and visualized with the relevant HRP-conjugated secondary antibodies (Silenus, Australia).

3.3 Results.

3.3.1) TSG16 transcript characterization, genomic structure and expression profile. We initially identified TSG16 from exon trapping experiments within the 16q24.3 LOH interval. These exons are now recognized as the LZ16 mRNA sequence of 1603 bp (NM_013275, Hs.42390). *In silico* analysis of LZ16 allowed the assembly of various overlapping human Unigene clusters and ESTs. Physically adjacent ESTs were linked by RT-PCR experiments yielding a partial TSG16 cDNA sequence of 2290 bp. Homologous mouse ESTs were identified at the 3' end of the gene allowing further extension of the transcript. Primer pairs were designed to these putative cDNA sequences and 3'-RACE experiments confirmed their presence (data not shown.) Similarly, these assembled cDNA sequences were subsequently used to identify 3' extending homologous human ESTs. The availability of the genomic sequence within this region of 16q24.3 (Powell *et al,* 2002) enabled further 5' sequence extension. This extensive cloning procedure identified a full-length TSG16 transcript of 9307 bp (Fig 1a).

The genomic structure of TSG16 was established by the alignment of full-length TSG16 with the available genomic sequence at 16q24.3 (accession numbers AC105409 and AC009098, Fig 7 Chapter 2). The TSG16 gene consists of 13 exons ranging in size from 85bp to 6.58kb and spans a genomic interval of approximately 263kb. A 113kb intron is present between exons 2 and 3 and all splice junctions conformed to the consensus sequence (Shapiro and Senapathy, 1987). TSG16 contains an ORF of 7995 bp with a start codon in exon 3 at position 462 and a stop codon in exon 13 at position 8454 (Fig 1a). This same stop codon was identified in the corresponding mouse cDNA clone. The homology between these two orthologues significantly decreased 3' of this stop codon, supporting the likelihood that this sequence is the 3' untranslated region (3'UTR.) The 3' UTR of 852 bp included a polyadenylation signal, AATAAA, at nucleotide positions 9288-9293, 13bp before the polyadenylation site. The start codon at position 462 conformed to the Kozak consensus sequence.

The gene prediction program, GeneFINDER, successfully recognized the open reading frame (Fig 1a) and splice junctions (Table 1a) of TSG16, the later subsequently confirmed by RT-PCR analysis from normal breast RNA. Primer pairs were designed to

span each of the twelve exon boundaries of TSG16 (Table 1b) and all amplicons were sequenced verified.

Examination of the EST distribution across the full-length transcript revealed an absence of ESTs at the 3' end of TSG16. This absence is partially a consequence of the long adenosine nucleotide (poly(A)) tracts present throughout the sequence of TSG16, acting as priming sites for oligo-dT in reverse transcription experiments. For example, blastn and blastp analyses identified LZ16 and subsequent examination revealed that LZ16 was an incomplete TSG16 cDNA, reverse transcribed from the poly(A)-rich tract "AAAAAAATAAAGTGAAAAA" starting at nucleotide 1832. In addition, the extra-large transcriptional size of TSG16 exceeds the limits of conventional reverse transcription, further contributing to the database abundance of partial TSG16 transcripts.

The expression profile of the TSG16 transcript was determined by Northern blot analysis (Fig 2a). A cDNA fragment of 919 bp, encompassing exons 6 to 9, identified a predominant 9kb transcript expressed across all tissues examined, with abundant expression in the heart, placenta, kidney and pancreas. Smaller transcripts of 3.2, 3.0 and 1.0kb were also detected with a lesser intensity to that of the 9kb transcript. cDNA probes generated from various parts of the TSG16 transcript also identified this predominant 9 kb band (data not shown). The smaller bands are considered to represent alternatively spliced transcripts or homologues present elsewhere in the genome. For example, Blastn analyses of the Expressed Sequence Tag (EST) and High Throughput Genomic Sequence (htgs) databases with full-length TSG16 identified three alternatively spliced TSG16 isoforms; Hs.380715, T17 and Hs.457500. Hs.380715 exhibits 1460 bp of sequence with a 472 bp ORF identical to that of TSG16 from exon 1 to exon 4, followed by a unique 3' terminal exon located within intron 4. Similarly, the transcriptional isoform T17 selectively utilizes a unique 3' terminal exon situated within intron 7 of TSG16. The mRNA transcripts identified by T17 were similar to those detected by exons 6-9 of TSG16, except for the 9kb transcript (data not shown). The third isoform, represented by a partial cDNA sequence Hs.457500, excludes exons 8 and 9 from the TSG16. An additional gene, Hs.26975 (EST AA995728) was shown to reside within intron 4 of TSG16. Sequence analysis and subsequent RT-PCR experiments established Hs.26975 as

an independent transcript, transcribed in the opposite direction to that of TSG16 (data not shown).

In addition, homologous TSG16 ESTs have been generated from a range of different tissue cDNA libraries, suggesting a ubiquitous expression profile.

3.3.2) SSCA of TSG16.

3.3.2.1) Sporadic Breast Tumours. Two series of breast cancer patients were analysed for this study. Collaborators, according to the World Health Organization criteria, carried out histopathological classification of each tumour specimen and the subsequent LOH analysis (Departments of Pathology and Human and Clinical Genetics, Leiden University Medical Center, Leiden, the Netherlands; Department of Hematology and Genetic Pathology, Flinders Medical Center, Flinders University, South Australia, Australia). Tumour tissue DNA and peripheral blood DNA from the same individual was prepared as previously described (Devilee *et al*, 1991).

Series 1 consisted of 189 patients operated on between 1986 and 1993 in three Dutch hospitals, a Dutch University and two peripheral centers. Tumour tissue was snap frozen within a few hours of resection. For DNA isolation, a tissue block was selected only if it contained at least 50% of tumour cells following examination of haematoxillin and eosin stained tissue sections by a pathologist. Tissue blocks that contained fewer than 50% of tumour cells were omitted from further analysis.

Series 2 consisted of 123 patients operated on between 1987 and 1997 at the Flinders Medical Center in Adelaide, Australia. Tumours specimens were either snap frozen or embedded into paraffin blocks. Prior to DNA isolation, tumour cells were microdissected from tissue sections mounted on glass slides so as to yield at least 80% tumour cells. In some instances, no peripheral blood was available such that pathologically identified paraffin embedded non-malignant lymph node was used instead.

A total of 45 genetic markers were used for the LOH analysis of breast tumour and matched DNA samples. Figure 3, adapted from Cleyton-Jansen *et al,* (2001) summarizes the patterns of 16q LOH in the tumours used for TSG16 SSCP analysis. Details regarding all markers can be obtained from NCBI (3). The physical order of the markers was determined by a combination of PCR based mapping on the chromosome 16

somatic cell hybrid map (Callen *et al*, 1995) and genomic sequence information. All the tumours examined exhibited 16q LOH and categorized as 16q24.3 loss, partial loss of 16q or complex loss (alternating loss and retention of markers).

The TSG16 coding sequence, surrounding splice sites and short flanking regions of intronic sequence were screened for mutation in 46 sporadic breast tumour DNA samples. All band shifts detected were subsequently screened in the DNA from peripheral blood lymphocytes of the same patient.

No mutations were found that were restricted to the tumour DNA. Twelve different mutations were found to be apparently homozygous in tumours but were heterozygous in the constitutional DNA from the same patient. This is expected as these tumours had LOH for this region of chromosome 16. Seven of the twelve mutations were silent polymorphisms as they were present in normal individuals and did not cause an amino acid change. Five mutations resulted in an amino acid change and four were polymorphic. Of particular interest was the mutation of cytosine to thymidine at position 1354 that resulted in an amino acid change from threonine to methionine (Table 5a). This threonine residue is conserved in both the human TSG18 homologue and the mouse ortholog of TSG16. This change is positioned 64 bp 3' of the ANK domains and was absent from 330 normal chromosomes. It was hypothesized that this low frequency missense mutation may alter the function of the encoding protein and thus contribute to the disease phenotype. Known tumour suppressor genes such as *BRCA1* and *SYK*, often exhibit truncation or frame shift mutations, resulting in altered gene expression (Hedenfalk *et al,* 2002). Missense mutations generally do not affect the expression levels of the transcript (Hupp, 1999), an observation supported by the uniform expression profile of TSG16 in several breast cancer cell lines exhibiting 16q LOH (2.3.3.3, Fig 8 listed as Hs.42390). The biological significance of this mutation awaits functional confirmation as this change maybe a consequence of a rare polymorphism, with LOH deleting the most predominantly observed allele. The possibility of additional mutations going undetected was an unavoidable limitation of SSCP analysis.

3.3.2.2) kConFab. Circumstantial evidence suggested that TSG16 might be involved in familial breast cancer. Since the region containing the ankyrin domains of TSG16 showed significant homology to BARD1, the BRCA1 interacting protein, a role in the BRCA1 pathway may be possible. Furthermore, an unconfirmed report of linkage analysis in familial breast cancer identified 16q24.3 as a disease candidate region (Giles *et al,* 1992). Approximately 50% of the ANKRD11 gene was screened with the regions selected based on the variants found in the tumour screening. These included the region containing the ankyrin domains and a major proportion of exon 9. Peripheral blood lymphocyte DNA samples from individuals with breast cancer from each of 100 families ascertained because of multiple cases of breast cancer were screened. All individuals with breast cancer have been previously tested by the kConFab consortium and did not possess BRCA1 and BRCA2 mutations. A number of variants on SSCA were detected that were polymorphisms as they were present in normal individuals.

3.3.2.2.1) Family 0019.99.006 (Fig 4). A heterozygous mutation was detected in one patient at nucleotide 5729 that resulted in a CCC triplet coding for proline being mutated to TCC that codes for serine, a non-conserved amino acid change (Table 5b). This mutation was not detected in chromosomes from unrelated normal individuals. From this pedigree DNA was available from two additional females with breast cancer and all were also heterozygous for this C to T mutation. DNA was extracted from a paraffin section of breast tumour derived from one of these individuals. Subsequent PCR amplification and sequencing established that the tumour was also heterozygous, (C/T)CT. Therefore it is unlikely that this mutation at nucleotide 5729 is related to the risk of breast cancer as it would be expected that the normal allele would be lost in the tumour.

3.3.2.2.2) Family 0015.99.002 (Fig 5). Two females with breast cancer were heterozygous for a G to A change that resulted in a non-conserved amino acid change from alanine to threonine. This change was detected in 1/200 normal chromosomes (0.5%). An additional 4 females are affected but DNA samples were unavailable for 3 of these individuals and the fourth did not carry the nucleotide change. An obligate male carrier was also identified and shown to exhibit the nucleotide change. The same change

was seen in another pedigree but there was only one DNA sample available from an individual with breast cancer (0007.99.007).

3.3.3) Cell growth suppression assay. Attempts were made to assess the effect of TSG16 on the cellular growth patterns of several different breast cancer cell lines. It was hypothesized that exogenous TSG16 expression may suppress the formation of colonies in soft agar and on plastic, a hallmark of known tumour suppressors (Wang *et al*, 1993). The 7995 bp open reading frame of TSG16 was found to be too large for efficiently retroviral packaging, and subsequent experiments were aborted. Myc-tagged TSG16 was clearly absent from whole cell lysates and in immuno-stained infected cells (data not shown). This gene delivery system was chosen, as breast cancer cell lines are relatively resistant to classical cationic transfection reagents.

As TSG16 exhibited no expression variability in breast cancer cell lines and only one disease specific mutation, it has been excluded as a frequent cause of breast cancer, but still considered as a candidate for rare forms of the disease. The identification of this large novel gene containing an ANK repeat domain prompted further genomic and protein analysis, in an attempt to elicit TSG16's cellular function.

3.3.4) Functional analyses of the TSG16 protein. The TSG16 ORF of 7995 nucleotides encodes a 2,664 amino acid protein with a predicted molecular weight of 298 kDa (Fig 1a & 1b). Protein sequence analyses identified an ankyrin repeat domain (ANK), a motif that has been identified in over 400 different proteins and function as sites for protein-protein interactions. The ANK containing proteins, BARD1 and Iκβ-R, exhibited the highest homology to TSG16. The TSG16 protein sequence between amino acids 151-307 was 40% identical to amino acids 410-569 of BARD1 and 43% identical to amino acids 362-477 of Iκβ-R. Detailed analyses of this region revealed three tandem ankyrin-like repeats between amino acids 161-276. In addition, the homology between TSG16/BARD1 and TSG16/Iκβ-R extended beyond the conserved amino acids within the ANK motif, and several protein domain prediction programs identified this same TSG16 ANK.

The Profilescan program identified ten bipartite nuclear localization signals (BNLS) between amino acids 457 and 1659 of TSG16 (Fig 1b). In an effort to confirm this predicted sub-cellular localization, the coding region of TSG16 was amplified from human breast RNA and cloned into a mammalian expression vector with a 3' in-frame myc peptide tag. HEK 293T cells were transfected and recombinant proteins exhibited nuclear dot localization, with the number of dots varying from 10 to 50 per nuclei (see section 3.3.9).

Further protein analyses identified six potential PEST sequences, five of which were located at the C-terminus of TSG16 (Fig 1b). PEST sequences function to label proteins for degradation via the ubiquitination pathway (Rechsteiner and Rogers, 1996).

3.3.5) TSG16 homologues. A blastn search of the htgs database identified two BAC clones from the X chromosome (accession numbers AC025083 and AC078956). These assembled genomic clones derived from Xq27.1 were identical to the TSG16 ORF from exon 9 to 13 inclusive. This chromosomal region is flanked by L1 elements suggesting that this DNA has been transposed and represents a true pseudogene. Analyses of this pseudogene revealed 2026 bp of sequence, no start codon and a stop codon followed by non-coding 3'UTR.

The amino acid sequence of TSG16 was used to screen the htgs database using the tblastn algorithm. Two BAC clones were identified (accession numbers AP001033 and AC015955) from chromosome 18 (18p11.3). The identity between TSG16 and the assembled BACs ranged from 30-84% across the length of TSG16. The full-length homologous gene, termed TSG18, was characterized using a similar *in silico* cloning approach to that of TSG16 (Table 6a). RT-PCR experiments from normal breast RNA confirmed the presence of this homologue (Table 6b).

Subsequent comparison of TSG18 to the assembled genomic sequence at 18p11.3 enabled the identified of the genomic structure (Table 6c). Full-length TSG18 consists of 9,035 bp of sequence, 13 exons and spans approximately 147kb of genomic DNA (Fig 6a). TSG18 exhibits an ORF of 6,189 bp encoding a 2062 amino acid protein with a start codon in exon 2 at position 215-217 and a stop codon in exon 11 at position 6401-6403 (Fig 6b).

The transcriptional size of TSG18 was confirmed by northern blot analysis (Fig 2b). A 1450 bp TSG18 cDNA probe, spanning the exon 6/7 boundary, identified two transcripts of approximately 9kb and 7kb, predominantly expressed in the skeletal muscle and placenta. Weaker expression was detected in the remaining tissues. The larger band corresponds to the full-length TSG18 transcript while the origin of the smaller band is unknown, probably resulting from alternative exon splicing or the use of an alternative polyadenylation signal.

A direct comparison between TSG16 and TSG18 indicates a high degree of conservation at both the physical and sequence level between the two proteins across their entire length. Both genes are of similar size (2,664 amino acids and 2,062 amino acids respectively), span large genomic intervals, share an extremely large exon (6,578 nucleotides and 4,721 nucleotides respectively) and have extremely high exon/intron structure conservation. Amino acid sequence alignment between the two genes indicates that both have the same 5' and 3' termini, with the difference in length dictated by the fact that TSG18 has a shorter "large" exon. Sequence identity between TSG16 and TSG18 ranged from 22-67% across their entire lengths (Figure 7) with highest homology occurring between the ankyrin domains present in each gene (74% identity and 87% similarity).

3.3.6) Yeast two-hybrid. A yeast two-hybrid (Y2H) screen was performed to gain insight into TSG16 gene function. Ms Hayley Spendlove performed this screen and below I have presented a summary of the results.

A standard Y2H screen (Display Systems Biotech) was used employing two reporter genes, leucine dependence and GFP (green fluorescence protein) expression. The TSG16 reporter constructs were tested for auto-activation (data not shown). The ANK of TSG16 was used as bait to screen a human breast cDNA library and this 1343 bp fragment identified several members of the PIAS (potent inhibitors of STAT transcription) protein family. Three of the five family members, PIAS1 (40%), PIASxα (15%) and PIAS3 (7%) were the most predominant sequences recovered from the screen (Table 7).

PIAS proteins perform an increasing repertoire of functions including the inhibition of the STAT signalling, regulation of steroid receptor-dependent transcriptional activation, inhibition of apoptosis and more recently, function as ligases in protein sumolyation. Subsequent experiments focused and investigated a possible involvement of TSG16 in STAT mediated transcription. Initial investigations examined the STAT aspect of PIAS protein function, as this is a well-established signalling pathway with central roles in differentiation, proliferation, development, apoptosis and inflammation.

Three different methologies were employed in an effort to confirm this TSG16/PIAS protein interaction: immunoprecipitiation of exogenously expressed proteins, co-localization studies and STAT transcriptional reporter assays. These analyses required the cloning of various expression and reporter constructs that are described in detail in Table 3.

3.3.7) Immunoprecipitation of TSG16 and PIAS1/3. Full length TSG16/pLNCX2/myc was co-transfected with PIAS1/pTarget or PIAS3/GST/pDEST27 into the human embryonic kidney cell line, 293T. 24 hours post-transfection, cells were stimulated with either LIF or IFNα for STAT3 transcriptional activation, or IFNγ for STAT1 for transcriptional activation. Cells were treated for 15 minutes before the media was removed and cells lysed. Proteins were immunoprecipitated with antibodies to their respective epitope tags or c-terminal specific antibodies for PIAS1. Figure 8a shows western blot analysis of cells co-transfected with TSG16 and PIAS1, immunoprecipitated with anti-myc or anti-PIAS1, and blotted with anti-myc. The protein band of approximately 298 kDa represents exogenously expressed TSG16 and confirms it can be immunoprecipitated with myc antibodies. PIAS1 was absent from the TSG16/myc immunoprecipitates (Fig 8a). These same blots were stripped and probed with anti-PIAS1. Anti-PIAS1 immunoprecipitated exogenously expressed PIAS1 but no corresponding TSG16 was detected (Fig 8b). The absence of interaction was not due to the constructs and/or transfection conditions as we could consistently detect the exogenously expressed proteins in whole cell lysates. Similarly, glutathione-sepharose was used to immunoprecipitate PIAS3/GST from co-transfected cells. Glutathione-

sepharose successfully immunoprecipitated the 98 kDa PIAS3/GST recombinant protein but failed to immunoprecipitate any corresponding TSG16 protein (data not shown).

It was possible that TSG16 alters the PIAS/STAT interaction, which itself is cytokine dependent. Therefore immunoprecipitiation experiments were performed in the presence of either IFNγ for STAT1 activation, or LIF for STAT3 activation. These cell lysates were probed with antibodies that recognize either STAT or tyrosine phosphorylated STAT. Both endogenous and exogenous levels of STAT1 were detected and phosphorylated in response to IFNγ (Fig 9b). Similarly, endogenous STAT3 was phosphorylated in response to LIF (Fig 9a). These findings allowed us to eliminate the lack of STAT activation as a possible reason for the absence of interaction.

Additional attempts were made to confirm this interaction by expressing the ANK motif of TSG16 as an alternative to the full-length protein. This eliminated any technical difficulties of working with the 298 kDa full-length protein and furthermore, this motif was originally used in the Y2H screen that identified the PIAS proteins. Subsequent immunoprecipitation experiments established that the ANK motif was stably expressed and can be immunoprecipitated with anti-flag antibodies, but failed to immunoprecipitate PIAS1 (Fig 10). In addition, PIAS1 failed to immunoprecipitate the ANK motif of TSG16 (Fig 11).

The PIAS3 rabbit polyclonal antibodies were suitable for western blot analysis but proved to be ineffective for immunoprecipitation experiments. Glutathione-sepharose successfully immunoprecipitated the PIAS3/GST recombinant protein and the ANK motif was absent from these immunoprecipitates (Fig 12). Similarly, immunoprecipitation of ANK with anti-flag failed to pull-down PIAS3/GST (Fig 13).

Immunoprecipitations can be sensitive to various experimental parameters such as buffer stringency and constituents, and together with the transient nature of some protein interactions, often make them difficult to detect. An extensive effort was made to optimize several experimental parameters including alternative lysis conditions with various salt concentrations (50 mM to 500 mM) (data not shown). To eliminate the possibility that any transient interactions going undetected, extensive cytokine stimulated time course experiments were undertaken. Cytokine induced STAT phosphorylation persisted for up to 1 hour (Fig 9). The development of TSG16 specific antibodies will

enable the examination of endogenous proteins at physiological concentrations, a strategy used to successfully demonstrate the original PIAS/STAT interactions.

3.3.8) STAT transcriptional reporter assays. As an independent indirect methodology to confirm the TSG16/PIAS interaction, gene reporter assays were conducted. It was hypothesized that TSG16 interacts with PIAS1 and/or PIAS3, and alters their STAT inhibitory function.

A dual luciferase reporter system consisting of a STAT3 reporter expressing *firefly* luciferase, and the pRL-CMV reporter expressing *Renilla* luciferase was developed. Transfection efficiencies were normalized to *Renilla* luciferase, whose expression was driven by the strong constitutive CMV promoter. The STAT3 reporter construct pAPRE-luc, possess a luciferase reporter gene containing four copies of APRE (acute phase response elements) and a minimal junB promoter. LIF stimulated 293T cells displayed a 57-fold increase in endogenous STAT3 transcriptional activity (Fig 14). Western blot analysis of these same cell lysates revealed endogenous levels of both STAT3 and phospho-STAT3 (data not shown).

To validate the assay, co-transfection of PIAS3 was performed to replicate the initial PIAS3 dependent inhibition of STAT3 transcription as reported by Chung *et al*, (1997). The PIAS3/pDEST12.2 construct inhibited STAT3/LIF-mediated APRE transcription in a concentration dependent manner (Fig 14). This inhibition of APRE depends on the amount of PIAS3 expression vector transfected and furthermore, was specific for PIAS3 as PIAS1 had no effect. PIAS3 inhibited both endogenous and exogenous STAT3 mediated transcription by up to 80%. All lysates were probed for exogenous protein expression verifying the transfection and stimulations were successful (data not shown).

Co-transfection of the TSG16/myc expression construct resulted in a dose dependent stimulation of the CMV driven *Renilla* luciferase reporter gene (data not shown). This constitutively active reporter was initially included to normalize all treatments for varying transfection efficiencies. This TSG16 induced activation of the CMV promoter driven *Renilla* luciferase gene was not anticipated, and it is conceivable

that TSG16 may alter all CMV driven transcription. The specificity of such responses is still to be established.

These findings invalidated this experimental approach, as the data generated from the STAT3 transcriptional assays was not interpretable in the presence of TSG16. Furthermore TSG16 may induce similar nonspecific effects on other gene promoters. Future investigations should examine the reproducibility and specificity of this TSG16 induced generalized transcriptional activation.

Gene reporter assays were initially devised to incorporate the use of the commercially available STAT1 and STAT3 reporters, pSTAT3-TA-luc and pGAS-TA-luc respectively (BD Biosciences). Both constructs were shown to be useless in gene reporter assays, when compared to data generated by pAPRE-luc (Fig 14), and the relevant literature (Rödel *et al*, 2000; Chung *et al*, 1997).

3.3.9) Co-localization studies. To provide additional evidence for the interaction between the PIAS and TSG16, cellular co-localization studies were performed. HEK 293T cells were co-transfected with combinations of TSG16-myc and PIAS1 or PIAS3. 24 hours later the cells recombinant proteins were detected with myc monoclonal antibodies for TSG16, and epitope specific antibodies for PIAS1 and PIAS3. TSG16 displayed nuclear dot localization, with the number of dots varying from 10 to 50 per nuclei (Fig 15a). Cell nuclei were visualized with DAPI. PIAS3 also exhibited nuclear dot localization, but these dots were smaller and more abundant than the TSG16 nuclear dots (Fig 15b). Double staining experiments with PIAS3/TSG16 revealed that these proteins clearly do not co-localize. PIAS1 co-localization was essentially the same as that of PIAS3 (data not shown). All TSG16/PIAS co-localization experiments were also performed in the presence of cytokine. Under these conditions all protein localization's appeared essentially the same as in the absence of cytokine. Additional co-localization studies were performed with the nuclear speckled PML proteins. PML, sp100 and other proteins constitute nuclear structures called PODs, which are disrupted in acute promyelocytic leukemia. Interestingly, sp100 was recovered from the previously described Y2H screen (3.3.6). The TSG16 nuclear dots are smaller and more abundant than the nuclear POD

proteins (Fig 15c). In summary, exogenously expressed TSG16 does not co-localize with PIAS1, PIAS3 or POD proteins in 293T cells.

3.4 Discussion.

3.4.1) TSG16 cloning, characterization and mutational analysis. The atypical gene structure and size of TSG16 has allowed it to remain uncharacterized and absent from all public EST and non-redundant databases. The large transcriptional size of TSG16 exceeds the limits of conventional reverse transcription and together with the large 263 kb of genomic sequence, both exceed the standard parameters used by many gene prediction programs. In addition, TSG16 exhibits internal poly(A) tracts that permit oligo-dT primer binding in reverse transcription. As a consequence of these limitations, TSG16 was initially represented as an abundance of partial cDNA sequences. The identification and cloning of full-length TSG16 was only possible with the availability of the genomic sequence at 16q24.3 and subsequent RT-PCR analysis.

Initially TSG16 was considered to be a candidate tumour suppressor mapping to the 16q24.3 breast cancer LOH interval. Preliminary analysis identified a putative ankyrin (ANK) domain with 40% homology to the ANK domain of the BRCA1-associated RING domain protein (BARD1), prompting further investigation into a possible link with breast cancer. Mutation analysis in paired breast tumour and normal DNA isolated from the blood lymphocytes of the same patient was undertaken to detect possible cancer specific mutations. One tumour had a mutation at nucleotide 1767, generating the nucleotide change of cytosine to thymidine, and the resulting amino acid change of threonine to methionine. The blood was heterozygous (C/T) and the tumour was homozygous (T). This change lies 64 bp 3' of the ANK and was absent from 100 unrelated normal DNA samples. This tumour-restricted, low frequency missense mutation awaits functional confirmation. Additional amino acid changes were also detected in tumour samples, but these were also present in the normal population and considered polymorphisms.

The screening of non-*BRCA1* and *BRCA2* familial breast cancer patients uncovered a nucleotide change present in the blood of 3 affected individuals from the same family. This family restricted nucleotide change was absent from 100 unrelated

normal DNA samples analysed. From this pedigree tumour DNA was available from one individual with breast cancer and subsequent analysis revealed a heterozygous genotype. Therefore it is unlikely that this mutation at nucleotide 5972 is related to the risk of breast cancer, as it would be expected that the normal allele would be lost in the tumour.

The expression profile of TSG16 has been previously determined in a panel of breast cancer cell lines exhibiting 16q24.3 LOH (Chapter 2, represented as Hs.42390). TSG16 was shown to exhibit a uniform expression profile across the panel of breast cancer cell lines when compared to both normal breast tissue, and the non-tumorgenic breast epithelium cell line, MCF12A. The established tumour suppressor genes *SYK* and *CDKN2A*, both exhibited large variability of expression consistent with the working hypothesis that tumour suppressor genes will show expression variability in breast cancer cell lines. As missense mutations generally do not affect the expression levels of the transcript (Hupp, 1999) the previously described TSG16 missense mutation complies with the uniform expression profile observed in breast cancer cell lines.

Attempts to use a retroviral gene delivery system infect breast cancer cell lines MCF-7 and SKBR-3 were unsuccessful. Colony formation assays for potential TSG16 tumour suppressor function were hampered by the large 7995 bp open reading frame of this gene. Adeno-viral mediated gene delivery has previously been used for large genes, such as *BRCA1*, and could be employed for any future functional analysis.

3.4.2) TSG16 homologues and orthologues. The isolation and characterization of TSG16, and its chromosome 18 homologue TSG18, forms the basis of a new novel protein family. A direct comparison between TSG16 and TSG18 indicates both genes are of similar size (2,664 amino acids and 2,062 amino acids respectively), span large genomic intervals, share an extremely large exon (6,578 nucleotides and 4,721 nucleotides respectively) and have extremely high exon/intron structure conservation. Interestingly, chromosome 18p11.3 has been shown to exhibit significant LOH in sporadic breast tumours (Kittiniyom *et al,* 2001). 68% (27/43) of tumours examined exhibited LOH, 56% of which were early ductal carcinomas, consistent an early event in breast cancer

initiation. The 18p11.3 LOH was clonally inherited in most tumours examined, and subsequent analysis revealed 81% of these same tumours exhibited 16q LOH.

The nucleotide sequence of both TSG16 and TSG18 exhibit significant homology to cDNA clones derived primates and rodents, suggesting a recent evolutionary history with a function restricted to mammals. Interestingly, *BRCA1* also exhibits a similar evolutionary origin. The ANK domain of the mouse orthologue exhibits 91% homology to the ANK domain of TSG16. A putative full-length mouse orthologue can be constructed from overlapping mouse ESTs and predicted sequences based on TSG16 homology.

3.4.3) TSG16 functional analysis. In an attempt to elicit the cellular function of TSG16 a yeast two-hybrid screen was undertaken using the ANK repeat domain as bait. This screen identified three members of the PIAS protein family, PIAS1, PIAS3 and PIASxα (Spendlove, unpublished data). These clones proved to be the most predominant sequences recovered from the screen. The PIAS proteins have a growing repertoire of functions and subsequent analysis focused on a possible involvement of TSG16 in STAT mediated transcription.

The PIAS protein family was originally identified as Potent Inhibitors of Activated STAT (PIAS). STAT proteins are transcription activators that are regulated by classical receptor mediated cell signalling (Reviewed in Aaronson and Horvath 2002). Moreover, ligand binding induces dimerisation of the receptors and activation of JAK kinases. This in turn leads to STAT phosphorylation, formation of STAT1 or STAT3 homodimers, or alternatively STAT1/STAT3 heterodimers. These dimers then translocate to the nucleus where they are specifically recruited to their enhancer element and activate transcription. Negative regulation of this pathway can occur at the JAK kinase level through the binding of either Suppressor Of Cytokine Signalling (SOCS) or CIS proteins. Additional negative regulation occurs at the STAT level through interaction with PIAS (Liao *et al*, 2000). This interaction occurs through N-terminus of PIAS and to date only two of the five PIAS family members have been shown to inhibit STAT transcription, PIAS1 and PIAS3.

PIAS1 and PIAS3 specifically inhibit the transcriptional responses of STAT1 and STAT3 respectively (Liao *et al*, 2000; Chung *et al*, 1997). It was hypothesized that TSG16 interacts with both PIAS1 and PIAS3, and alters their ability to inhibit STAT transcriptional responses. In an effort to validate this hypothesis, 293T cells were selected for analysis as they exhibit high endogenous levels of both STAT3 and STAT1, which undergo phosphorylation and subsequent activation in response to physiological concentrations of LIF and IFNγ, respectively.

Immunoprecipitation experiments with combinations of exogenously expressed TSG16 or the ANK repeat domain of TSG16, with PIAS1 or PIAS3, failed to confirm this physical interaction in 293T cells. An extensive effort was made to optimize various experimental parameters, but with no success. Published data also utilized 293T cells in both STAT1/PIAS1 and STAT3/PIAS3 transcriptional reporter assays, however the accompanying immunoprecipitation from these same cells was not presented (Liao *et al*, 2000; Chung *et al*, 1997; Lui *et al*, 1998). Instead the interaction was demonstrated with either endogenous immunoprecipitation from Daudi cells, or *in vitro* GST immunoprecipitations. Furthermore, the human Daudi lymphoblastoid cell line was not subjected to the corresponding transcriptional reporter assay. The development of TSG16 specific antibodies will enable analysis of the endogenous protein, and subsequent immunoprecipitation will not be limited to cells that are transfection compliant. It is possible that only a small proportion of the exogenously expressed TSG16 and PIAS proteins interact and the assays used were not sufficiently sensitive for their detection.

Gene reporter assays were performed to assess the effect of TSG16 on PIAS mediated inhibition of STAT transcription. These assays were adopted from the original reports showing dose dependent inhibition of STAT1 and STAT3 transcription by PIAS1 and PIAS3 respectively (Liao *et al*, 2000, Chung *et al*, 1997; Lui *et al*, 1998). Comparable levels of cytokine induced endogenous STAT3 activation was achieved, which was dramatically reduced by co-transfection with PIAS3. The finding that TSG16 stimulates the constitutive active CMV promoter from the *Renilla* luciferase reporter, originally included for transfection efficiency normalization, was totally unexpected. Based on this finding the data generated from the STAT reporters could not be interpreted in the presence of TSG16. These observations raise the possibility that TSG16

may also non-specifically affect the promoter of the STAT reporter constructs and furthermore, act as a generalized transcriptional activator.

The zinc finger protein Gfi-1 has recently been shown to perform a similar function to what was originally proposed for TSG16. A yeast two-hybrid screen originally identified PIAS3 as an interacting protein of Gfi-1 and this was subsequently confirmed in HepG2 cells by immunoprecipitation and co-localization. (Rödel *et al*, 2000). Furthermore, Gfi-1 amplified STAT3 transcriptional responses, including IL-6-dependent T-cell activation, by binding and sequestering PIAS3. In summary, the PIAS proteins were identified as potential TSG16 interacting proteins from a yeast two-hybrid screen. It was not possible to confirm the PIAS/TSG16 interaction or demonstrate a functional effect of TSG16 on PIAS mediated STAT inhibition in 293T cells.

3.4.4) STAT protein function. STAT proteins have roles in many different cellular processes including differentiation, proliferation, development, apoptosis and inflammation (Reviewed in Darnell, 1997; Levy and Darnell, 2002). The identification of any new proteins that regulate STAT function will have potential implications for all such processes. Seven mammalian STAT proteins have been identified which display either specific or general functions in cytokine signalling. STAT1, STAT3 and STAT5 are all stimulated by many different cytokines and gene-knock out experiments have begun to elicited their *in vivo* function. STAT1 functions in interferon signalling, as knockout mice show no innate response to viral or bacterial infection (Meraz *et al*, 1996). STAT3 knockouts exhibited embryonic lethality (Darnell, 1997) and a survey of the relevant literature implicates STAT3 in keratinocyte migration, T cell apoptosis, IL-10-mediated signalling in macrophages, and the induction of apoptosis in the involuting breast (Reviewed in Levy and Lee, 2002). STAT5 knockout mice are viable with phenotypic defects that are tissue specific, including defects in mammary gland development and lactation during pregnancy (Liu *et al*, 1997), as well as sexually dimorphic pattern of liver gene expression (Udy *et al*, 1997), infertility and immune dysfunction (Teglund *et al*, 1998).

Frequently abnormal activity of certain STAT family members, particularly STAT1, STAT3 and STAT5, has been associated with a wide variety of human

malignancies, including hematologic, breast, head and neck, and prostrate cancers (for review see Bromberg, 2002). STAT3 has proven oncogenic potential, with constitutively active mutants inducing cellular transformation and subsequent tumours *in vivo* (Turkson and Jove, 2000). Many STAT3 target genes are known, including those encoding the antiapoptotic proteins Bcl-xl, Mcl-1 and Bcl-2, the proliferation-associated proteins Cyclin D1 (CCND1) and MYC, and the proangiogenic factor VEGF (for review see Bromberg, 2002). The STAT signalling pathway is an attractive target for therapeutic intervention, and strategies designed to inhibit STAT activation and STAT mediated gene transcription may play an important role in the next generation of anti-cancer therapies.

3.4.5) PIAS and protein sumoylation PIAS proteins exhibit many different cellular functions including the regulation of steroid receptor-dependent transcriptional activation, inhibition of apoptosis and function as ligases in protein sumolyation.

Sumoylation represents one of the most important post-translational modifications, modulating a wide spectrum of proteins that participate in protein translocation, signal transduction and cell growth. The SUMO modification (sumolyation) pathway is mechanistically similar to that of ubiquitination, but the enzymes of the two processes are distinct. In contrast to ubiquination, sumolyation does not tag proteins for degradation, but seems to enhance their stability or modulate their subcellular compartmentalization (Desterro *et al*, 1998). The number of proteins modified by SUMO is rapidly expanding and currently includes p53, c-Jun, IκBα, Sp100, PML and several nuclear hormone receptors. Mammalian SUMO-1 and its close relatives, SUMO-2 and SUMO-3, are about 18% identical to ubiquitin and are conjugated via an enzymatic cascade that requires a SUMO-specific, heterodimeric E1-activating enzyme (Aos1/Uba2) and a single E2-type conjugating enzyme, Ubc-9. By analogy with the ubiquination process, the PIAS protein family was identified as the E3-ligase in sumolyation. Subsequent experiments have demonstrated a direct involvement of PIAS in the sumolyation of several transcription factors. For example, PIAS proteins can act as specific SUMO ligases in p53 and c-Jun sumoylation and furthermore, PIAS exerts strong repressive effects on p53-dependent transactivation (Schmidt *et al*, 2002). Similarly, PIAS functions as a SUMO-1 ligase in the regulation of androgen,

glucocorticoid, progesterone and estrogen steroid receptor mediated transcription (Kotaja *et al*, 2002). Depending on the cell and promoter context, PIAS family members can either activate or repress transcription.

The previously described yeast two-hybrid screen that originally identified the PIAS proteins also identified Ubc-9, the sumo E2-type conjugating enzyme. In view of this, it is reasonable to hypothesize that TSG16 may also function in protein sumolyation. Furthermore, the sp100 protein was also recovered from the yeast two-hybrid screen, a sumo-regulated nuclear protein that together with the PML protein, form nuclear structures called PODs (Doucas, 2000). PML bodies play central roles in autoimmunity, apoptosis and proteasome-dependent protein degradation (Mattsson *et al*, 2001). The sumolyation of Sp100 and PML results in altered intra-nuclear localization and furthermore, PML bodies undergo disruption to a microgranular form in promyelocytic leukemia (Zhong *et al,* 2000). The PML protein is expressed as an oncogenic fusion protein with the retinoic acid receptor α, and the PML bodies can be reconstituted with retinoic acid treatment. Immuno-fluorescence examination of exogenously expressed PML and TSG16 revealed both were nuclear but do not co-localize in 293T cells.

In summary I have failed to confirm the TSG16/PIAS protein interaction in 293T cells. The development of TSG16 specific antibodies and subsequent analysis of endogenous protein interactions may overcome the limitations of this over-expression system. Future experiments should also aim to confirm the TSG16 mediated transcriptional activation of the CMV promoter.

Figures

Figure 1: Nucleotide (A) and deduced amino acid sequence (B) of TSG16. Start and stop codons are shown (bold and underlined). The first nucleotide of each exon is indicated (bold and underlined). The ankyrin repeat domain is underlined in red, the bipartite nuclear localization signals (BNLS) are underlined in blue and the PEST sequences are underlined in green. The 7795 bp open reading frame encodes a 2664 amino acid protein.

A) TSG16 Nucleotide sequence

```
AGAGGCCGCCCTGAGACGGTGCGCGATGGACCGAGGGCCCCAGCCGGGGA          50
GGCGCCGCCGCCGAGCCCGCGGCCAGACGCCCCATCAGTAGCGTCCGCAC
CGGGAGCCGCGGCTCTCGCCCGAGCCGTGGGCGCGCCCGAGGGGCGGGCT
CGCCTCCCGCCGTCCCTCGCAGCTCTGCCGGGCCCGAGCCCGCGCCGCCG
CCGCCGCCGCCTTGCCGCTCGGGCCGCGCGGCCCGGGAAACGCGGCCGCG          250
GGCTGCATGGGCAGCGCCCGCGCCCCGCCGCTGAGCCGTCGCGGAGCCGC
GCAGCCCTCGGAGCACGAATATATACAGCCCTGCTCTGGGACACACCTCC
ATTGGATTTAAAAGACAGTCCTCGTCAGCACTGACTTTCAGCTATGGAAT
CGCAGACGGTTGATGATGAAGCGCCGGCCGTGTAAATGAAGATCGGGTGA
GGAGCAGGACGATGCCCAAGGGTGGGTGCCCTAAAGCACCACAGCAGGAA          500
GAGCTTCCCCTCAGCAGCGACATGGTGGAGAAGCAGACTGGGAAAAAGGA
TAAAGATAAAGTTTCTCTAACCAAGACCCCAAAACTGGAGCGTGGCGATG
GCGGGAAGGAGGTGAGGGAGCGAGCCAGCAAGCGGAAGCTGCCCTTCACC
GCGGGCGCCAATGGGGAGCAGAAGGACTCGGACACAGAAGAGAAGCAGGG
CCCTGAGCGGAAGAGGATTAAGAAGGAGCCTGTCACCCGGAAGGCCGGGC          750
TGCTGTTTGGCATGGGGCTGTCTGGAATCCGAGCCGGCTACCCCCTCTCC
GAGCGCCAGCAGGTGGCCCTTCTCATGCAGATGACGGCCGAGGAGTCTGC
CAACAGCCCAGTGGACACAACACCAAAGCACCCCTCCCAGTCTACAGTGT
GTCAGAAGGGAACGCCCAACTCTGCCTCAAAAACCAAAGATAAAGTGAAC
AAGAGAAACGAGCGTGGAGAGACCCGCCTGCACCGAGCCGCCATCCGCGG          1000
GGACGCCCGGCGCATCAAAGAGCTCATCAGCGAGGGGGCAGACGTCAACG
TCAAGGACTTCGCAGGCTGGACGGCGCTGCACGAGGCCTGTAACCGGGGC
TACTACGACGTCGCGAAGCAGCTGCTGGCTGCAGGTGCGGAGGTGAACAC
CAAGGGCCTAGATGACGACACGCCTTTGCACGACGCTGCCAACAACGGGC
ACTACAAGGTGGTGAAGCTGCTGCTGCGGTACGGAGGGAACCCGCAGCAG          1250
AGCAACAGGAAAGGCGAGACGCCGCTGAAAGTGGCCAACTCCCCCACGAT
GGTGAACCTCCTGTTAGGCAAAGGCACTTACACTTCCAGCGAGGAGAGCT
CGACGGAGAGCTCAGAAGAGGAAGACGCACCATCCTTCGCACCTTCCAGT
```

TCAGTCGACGGCAACAACACGGACTCCGAGTTCGAAAAAGGCCTCAAGCA
CAAGGCCAAGAACCCAGAGCCACAGAAGGCCACGGCCCCCGTCAAGGACG 1500
AGTATGAGTTTGATGAGGACGACGAGCAGGACAGGGTTCCTCCGGTGGAC
GACAAGCACCTATTGAAAAAGGACTACAGAAAAGAAACGAAATCCAATAG
TTTTATCTCTATACCCAAAATGGAGGTTAAAAGTTACACTAAAAATAACA
CGATTGCACCAAAGAAAGCGTCCCATCGTATCCTGTCAGACACGTCGGAC
GAGGAGGACGCGAGTGTCACCGTGGGGACAGGAGAGAAGCTGAGACTCTC 1750
GGCACATACGATATTGCCTGGTAGTAAGACACGAGAGCCTTCTAATGCCA
AGCAGCAGAAGGAAAAAAATAAAGTGAAAAAGAAGCGAAAGAAAGAAACA
AAAGGCAGAGAGGTTCGCTTCGGAAAGCGGAGCGACAAGTTCTGCTCCTC
GGAGTCGGAGAGCGAGTCCTCAGAGAGTGGGGAGGATGACAGGGACTCTC
TGGGGAGCTCTGGCTGCCTCAAGGGGTCCCCGCTGGTGCTGAAGGACCCC 2000
TCCCTGTTCAGCTCCCTCTCTGCCTCCTCCACCTCGTCTCACGGGAGCTC
TGCCGCCCAGAAGCAGAACCCCAGCCACACAGACCAGCACACCAAGCACT
GGCGGACAGACAATTGGAAAACCATTTCTTCCCCGGCTTGGTCAGAGGTC
AGTTCTTTATCAGACTCCACAAGGACGAGACTGACAAGCGAGTCTGACTA
CTCCTCTGAGGGCTCCAGTGTGGAATCGCTGAAGCCAGTGAGGAAGAGGC 2250
AGGAGCACAGGAAGCGAGCCTCCCTGTCGGAGAAGAAGAGCCCCTTCCTG
TCCAGCGCGGAGGGCGCTGTCCCCAAACTGGACAAGGAGGGGAAAGTTGT
CAAAAAACATAAAACAAAACACAAACACAAAAACAAGGAGAAGGGACAGT
GTTCCATCAGCCAAGAGCTGAAGTTGAAAAGTTTTACTTACGAATATGAG
GACTCCAAGCAGAAGTCAGATAAGGCTATACTGTTAGAGAATGATCTTTC 2500
CACTGAAAACAAGCTAAAAGTGTTAAAGCACGATCGCGACCACTTTAAAA
AAGAAGAGAAACTTAGCAAAATGAAATTAGAAGAAAAAGAATGGCTCTTT
AAAGATGAAAAATCACTGAAGAGAATCAAAGACACAAACAAAGACATCAG
CAGGTCTTTCCGAGAAGAGAAAGACCGTTCGAATAAAGCAGAAAAGGAGA
GATCGCTGAAGGAAAAGTCTCCGAAAGAAGAAAAACTGAGACTGTACAAA 2750
GAGGAGAGAAAGAAGAAATCAAAAGACCGGCCCTCAAAATTAGAGAAGAA
GAATGATTTAAAAGAGGACAAAATTTCAAAAGAGAAGGAGAAGATTTTTA
AAGAAGATAAAGAAAAACTCAAAAAAGAAAAGGTTTATAGGGAAGATTCT
GCTTTTGACGAATATTGTAACAAAAATCAGTTTCTGGAGAATGAAGACAC
CAAATTTAGCCTTTCTGACGATCAGCGAGATCGGTGGTTTTCTGACTTGT 3000
CCGATTCATCCTTTGATTTCAAAGGGGAGGACAGCTGGGACTCGCCAGTG
ACAGACTACAGGGACATGAAGAGCGACTCTGTGGCCAAGCTCATCTTGGA
GACGGTGAAGGAGGACAGCAAGGAGAGGAGGCGGGACAGCCGGGCCCGGG
AGAAGCGAGACTACAGAGAGCCCTTCTTCCGAAAGAAGGACAGGGACTAT
TTGGATAAAAACTCTGAGAAGAGGAAAGAGCAGACTGAAAAGCATAAAAG 3250

TGTCCCTGGCTACCTTTCGGAAAAGGACAAGAAGAGGAGAGAGTCCGCAG
AGGCCGGGCGGGACAGAAAGGACGCCCTGGAGAGCTGCAAGGAGCGCAGG
GACGGCAGGGCCAAGCCCGAGGAGGTGCACCGGGAGGAGCTGAAGGAGTG
TGGCTGCGAGAGTGGCTTCAAGGACAAGTCCGACGGCGACTTTGGGAAGG
GCCTGGAGCCGTGGGAACGGCACCACCCAGCACGAGAGAAGGAGAAGAAG 3500
GATGGCCCCGATAAGGAAAGGAAGGAGAAGACAAAACCAGAAAGATACAA
AGAGAAATCCAGTGACAAGGACAAAAGTGAGAAATCAATCCTGGAAAAAT
GTCAGAAGGACAAAGAATTTGATAAATGTTTTAAAGAGAAAAAAGATACC
AAGGAAAAACATAAAGACACACATGGCAAAGACAAAGAAAGGAAAGCGTC
TCTCGACCAAGGGAAAGAGAAGAAGGAGAAGGCTTTCCCTGGGATCATCT 3750
CAGAAGACTTCTCTGAAAAAAAAGATGACAAGAAAGGCAAAGAGAAAAGC
TGGTACATCGCAGACATCTTCACAGATGAGAGTGAGGACGACAGAGACAG
CTGCATGGGGAGCGGGTTCAAGATGGGAGAGGCCAGCGACTTGCCGAGGA
CGGACGGCCTCCAGGAGAAGGAGGAAGGACGGGAGGCCTATGCCTCCGAC
AGACACAGGAAGTCTTCTGACAAGCAGCACCCTGAGAGGCAGAAGGACAA 4000
GGAGCCCAGAGACAGGAGAAAGGACCGAGGGGCTGCCGACGCGGGGAGAG
ACAAAAAAGAGAAAGTCTTTGAAAAGCACAAGGAGAAGAAGGATAAAGAG
TCCACAGAAAAGTACAAGGACAGGAAGGACAGAGCCTCAGTGGACTCCAC
GCAAGATAAGAAAAATAAACAGAAGCTCCCCGAGAAGGCTGAAAAGAAGC
ACGCTGCCGAAGACAAGGCTAAAAGCAAACACAAAGAGAAGTCGGACAAA 4250
GAACATTCCAAGGAGAGGAAGTCCTCGAGAAGTGCCGACGCGGAAAAAAG
CCTGCTTGAAAAGTTGGAAGAAGAGGCTCTCCATGAGTACAGAGAAGACT
CCAACGATAAAATCAGCGAGGTCTCCTCTGACAGCTTCACGGACCGAGGG
CAGGAGCCGGGGCTGACTGCCTTCCTGGAGGTCTCTTTCACGGAGCCACC
TGGAGACGACAAGCCGAGGGAGAGCGCCTGCCTCCCTGAGAAGCTGAAAG 4500
AGAAGGAGAGGCACAGACACTCCTCATCTTCATCCAAGAAGAGCCACGAC
CGAGAGCGAGCCAAGAAAGAGAAGGCCGAGAAGAAAGAGAAGGGCGAAGA
TTACAAGGAGGGCGGTAGCAGGAAGGACTCCGGCCAGTACGAAAAGGACT
TCCTGGAGGCGGATGCTTACGGAGTTTCTTACAACATGAAAGCTGACATA
GAAGATGAGCTAGATAAAACCATTGAATTGTTTTCTACCGAAAAGAAAGA 4750
TAAAAATGATTCCGAGAGAGAACCTTCCAAGAAAATAGAAAAGGAACTAA
AGCCTTATGGATCTAGTGCCATCAACATCCTAAAAGAGAAGAAGAAGAGA
GAGAAACACAGGGAGAAATGGAGAGACGAGAAGGAGAGGCACCGGGACAG
GCATGCGGATGGGCTGCTGCGGCATCACAGGGACGAGCTCCTGCGGCATC
ACAGGGACGAGCAGAAGCCCGCCACCAGGGACAAGGACAGCCCGCCCCGC 5000
GTGCTCAAAGACAAGTCCAGGGACGAGGGCCCGAGGCTCGGCGATGCCAA
ACTGAAGGAGAAATTCAAGGACGGTGCAGAGAAAGAAAAGGGCGACCCAG

TGAAGATGAGCAACGGGAATGATAAGGTAGCGCCATCCAAAGACCCAGGC
AAGAAAGACGCCAGGCCCAGGGAGAAGCTCCTGGGGGACGGCGACCTGAT
GATGACCAGCTTCGAGAGGATGCTGTCCCAGAAGGACCTGGAGATCGAGG 5250
AGCGCCACAAGCGGCACAAGGAGAGGATGAAGCAAATGGAGAAGCTGAGG
CACCGGTCCGGAGACCCCAAGCTCAAGGAGAAGGCGAAGCCGGCAGACGA
CGGGCGGAAGAAGGGTCTGGACATTCCTGCTAAGAAACCGCCGGGGCTGG
ACCCTCCATTTAAAGACAAAAAGCTCAAAGAGTCGACTCCTATTCCACCT
GCCGCGGAAAATAAGCTACACCCAGCATCAGGTGCAGACTCCAAAGACTG 5500
GCTGGCAGGCCCTCACATGAAAGAGGTCCTGCCTGCGTCCCCCAGGCCTG
ACCAGAGCCGGCCCACTGGCGTGCCCACCCCTACGTCGGTGCTATCCTGC
CCCAGCTACGAGGAGGTGATGCACACGCCCAGGACCCCGTCCTGCAGCGC
CGATGACTACGCGGACCTCGTGTTCGACTGCGCCGACTCGCAGCACTCCA
CGCCCGTGCCCACCGCTCCCACCAGCGCCTGCTCCCCCTCCTTTTTCGAC 5750
AGGTTCTCCGTGGCTTCAAGTGGGCTTTCGGAAAACGCCAGCCAGGCTCC
TGCCAGGCCTCTCTCCACAAACCTTTACCGCTCGGTCTCTGTCGACATTA
GGAGGACCCCCGAGGAAGAATTCAGCGTCGGAGACAAGCTCTTCAGGCAG
CAGAGCGTTCCTGCTGCCTCCAGCTACGACTCTCCCATGCCACCCTCGAT
GGAAGACAGGGCGCCCCTGCCCCCGGTTCCCGCGGAGAAGTTTGCCTGCT 6000
TGTCGCCAGGGTACTACTCCCCAGACTATGGCCTCCCGTCGCCCAAAGTC
GACGCTTTGCACTGCCCACCGGCTGCCGTTGTCACTGTCACCCCGTCTCC
AGAGGGCGTCTTCTCAAGTTTACAAGCAAAACCTTCCCCTTCCCCCAGAG
CCGAGCTGCTGGTTCCTTCCCTCGAAGGGGCCCTTCCCCCGGACCTGGAC
ACCTCCGAGGACCAGCAGGCGACGGCCGCCATCATCCCCCCGGAGCCCAG 6250
CTACCTGGAGCCGCTGGACGAGGGTCCCTTCAGCGCCGTCATCACCGAGG
AGCCCGTTGAGTGGGCCCACCCCTCCGAGCAGGCGCTTGCCTCTAGCCTG
ATCGGGGGCACCTCTGAAAACCCTGTGAGCTGGCCTGTGGGCTCGGACCT
CCTGCTGAAGTCTCCACAGAGATTCCCCGAGTCCCCAAAGCGTTTCTGCC
CCGCGGACCCCCTCCACTCTGCCGCCCCAGGGCCCTTCAGCGCCTCGGAG 6500
GCGCCGTACCCCGCCCCTCCCGCCTCTCCTGCCCCGTACGCTCTGCCCGT
CGCTGAGCCGGGGCTGGAGGACGTCAAGGACGGAGTGGACGCCGTCCCCG
CCGCCATCTCCACCTCAGAGGCGGCTCCCTACGCCCCTCCCTCCGGGCTG
GAGTCCTTCTTCAGCAACTGCAAGTCACTTCCGGAAGCCCCGCTGGACGT
GGCCCCCGAGCCCGCCTGTGTAGCCGCTGTGGCTCAGGTGGAGGCTCTGG 6750
GGCCCCTGGAAAATAGCTTCCTGGACGGCAGCCGCGGCCTGTCTCACCTC
GGCCAGGTGGAGCCGGTGCCCTGGGCGGACGCCTTCGCCGGCCCCGAGGA
CGACCTGGACCTGGGGCCCTTCTCCCTGCCGGAGCTTCCCCTGCAGACTA
AAGATGCCGCAGATGGTGAAGCGGAACCCGTGGAAGAAAGTCTTGCTCCT

CCAGAAGAGATGCCTCCAGGGGCCCCCGGGGTCATAAACGGTGGGGATGT 7000
TTCCACCGTAGTGGCTGAGGAGCCGCCGGCACTGCCTCCTGACCAGGCCT
CCACCCGGCTCCCTGCAGAGCTCGAGCCTGAGCCCTCAGGGGAGCCAAAG
CTGGACGTGGCTCTAGAAGCTGCGGTGGAGGCGGAGACGGTGCCGGAAGA
GAGGGCCCGTGGGGATCCGGACTCCAGCGTGGAGCCCGCGCCCGTTCCCC
CAGAACAGCGCCCACTGGGGAGCGGAGACCAGGGGGCTGAGGCTGAAGGC 7250
CCCCCCGCCGCGTCCCTCTGTGCCCCTGACGGCCCCGCCCCGAACACTGT
GGCACAAGCTCAGGCCGCAGACGGTGCCGGCCCCGAGGACGACACTGAGG
CCTCCCGTGCCGCCGCCCCAGCCGAAGGCCCTCCTGGCGGCATCCAGCCG
GAAGCCGCAGAACCAAAACCCACGGCCGAAGCCCCGAAGGCCCCCCGAGT
GGAGGAGATCCCTCAGCGCATGACCAGGAACCGGGCGCAGATGCTCGCGA 7500
ACCAGAGCAAGCAGGGCCCGCCCCCCTCCGAGAAGGAGTGCGCCCCCACC
CCTGCCCCGGTCACCAGGGCCAAGGCCCGCGGCTCCGAGGACGACGACGC
CCAGGCCCAGCATCCGCGCAAACGCCGCTTTCAGCGCTCCACCCAGCAGC
TGCAGCAGCAGCTGAACACGTCCACGCAGCAGACGCGGGAGGTGATCCAG
CAGACGCTGGCCGCCATCGTGGACGCCATCAAGCTGGATGCCATCGAGCC 7750
CTACCACAGCGACAGGGCCAACCCCTACTTCGAATACCTGCAGATCAGGA
AGAAGATCGAGGAGAAGCGCAAGATCCTGTGCTGTATCACGCCGCAGGCG
CCCCAGTGCTACGCCGAGTACGTCACCTACACGGGCTCCTACCTCCTGGA
CGGCAAGCCGCTCAGCAAGCTCCACATCCCCGTGATCGCACCCCCTCCCT
CCCTGGCGGAGCCCCTGAAGGAGCTGTTCAGGCAGCAGGAGGCCGTCCGG 8000
GGAAAGCTGCGTCTACAGCACAGCATCGAGAGGGAGAAGCTGATCGTATC
CTGTGAGCAGGAGATTCTGCGGGTTCACTGCCGGGCGGCCAGGACCATCG
CCAACCAGGCAGTGCCATTCAGCGCCTGCACGATGCTGCTGGACTCCGAG
GTCTACAACATGCCCCTGGAGAGCCAGGGTGACGAGAACAAGTCAGTGCG
CGACCGTTTCAACGCCCGCCAGTTCATCTCCTGGCTCCAGGACGTGGATG 8250
ACAAGTATGACCGCATGAAGACTTGCCTCCTCATGCGGCAGCAGCACGAG
GCCGCGGCCCTGAACGCCGTGCAGAGGATGGAGTGGCAGCTGAAGGTGCA
GGAACTGGACCCCGCCGGGCACAAGTCCCTGTGCGTGAACGAGGTGCCCT
CCTTCTACGTGCCCATGGTCGACGTCAACGACGACTTTGTATTGTTGCCG
GCATTGACACCGCGGGACGGCCGCAGGACGCAGGCGAGGGCCGCACGGCTG 8500
CCCAGGACTGCTGCTGAGCCCCAGGGGCGGAGGAGGGAGCGCCCTGTCCA
CCCGGGCGGGGAGAGACCCCGAGAGAGACTGCACTTGGCCACAGCGCCTC
CATCCCTTCCAGACTGGTCCAGACGTCGGGGAGGTGAAACACGTCTCTTC
TCTACCAGCTGCCGCGGCGGGGGCAAAGCCCCCAGAGCCTCACTGGCCCC
GGCGGGATGAGGAGACCATGCCGGGCGCGCACACACGGCAGCTTCTGTTT 8750
GAAAACGGCGACCTTCGGGCCCTTTTCTCCAGCAACTGCAGAGGCATTTC

AGGAGTTGGAGAACGGAGTACATTTTTAAAATTTTTTGTTCCATTCTGAT
CAACTAAAGAAAATTAAGGTGTCCACCTCTCACAAACAGTGACCTGCAGG
TCGGAGGGGCAGGAGCCCCATCCCAGCCGTGGTTCTGTTGCCGCCGAGCT
GCGACGGCCTCAGACTGGGCTTCAGACCTGGTGCGGAGCCAGGCGGAGGG 9000
ACACAGAGCCAGGACTCCCAGCCGTATTGAAATGGAGTCAAATCCGCGTG
GTTTGTATCTTCTTTCCTTTAATCTTGGGCATTTTGTTTCTCTTTCTGAG
AACGTAACTTGAAACACTACAGAGCCAATAATCATATAAAAAGTGTATCC
GTAAACGCTGTACAGTTTTATATACCTTCCCAATGTACTTTTGCTGCCTT
TTAAATAATTTATTTTGCAACCCATTACTGCCCAGTTGTAAATAACTTAT 9250
GTACTGTAACATCACTTTCAGTTATGTGTAAAAAAATAAATAAATTAATT
AAAATGC AAAAAAAAAAAAAAAAAAAAA

B) TSG16 amino acid sequence.

MPKGGCPKAPQQEELPLSSDMVEKQTGKKDKDKVSLTKTPKLERGDGGKE 50
VRERASKRKLPFTAGANGEQKDSDTEEKQGPERKRIKKEPVTRKAGLLFG 100
MGLSGIRAGYPLSERQQVALLMQMTAEESANSPVDTTPKHPSQSTVCQKG 150
TPNSASKTKDKVNKRNERGETRLHRAAIRGDARRIKELISEGADVNVKDF 200
AGWTALHEACNRXYYDVAKQLLAAGAEVNTKGLDDDTPLHDAANNGHYKV 250
VKLLLRYGGNPQQSNRKGETPLKVANSPTMVNLLLGKGTYTSSEESSTES 300
SEEEDAPSFAPSSSVDGNNTDSEFEKGLKHKAKNPEPQKATAPVKDEYEF 350
DEDDEQDRVPPVDDKHLLKKDYRKETKSNSFISIPKMEVKSYTKNNTIAP 400
KKASHRILSDTSDEEDASVTVGTGEKLRLSAHTILPGSKTREPSNAKQQK 450
EKNKVKKKRKKETKGREVRFGKRSDKFCSSESESESSESGEDDRDSLGSS 500
GCLKGSPLVLKDPSLFSSLSASSTSSHGSSAAQKQNPSHTDQHTKHWRTD 550
NWKTISSPAWSEVSSLSDSTRTRLTSESDYSSEGSSVESLKPVRKRQEHR 600
KRASLSEKKSPFLSSAEGAVPKLDKEGKVVKKHKTKHKHKNKEKGQCSIS 650
QELKLKSFTYEYEDSKQKSDKAILLEXDLSTENKLKVLKHDRDHFKKEEK 700
LSKMKLEEKEWLFKDEKSLKRIKDTNKDISRSFREEKDRSNKAEKERSLK 750
EKSPKEEKLRLYKEERKKKSKDRPSKLEKKNDLKEDKISKEKEKIFKEDK 800
EKLKKEKVYREDSAFDEYCNKNQFLENEDTKFSLSDDQRDRWFSDLSDSS 850
FDFKGEDSWDSPVTDYRDMKSDSVAKLILETVKEDSKERRRDSRAREKRD 900
YREPFFRKKDRDYLDKNSEKRKEQTEKHKSVPGYLSEKDKKRRESAEAGR 950
DRKDALESCKERRDGRAKPEEVHREELKECGCESGFKDKSDGDFGKGLEP 1000
WERHHPAREKEKKDGPDKERKEKTKPERYKEKSSDKDKSEKSILEKCQKD 1050

KEFDKCFKEKKDTKEKHKDTHGKDKERKASLDQGKEKKEKAFPGIISEDF 1100
SEKKDDKKGKEKSWYIADIFTDESEDDRDSCMGSGFKMGEASDLPRTDGL 1150
QEKEEGREAYASDRHRKSSDKQHPERQKDKEPRDRRKDRGAADAGRDKKE 1200
KVFEKHKEKKDKESTEKYKDRKDRASVDSTQDKKNKQKLPEKAEKKHAAE 1250
DKAKSKHKEKSDKEHSKERKSSRSADAEKSLLEKLEEEALHEYREDSNDK 1300
ISEVSSDSFTDRGQEPGLTAFLEVSFTEPPGDDKPRESACLPEKLKEKER 1350
HRHSSSSSKKSHDRERAKKEKAEKKEKGEDYKEGGSRKDSGQYEKDFLEA 1400
DAYGVSYNMKADIEDELDKTIELFSTEKKDKNDSEREPSKKIEKELKPYG 1450
SSAINILKEKKKREKHREKWRDEKERHRDRHADGLLRHHRDELLRHHRDE 1500
QKPATRDKDSPPRVLKDKSRDEGPRLGDAKLKEKFKDGAEKEKGDPVKMS 1550
NGNDKVAPSKDPGKKDARPREKLLGDGDLMMTSFERMLSQKDLEIEERHK 1600
RHKERMKQMEKLRHRSGDPKLKEKAKPADDGRKKGLDIPAKKPPGLDPPF 1650
KDKKLKESTPIPPAAENKLHPASGADSKDWLAGPHMKEVLPASPRPDQSR 1700
PTGVPTPTSVLSCPSYEEVMHTPRTPSCSADDYADLVFDCADSQHSTPVP 1750
TAPTSACSPSFFDRFSVASSGLSENASQAPARPLSTNLYRSVSVDIRRTP 1800
EEEFSVGDKLFRQQSVPAASSYDSPMPPSMEDRAPLPPVPAEKFACLSPG 1850
YYSPDYGLPSPKVDALHCPPAAVVTVTPSPEGVFSSLQAKPSPSPRAELL 1900
VPSLEGALPPDLDTSEDQQATAAIIPPEPSYLEPLDEGPFSAVITEEPVE 1950
WAHPSEQALASSLIGGTSENPVSWPVGSDLLLKSPQRFPESPKRFCPADP 2000
LHSAAPGPFSASEAPYPAPPASPAPYALPVAEPGLEDVKDGVDAVPAAIS 2050
TSEAAPYAPPSGLESFFSNCKSLPEAPLDVAPEPACVAAVAQVEALGPLE 2100
NSFLDGSRGLSHLGQVEPVPWADAFAGPEDDLDLGPFSLPELPLQTKDAA 2150
DGEAEPVEESLAPPEEMPPGAPGVINGGDVSTVVAEEPPALPPDQASTRL 2200
PAELEPEPSGEPKLDVALEAAVEAETVPEERARGDPDSSVEPAPVPPEQR 2250
PLGSGDQGAEAEGPPAASLCAPDGPAPNTVAQAQAADGAGPEDDTEASRA 2300
AAPAEGPPGGIQPEAAEPKPTAEAPKAPRVEEIPQRMTRNRAQMLANQSK 2350
QGPPPSEKECAPTPAPVTRAKARGSEDDDAQAQHPRKRRFQRSTQQLQQQ 2400
LNTSTQQTREVIQQTLAAIVDAIKLDAIEPYHSDRANPYFEYLQIRKKIE 2450
EKRKILCCITPQAPQCYAEYVTYTGSYLLDGKPLSKLHIPVIAPPPSLAE 2500
PLKELFRQQEAVRGKLRLQHSIEREKLIVSCEQEILRVHCRAARTIANQA 2550
VPFSACTMLLDSEVYNMPLESQGDENKSVRDRFNARQFISWLQDVDDKYD 2600
RMKTCLLMRQQHEAAALNAVQRMEWQLKVQELDPAGHKSLCVNEVPSFYV 2650
PMVDVNDDFVLLPA. 2664

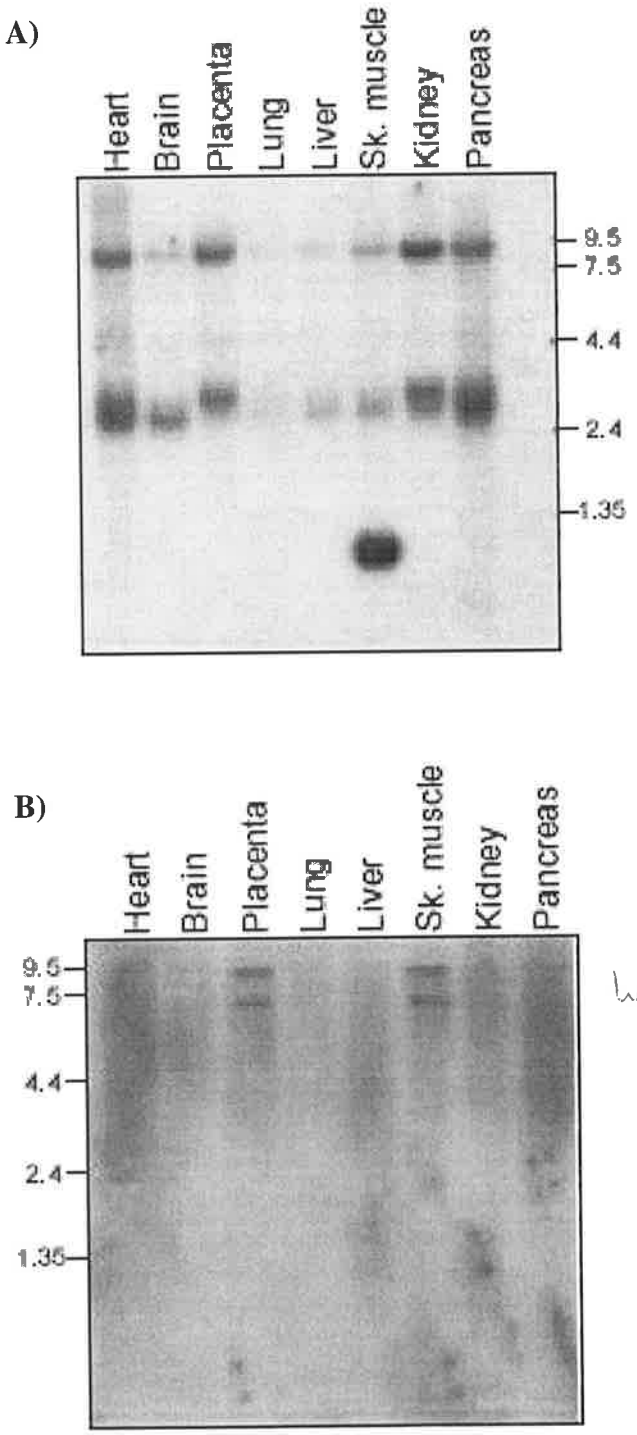


Figure 2: Expression of TSG16 (**A**) and TSG18 (**B**) in normal adult tissues. cDNA probes were hybridized to multiple human tissue northern blots containing poly(A)$^+$ mRNA (2μg each lane) (BD Biosciences). The TSG16 hybridization with a cDNA fragment of 919 bp, encompassing exons 6 to 9, identified a predominant 9 kb transcript expressed across all tissues examined, with abundant expression in the heart, placenta, kidney and pancreas. Smaller transcripts of 3.2, 3.0 and 1.0kb were also detected with a lesser intensity to that of the 9 kb transcript (see text for details). The TSG18 hybridization produced two bands of 7kb and 9 kb, predominantly expressed in the skeletal muscle and placenta

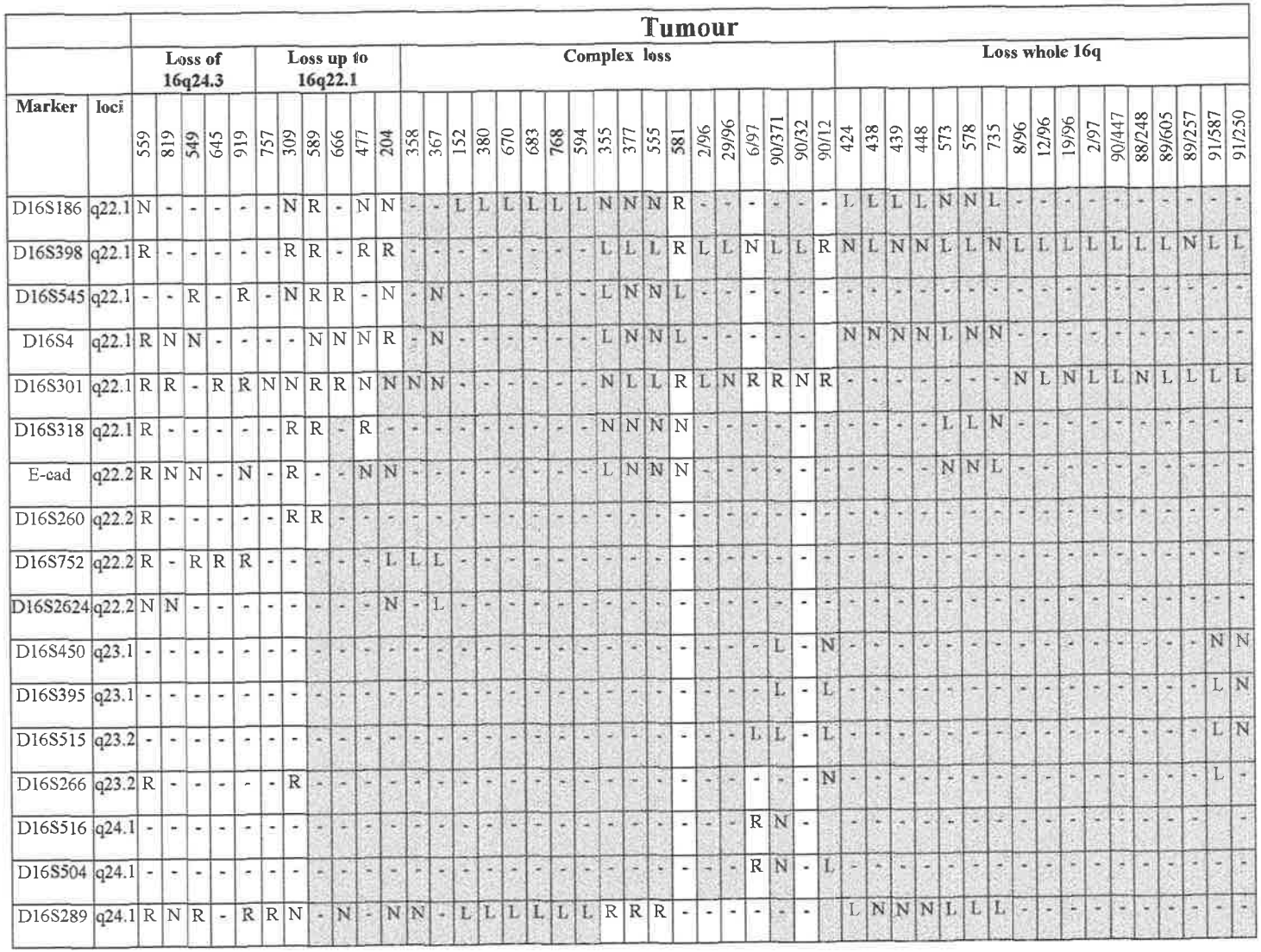


Figure 3: Schematic representation of the 46 tumours used for SSCP analysis of TSG16. Tumour identification numbers are shown at the top of each column (adapted from Clenton-Jansen *et al*, 2001). All tumours show 16q LOH.

Marker	loci	559	819	549	645	919	757	309	589	666	477	204	358	367	152	380	670	683	768	594	355	377	555	581	2/96	29/96	6/97	90/371	90/32	90/12	424	438	439	448	573	578	735	8/96	12/96	19/96	2/97	90/447	88/248	89/605	89/257	91/587	91/250
D16S507	q24.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	R	R	N	L	R	L	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	L	N
D16S511	q24.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	R	L	-	-	-	-	-	-	-	-	-	-	-	-	L	-	-	-	-	N	L
D16S402	q24.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	L	L	-	N	-	-	-	-	-	-	-	-	-	L	L	N	N	L	L	L	L	L	L
D16S449	q24.1	N	R	R	-	R	R	-	L	-	-	L	N	L	-	-	-	-	-	-	R	R	R	-	-	-	-	-	-	-	L	L	L	L	L	L	N	-	-	-	-	-	-	-	-	-	-
D16S539	q24.2	-	-	-	-	-	-	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	L	L	L	L	L	L	-	-	-	-	-	-	-	L	L	L	L	L	L	L	N	L	L
D16S498	q24.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	L	-	-	-	-	-	-	-	-	-	-	-	-	L	L	L	N	N	N	L	L	L	N
D16S3077	q24.3	R	-	R	-	R	-	L	N	L	L	L	N	L	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
D16S3063	q24.3	N	-	R	R	N	N	N	L	N	R	L	L	L	-	-	-	-	-	-	-	N	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
D16S3048	q24.3	-	-	N	-	-	-	L	-	-	-	-	-	-	-	-	-	-	-	-	-	R	R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
D16S305	q24.3	R	N	L	L	L	L	L	N	N	R	L	N	L	-	-	-	-	-	-	R	R	R	R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
D16S1320	q24.3	N	R	L	L	N	N	L	L	L	-	L	L	-	-	-	-	-	-	-	N	L	L	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
D16S413	q24.3	N	-	N	N	L	L	L	-	L	N	L	N	N	-	-	-	-	-	-	L	N	N	N	-	-	-	-	-	-	N	N	N	L	N	N	L	-	-	-	-	-	-	-	-	-	-
D16S2621	q24.3	R	N	N	L	-	-	L	L	N	L	L	L	N	-	-	-	-	-	-	L	L	L	R	L	-	-	-	-	-	-	-	-	-	-	-	-	L	L	L	L	L	N	N	L	L	N
D16S7	q24.3	N	-	L	-	-	-	L	L	N	N	N	-	L	-	-	-	-	-	-	N	L	L	L	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
D16S3023	q24.3	R	-	N	-	-	-	-	-	-	-	-	-	L	-	-	-	-	-	-	L	-	-	L	-	-	-	-	-	-	L	L	L	N	N	L	L	-	-	-	-	-	-	-	-	-	-
APRT	q24.3	R	N	N	-	-	-	N	N	N	N	L	-	N	-	-	-	-	-	-	N	N	N	L	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
D16S3026	q24.3	R	N	N	N	N	N	-	N	N	N	L	-	N	-	-	-	-	-	-	L	N	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
D16S3121	q24.3	N	L	N	L	-	-	L	L	-	N	L	-	N	-	-	-	-	-	-	L	L	L	-	-	-	-	-	-	-	L	L	L	N	L	L	L	-	-	-	-	-	-	-	-	-	-
D16S3407	q24.3	L	N	-	-	N	-	L	N	L	-	-	L	L	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
D16S303	q24.3	L	L	L	L	L	L	N	N	L	L	L	-	N	L	L	L	L	L	L	L	L	L	N	-	L	L	L	N	N	N	L	L	L	L	N	L	L	-	-	N	N	L	N	N	L	N

Figure 4: Pedigree 0019.99.006.

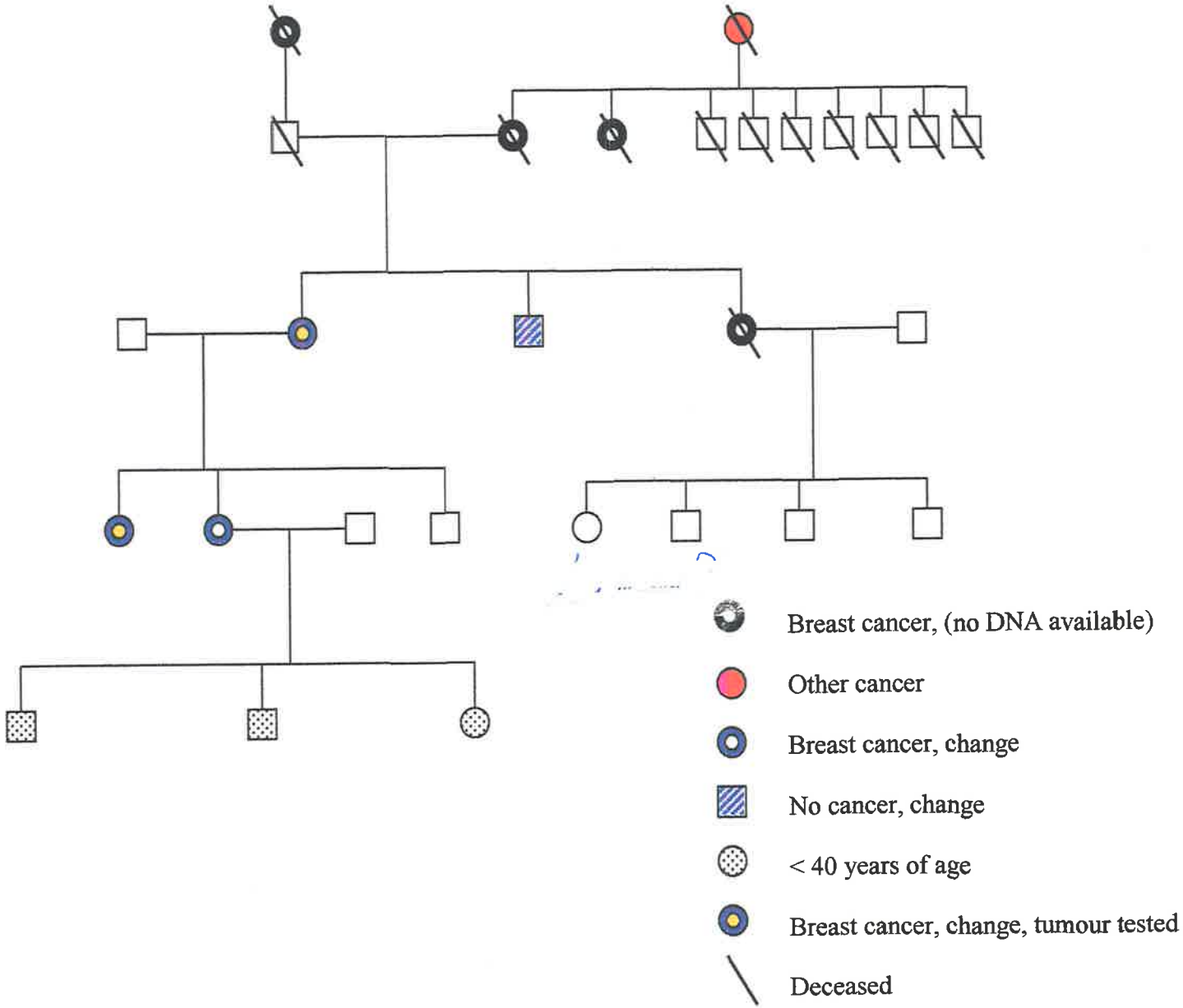
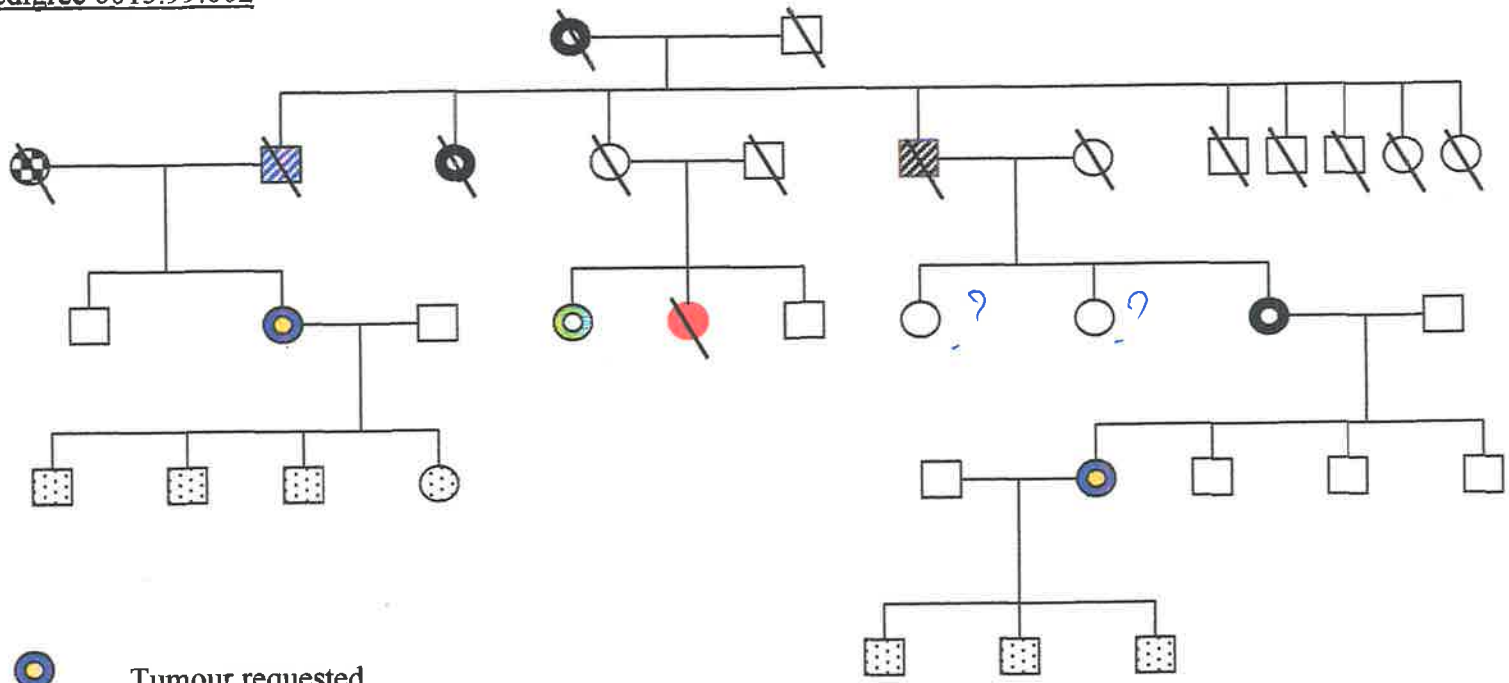


Figure 5: Pedigree 0015.99.002



Tumour requested

No cancer, no change

Breast cancer, (no DNA available)

Breast cancer, change

Breast cancer, no change

Obligate carrier, change

Other cancer

Obligate carrier, (no DNA available)

< 40 years of age

Deceased

Figure 6: Nucleotide (A) and deduced amino acid sequence (B) of TSG18. Start and stop codons are shown (bold and underlined). The first nucleotide of each exon is indicated (bold and underlined). The ankyrin repeat domain is underlined in red, the bipartite nuclear localization signals (BNLS) are underlined in blue and the PEST sequences are underlined in green. The 6186 bp open reading frame encodes a 2062 amino acid protein.

A) TSG18 nucleotide sequence .

GCGAGGAGGCTGTGGTGAGAGACGGACCGAGACCGGAGATGTTCTCAAGC 50

CCGGCTCCGGCGGCTTTACAGGCGGCTGCAGCGGCGACGAAGACAACGAC 100

AGCGACGGCTACGCCGAAGCACTCGTTCCGGGGGGTGAAGCCTCCTGCGCC 150

GGCCTTGCCTCGG**A**TCCAGGATGAGAAGACTGATAAAAGAAGAAGCTAGC 200

TGAACAGCTGTAAA**ATG**CCCAAATCTGGGTTCACAAAACCAATTCAGAGT 250

GAAAATTCTGACAGTGACAGCAATATGGTAGAGAAACCATATGGAAGAAA 300

G**A**GTAAAGACAAGATTGCATCCTACAGCAAAACTCCAAAAATTGAACGAA 350

GTGATGTGAGCAAGGAGATGAAAGAGAAATCATCCATGAAACGTAAACTT 400

CCTTTTACTATTAGCCCATCAAGAAATGAAGAACGAGATTCAGACACAG**A** 450

TTCAGATCCAGGACATACAAGTGAAAATTGGGGGGAGAGACTTATATCTT 500

CTTACAGGACATACTCAG**A**GAAAGAAGGTCCAGAAAAGAAGAAGACAAAA 550

AAGGAAGCTGGAAATAAGAAATCCACACCAGTTAGCATTCTTTTTGGTTA 600

TCCACTCTCTGAGCGAAAACAGATGGCACTTCTTATGCAGATGACAGCAA 650

GAGACAACAGTCCAG**A**TTCCACACCAAATCATCCATCACAAACAACGCCT 700

GCCCAAAAGAAAACTCCCAGTTCTACATCTCGACAGAAAGATAAAGTTAA 750

TAAAAGAAATGAACGTGGTGAAACTCCTTTACACATGGCTGCTATTCGAG 800

GAGATGTGAAACAAGTTAAAGAATTAATAAGTTTAGGGGCAAATGTGAAT 850

GTGAAAGATTTTGCAG**G**TTGGACACCACTGCATGAAGCTTGCAATGTTGG 900

ATATTACGATGTTGCTAAGATACTTATAGCAGCTGGAGCAGATGTTAACA 950

CACAAGGATTAGATGATGACACTCCACTCCATGATTCTGCTAGTAGTGGG 1000

CACAGAGAT**A**TAGTAAAGCTGTTACTTCGTCACGGTGGAAATCCATTTCA 1050

AGCTAATAAACATGGGGAGCGTCCAGTGGATGTAGCAGAAACAGAGGAGT 1100

TGGAGTTGCTACTAAAAAGAGAGGTGCCTTTATCTGATGATGATGAAAGT 1150

TACACAG**A**TTCCGAAGAGGCTCAATCTGTAAATCCTTCTAGTGTTGATGA 1200

AAATATTGACTCTGAAACAGAGAAAGACTCTCTCATCTGTGAAAGTAAAC 1250

AGATACTTCCCAGTAAAACACCTCTTCCATCTGCCCTTGATGAGTATGAG 1300

TTCAAAGATGATGATGATGAAGAAATTAATAAGATGATTGATGATAGGCA 1350

TATTCTTAGGAAAGAACAACGAAAAGAAAATGAACCTGAAGCAGAAAAAA 1400
CTCATTTATTTGCAAAACAGGAGAAAGCCTTCTATCCTAAATCATTTAAA 1450
AGTAAAAAACAAAAGCCATCTAGGGTCTTATATTCAAGTACTGAAAGTTC 1500
TGATGAAGAAGCTCTTCAGAATAAAAAGATTTCTACTTCATGTTCCGTCA 1550
TCCCTGAAACATCAAATTCTGATATGCAAACCAAAAAGGAATATGTAGTT 1600
TCAGGTGAACACAAACAGAAAGGCAAAGTTAAAAGAAAATTGAAAAATCA 1650
GAATAAAAATAAAGAGAACCAAGAGCTAAAGCAAGAAAAGGAAGGAAAAG 1700
AAAATACAAGAATAACAAACTTGACAGTAAATACTGGACTAGATTGTTCA 1750
GAAAAGACCAGAGAGGAGGGGAACTTTAGGAAATCTTTTAGCCCAAAAGA 1800
TGATACTTCATTACATTTATTTCATATTTCCACTGGTAAATCTCCCAAAC 1850
ATTCTTGTGGATTAAGTGAAAAACAGTCAACACCACTAAAACAAGAACAT 1900
ACTAAAACATGTTTATCACCAGGAAGTTCTGAAATGTCATTACAGCCTGA 1950
TCTTGTTCGGTATGATAATACAGAATCTGAATTCTTGCCAGAAAGTTCAA 2000
GTGTAAAATCTTGTAAGCATAAGGAAAAAAGCAAACATCAGAAAGATTTC 2050
CACTTAGAATTTGGTGAAAAATCAAATGCCAAAATAAAGGATGAAGATCA 2100
TAGTCCAACATTTGAAAATTCAGATTGCACACTGAAAAAAATGGATAAAG 2150
AAGGTAAAACATTAAAAAAACATAAATTGAAGCATAAAGAGAGGGAAAAA 2200
GAAAAGCATAAAAAAGAAATTGAAGGTGAAAAGGAAAAATACAAAACTAA 2250
GGATAGTGCCAAAGAACTGCAGAGGAGTGTGGAATTTGATAGAGAATTTT 2300
GGAAAGAGAATTTTTTTAAAAGTGATGAAACTGAAGATCTCTTTTTAAAT 2350
ATGGAACATGAATCCTTAACATTAGAAAAAAAATCAAAATTGGAAAAAAA 2400
CATCAAAGATGATAAATCAACCAAGGAAAAGCATGTGTCAAAAGAGAGGA 2450
ACTTTAAAGAGGAACGAGACAAGATTAAAAAGGAAAGCGAGAAATCTTTT 2500
AGGGAGGAAAAAATAAAAGATCTAAAAGAAGAGAGAGAAAACATACCCAC 2550
AGATAAAGACTCAGAATTTACTTCTTTGGGTATGAGTGCCATTGAGGAAT 2600
CTATAGGGCTTCATTTAGTGGAAAAGGAAATAGACATTGAAAAACAAGAA 2650
AAGCATATAAAGGAAAGTAAAGAAAAACCTGAGAAGCGATCTCAAATTAA 2700
AGAAAAGGACATTGAGAAGATGGAAAGAAAAACCTTTGAAAAAGAAAAGA 2750
AGATAAAACATGAGCATAAGTCAGAAAAAGACAAATTAGATCTTAGTGAA 2800
TGTGTTGATAAAATAAAAGAAAAGGACAAGCTATATTCGCATCACACAGA 2850
AAAATGCCATAAAGAAGGTGAGAAGAGCAAAAATACTGCTGCTATTAAAA 2900
AAACTGACGACAGAGAGAAAAGTAGAGAAAAGATGGATAGGAAACATGAC 2950
AAAGAAAAGCCTGAAAAAGAGAGGCATCTAGCAGAAAGCAAAGAAAAGCA 3000
CTTGATGGAGAAAAAAAATAAACAATCAGATAATAGTGAATACAGTAAAT 3050
CAGAAAAAGGCAAAAATAAAGAAAAAGACAGGGAGCTAGATAAAAAGGAA 3100
AAATCTAGAGATAAAGAAAGTATAAATATAACTAACTCCAAACACATACA 3150
GGAAGAAAAAAAAATCAAGTATAGTAGACGGTAATAAAGCACAACATGAAA 3200
AACCCTTATCCCTTAAAGAAAAAACAAAAGATGAACCTTTGAAAACTCCA 3250

GATGGAAAAGAAAAAGATAAAAAAGATAAAGATATAGATAGATACAAAGA 3300
ACGAGACAAACATAAAGATAAAATTCAAATAAATAGCTTACTCAAACTAA 3350
AATCTGAAGCAGATAAGCCTAAACCTAAGTCATCACCAGCATCAAAAGAT 3400
ACCCGACCTAAAGAAAAGAGGTTAGTGAATGATGATTTAATGCAGACAAG 3450
TTTTGAACGAATGCTAAGCCTTAAAGACCTAGAAATAGAACAGTGGCACA 3500
AAAAACATAAGGAAAAAATTAAGCAAAAAGAAAAGGAACGGTTGAGAAAC 3550
CGAAACTGTTTAGAACTTAAAATAAAAGATAAAGAAAAAACAAAGCATAC 3600
ACCAACTGAATCCAAAAATAAAGAACTTACTAGGTCAAAGAGTTCAGAAG 3650
TGACTGATGCATATACCAAGGAGAAACAACCTAAAGATGCTGTAAGTAAC 3700
AGATCACAATCTGTTGACACCAAAAATGTAATGACTTTAGGGAAGTCATC 3750
TTTTGTTTCAGATAATAGCTTAAACAGGTCTCCTAGATCAGAAAATGAAA 3800
AGCCGGGTCTCAGCTCCAGATCTGTATCCATGATTTCTGTTGCTAGTTCA 3850
GAAGATTCCTGCCATACTACAGTGACAACCCCAAGGCCTCCAGTTGAGTA 3900
TGACTCTGACTTTATGTTAGAGAGTTCAGAATCCCAAATGTCCTTTTCCC 3950
AGTCACCTTTTTTGTCAATTGCCAAATCTCCTGCTCTTCATGAAAGGGAA 4000
TTGGACAGCCTGGCTGACTTGCCGGAGCGGATTAAACCACCATATGCAAA 4050
CAGACTTTCAACATCCCATCTTAGGTCATCTTCTGTAGAAGATGTTAAAC 4100
TAATTATAAGCGAGGGGAGACCTACCATAGAAGTTCGAAGATGTAGCATG 4150
CCTTCTGTCATTTGTGAACATACCAAACAATTCCAAACAATATCAGAAGA 4200
GAGCAATCAAGGTAGCTTATTAACTGTGCCAGGAGATACTAGTCCTTCTC 4250
CCAAACCTGAGGTATTCTCAAATGTGCCTGAAAGAGACCTTTCAAATGTA 4300
TCTAACATACATTCCAGTTTTGCAACTTCTCCAACTGGAGCTTCAAACAG 4350
CAAGTATGTTTCAGCTGATAGAAATCTCATCAAGAATACTGCCCCAGTGA 4400
ACACTGTAATGGACAGTCCAGTGCATTTAGAGCCATCTAGTCAGGTTGGT 4450
GTGATCCAGAATAAATCATGGGAGATGCCTGTTGATAGACTAGAGACATT 4500
AAGCACCAGAGACTTTATCTGCCCAAATTCTAACATACCTGATCAAGAAT 4550
CCTCTCTTCAGAGTTTTTGTAATTCTGAAAATAAGGTATTGAAAGAAAAT 4600
GCTGATTTTTTATCCCTGCGCCAGACTGAACTGCCAGGAAACTCTTGTGC 4650
TCAGGATCCGGCATCCTTTATGCCTCCACAGCAGCCTTGCTCTTTCCCCA 4700
GCCAATCACTTTCAGATGCTGAATCGATTTCTAAACATATGTCTTTGTCA 4750
TATGTTGCTAATCAAGAGCCAGGTATTTTACAACAAAAAAATGCAGTTCA 4800
GATTATTAGTTCTGCTTTAGATACTGATAATGAATCTACAAAAGATACAG 4850
AAAATACTTTTGTCCTAGGAGATGTTCAAAAAACAGATGCCTTTGTCCCA 4900
GTGTACTCTGACAGCACTATTCAAGAAGCATCACCAAACTTTGAGAAAGC 4950
TTATACTTTACCTGTGTTACCATCAGAAAAGGACTTTAATGGAAGTGATG 5000
CCTCTACCCAGCTAAATACACATTATGCATTTAGCAAACTAACTTACAAG 5050
TCTTCCAGTGGCCATGAAGTTGAGAATAGCACAACTGATACTCAGGTCAT 5100
TTCACATGAAAAAGAAAACAAACTGGAGAGTTTGGTTTTAACTCATTTGA 5150

GTAGGTGTGATTCTGATTTATGTGAAATGAATGCAGGGATGCCAAAAGGA 5200
AACCTAAATGAACAAGATCCAAAACATTGTCCTGAAAGTGAAAAGTGTTT 5250
GCTTTCCATAGAAGATGAGGAATCTCAACAAAGCATTTTATCAAGTCTGG 5300
AAAACCATTCACAGCAGTCAACTCAACCAGAAATGCATAAATATGGTCAG 5350
TTAGTTAAAGTAGAATTAGAAGAAAATGCCGAAGATGATAAAACTGAAAA 5400
CCAAATCCCTCAAAGAATGACTAGAAACAAAGCAAATACAATGGCAAATC 5450
AAAGCAAACAGATTCTTGCTAGCTGTACACTATTATCAGAAAAAGACAGT 5500
GAATCCTCATCTCCTAGAGGAAGAATAAGATTAACTGAAGATGACGATCC 5550
TCAAATTCACCATCCACGGAAAAGGAAAGTGTCACGTGTACCTCAGCCTG 5600
TGCAAGTGAGTCCCTCTTTACTACAAGCAAAAGAGAAAACTCAGCAATCT 5650
CTGGCAGCCATTGTAGATTCTCTAAAACTAGATGAGATTCAGCCATACAG 5700
TTCAGAGAGAGCAAATCCATATTTTGAATACTTGCACATAAGGAAAAAAA 5750
TAGAAGAAAAACGCAAATTACTGTGTAGTGTGATTCCTCAAGCACCTCAG 5800
TACTATGACGAATATGTAACATTTAACGGATCATATCTCCTGGATGGAAA 5850
CCCCTTAAGCAAGATTTGTATTCCCACA**A**TTACACCACCACCTTCACTGT 5900
CAGATCCACTTAAAGAGCTTTTTCGACAACAGGAAGTTGTAAGGATGAAA 5950
CTACGTTTGCAACACAGTATTGAAAGG**G**AAAAACTCATTGTATCCAACGA 6000
ACAAGAAGTTTTGCGAGTTCATTACAGAGCTGCAAGAACATTGGCAAATC 6050
AAACACTGCCATTTAGTGCTTGTACTGTTTTGCTGGATGCCGAAGTATAC 6100
AATGTACCATTGGACTCTCAG**T**CTGATGACAGTAAAACTTCTGTGAGGGA 6150
TCGCTTTAATGCAAGACAATTCATGTCTTGGTTACAAGATGTGGATGATA 6200
AATTTGACAAATTAAAG**A**CCTGTCTTTTAATGAGGCAACAACATGAAGCT 6250
GCGGCTTTAAATGCTGTCCAGAGGTTAGAATGGCAGCTCAAACTCCAGGA 6300
ACTTGATCCTGCCACCTATAAATCTATCAGCATTTACGAAATCCAGGAGT 6350
TTTATGTTCCCCTTGTTGATGTTAACGACGACTTTGAATTGACTCCTATA 6400
TAGCAGTCAGTACTTCCTGATGGTATTGTCCTAAACTGGTGATGCTCAAG 6450
CATTATACTGTGGAATACTGCCTTTTGACAAAAATACTCATGCCTTTACA 6500
ATTGTTAGTAAAGTTCGATTATAGTTGGTTATGTAGTAAACACTGTCATT 6550
TTATAAAAAATGAGAATTATTTTGGATCTTAGATCCAAACACAGTTTCTA 6600
ATAGAAAACTATTATTTATATTGGGAAAGGTAACTATTGCATTAGAGCAT 6650
GTTGGCAGACTGGTAGGTATTTAAAAAGTTGAGAATCTGCTAACAGCGCT 6700
GGAAGTTGTTAGCGCTCTAAGTAATAAGATAACCACTAGTATTCAAATCT 6750
CTTTCAGGTTTTATTAAAAAATATATATCAATAAACTAAAAGGTTCAATT 6800
CCTACCAAATAGTTTCTAATGTGGGAGAAAAACTTGGCACAAAATTTCTT 6850
CAGTTTATTATCTGTAAATTGTACAGTTTTCTTTTTGAAAGTTTTAATAT 6900
TGTCTTCCTTTTTAATAACTTATTTTATACATATTGTGCAGATGTAAATC 6950
TTGTAATTAATGGTCAAACTGTATAAAGGGATTGGTAGTCAAAACATGTA 7000

```
CAAAGAAATACCTGTAAAACTGTTTTGTCTCATGTTTTATTGGACCAAAG     7050
TTGTGGTTTGTATGGAGTGTAGTAGTAGTGTGTACAGGTAGAAAACTTTT     7100
AAATACAGCATGCAGGTGTTTCAGTTAGCTTGTTTTCATCACCATAACTG     7150
CAAAGATGTGGCTTAGTTGTATTGCATGCTTCCTATAATTTAACTCTCCA     7200
TAATTGATGCCTGCAGTAGTGTAAGGCATTTCATACTAGTCTCCTCTAGT     7250
AGACCTGTGACTTACTGTGTTGGACATATTATTTAGACTTAGTCATACAA     7300
AGAAACTTAGCTCTTTTTTCATCTCACAGTAAAGCCTATTTCCCCAGGAA     7350
AAAAATAAATGCCTTTGAATGAAAATTCTGAAATTGTAAATGTCTATTTT     7400
AATATTCACCTATGAAAGAATCTGTGAATATATGTAAATACGTTTAATAA     7450
ATTTTATTGGTCATGTTAAATCATTGTAAAACTTTTTTACATTGCTTAAT     7500
GTTTTAAGCTTAATAGCCTTTGCACTTTTAAAATAAAAACCAAGTATGCA     7550
AATCAAAGATATTTGGTAGTCAAAATAAGTAAAAGAAATATAGGAATATT     7600
CCAGTCAATCTCTGAAATGTTTATGAAAAACAGGTTAATATGTTGTATTT     7650
TTTTCCTTGTATCAAGATGCAAAACATAATTTGCAAAATTTTATAATTGA     7700
AATAAAACTTGGTATGCTGTTTTATAATAAATTAGCAATATACTTTAAAA     7750
AAAATCCAGTTTCTCCTATAACATGATGTAAATTAAATATTCAGGAATAT     7800
TTCAGGTCTGAAATCTGCTGCTGAACTTGTTAGACATTTTTGAAAGGAAA     7850
ATTAGGTTAGCGTACAATTTATCATAAATATATGAGGTAAAATTTGCAAA     7900
ACTTTCCACAGTACTTTCTTGAATTATAATACTGAATGTTTTCTGTTCCG     7950
ATAGAGAAAAGTGAAGCAAATATGTGAGGAATGAAGTGATCTAGTGCAAG     8000
TTGGTAATTCCAGCTAATGGGAGAGAATGTAAATGTGCAAAGGTACATAT     8050
TAAGCTTAGGGATTTTTGAATTTTTACGTCAGTTTATATTTTAATTCTTA     8100
CTGTACTTGGCATGCGCAATAAAGCAGCACATGTAAAAAAAAATACACAG     8150
TGCAAGGACTTTATTATAAAGGTTGGATGTAGTTTTCCTTAGAAATTTAG     8200
GTGAGAGATTTTGGCCTTTTTTATTGGATAAATTGGGCCAGAAAGGCAGG     8250
GTAGATTTCGAAGCACTGGTTCTGTTGAAGTTAAGTTTATTAAGCCTGGG     8300
AACATTAAAAGCTAATTTATAAAAGCAATACTTTTTAATATGAAAACTTA     8350
CTGCAAAGTTTGTTTATACTTTTGCCTAAAAAGGAAATTGGATGGGATAC     8400
TGTGGCAAATCATAAAAAACCAGATAATTGAACTTTGAAGTTATAGAAAA     8450
TCAGAGAGGGGTAAGTTTATAGGGCATTTTGTTCTGATGGTTCAACCAGA     8500
GGTCTGGGAAATAGCACTGTTGGCCCAAACAGAACAGGCTTTTAGAAGAT     8550
AAAAGCGACAAGAAGGAATCTGGTGAATTTTAGTCATCCCAGCTTTTTAG     8600
TCTTAACCACAGTTCTCACTCTCTTAAATGGTACCTCAAAAAGCTGGAGC     8650
CTCTCTGCCATGATTATGCTTCTACAAATTTCTTTTATAAAGAGACTCAA     8700
AGCTAATGATAGCTTAAAAGAAAAGTTAATGCCTTCTCATTGGAAATGTA     8750
TAATCAAATAAGTAGTTAAGGGCTTTTGGTATTAAAGATATTCTGAAGCT     8800
CTGAAATGCTAGAAAAAAATTTGGAATGGAGTATATGCCTGAAAAGGTTT     8850
TGGATTCAGAAAGAAAAAGGATGGTTAGTTTAATCAGTGATTCTTTTTAA     8900
```

```
ACTCTTCAAATATCATGAACAAGATACTAAATTGTACCTAAGGATTTGTA      8950
TTTCTTTACAATTTGTTCTAAATATCTGTTTAATGACTAGTTGATATTTG      9000
TGCATGTTATTTAATAAAGAGTTATATTTTTATAG                     9035
```

B) TSG18 amino acid sequence.

```
MPKSGFTKPIQSENSDSDSNMVEKPYGRKSKDKIASYSKTPKIERSDVSK      50
EMKEKSSMKRKLPFTISPSRNEERDSDTDSDPGHTSENWGERLISSYRTY      100
SEKEGPEKKKTKKEAGNKKSTPVSILFGYPLSERKQMALLMQMTARDNSP      150
DSTPNHPSQTTPAQKKTPSSTSRQKDKVNKRNERGETPLHMAAIRGDVKQ      200
VKELISLGANVNVKDFAGWTPLHEACNVGYYDVAKILIAAGADVNTQGLD      250
DDTPLHDSASSGHRDIVKLLLRHGGNPFQANKHGERPVDVAETEELELLL      300
KREVPLSDDDESYTDSEEAQSVNPSSVDENIDSETEKDSLICESKQILPS      350
KTPLPSALDEYEFKDDDDEEINKMIDDRHILRKEQRKENEPEAEKTHLFA      400
KQEKAFYPKSFKSKKQKPSRVLYSSTESSDEEALQNKKISTSCSVIPETS      450
NSDMQTKKEYVVSGEHKQKGKVKRKLKNQNKNKENQELKQEKEGKENTRI      500
TNLTVNTGLDCSEKTREEGNFRKSFSPKDDTSLHLFHISTGKSPKHSCGL      550
SEKQSTPLKQEHTKTCLSPGSSEMSLQPDLVRYDNTESEFLPESSSVKSC      600
KHKEKSKHQKDFHLEFGEKSNAKIKDEDHSPTFENSDCTLKKMDKEGKTL      650
KKHKLKHKEREKEKHKKEIEGEKEKYKTKDSAKELQRSVEFDREFWKENF      700
FKSDETEDLFLNMEHESLTLEKKSKLEKNIKDDKSTKEKHVSKERNFKEE      750
RDKIKKESEKSFREEKIKDLKEERENIPTDKDSEFTSLGMSAIEESIGLH      800
LVEKEIDIEKQEKHIKESKEKPEKRSQIKEKDIEKMERKTFEKEKKIKHE      850
HKSEKDKLDLSECVDKIKEKDKLYSHHTEKCHKEGEKSKNTAAIKKTDDR      900
EKSREKMDRKHDKEKPEKERHLAESKEKHLMEKKNKQSDNSEYSKSEKGK      950
NKEKDRELDKKEKSRDKESINITNSKHIQEEKKSSIVDGNKAQHEKPLSL      1000
KEKTKDEPLKTPDGKEKDKKDKDIDRYKERDKHKDKIQINSLLKLKSEAD      1050
KPKPKSSPASKDTRPKEKRLVNDDLMQTSFERMLSLKDLEIEQWHKKHKE      1100
KIKQKEKERLRNRNCLELKIKDKEKTKHTPTESKNKELTRSKSSEVTDAY      1150
TKEKQPKDAVSNRSQSVDTKNVMTLGKSSFVSDNSLNRSPRSENEKPGLS      1200
SRSVSMISVASSEDSCHTTVTTPRPPVEYDSDFMLESSESQMSFSQSPFL      1250
SIAKSPALHERELDSLADLPERIKPPYANRLSTSHLRSSSVEDVKLIISE      1300
GRPTIEVRRCSMPSVICEHTKQFQTISEESNQGSLLTVPGDTSPSPKPEV      1350
FSNVPERDLSNVSNIHSSFATSPTGASNSKYVSADRNLIKNTAPVNTVMD      1400
SPVHLEPSSQVGVIQNKSWEMPVDRLETLSTRDFICPNSNIPDQESSLQS      1450
FCNSENKVLKENADFLSLRQTELPGNSCAQDPASFMPPQQPCSFPSQSLS      1500
DAESISKHMSLSYVANQEPGILQQKNAVQIISSALDTDNESTKDTENTFV      1550
```

```
LGDVQKTDAFVPVYSDSTIQEASPNFEKAYTLPVLPSEKDFNGSDASTQL      1600
NTHYAFSKLTYKSSSGHEVENSTTDTQVISHEKENKLESLVLTHLSRCDS      1650
DLCEMNAGMPKGNLNEQDPKHCPESEKCLLSIEDEESQQSILSSLENHSQ      1700
QSTQPEMHKYGQLVKVELEENAEDDKTENQIPQRMTRNKANTMANQSKQI      1750
LASCTLLSEKDSESSSPRGRIRLTEDDDPQIHHPRKRKVSRVPQPVQVSP      1800
SLLQAKEKTQQSLAAIVDSLKLDEIQPYSSERANPYFEYLHIRKKIEEKR      1850
KLLCSVIPQAPQYYDEYVTFNGSYLLDGNPLSKICIPTITPPPSLSDPLK      1900
ELFRQQEVVRMKLRLQHSIEREKLIVSNEQEVLRVHYRAARTLANQTLPF      1950
SACTVLLDAEVYNVPLDSQSDDSKTSVRDRFNARQFMSWLQDVDDKFDKL      2000
KTCLLMRQQHEAAALNAVQRLEWQLKLQELDPATYKSISIYEIQEFYVPL      2050
VDVNDDFELTPI.                                            2062
```

Figure 7: A comparison between N-terminal amino acid sequence of TSG16 and TSG18. Highest homology was detected across the ANK motif exhibiting 74% identity and 87% similarity (***bold/italic***). Identical amino acids are boxed. The identity ranged from 22% to 67% across the entire protein lengths. The ANK consensus amino acid sequence is shown in red.

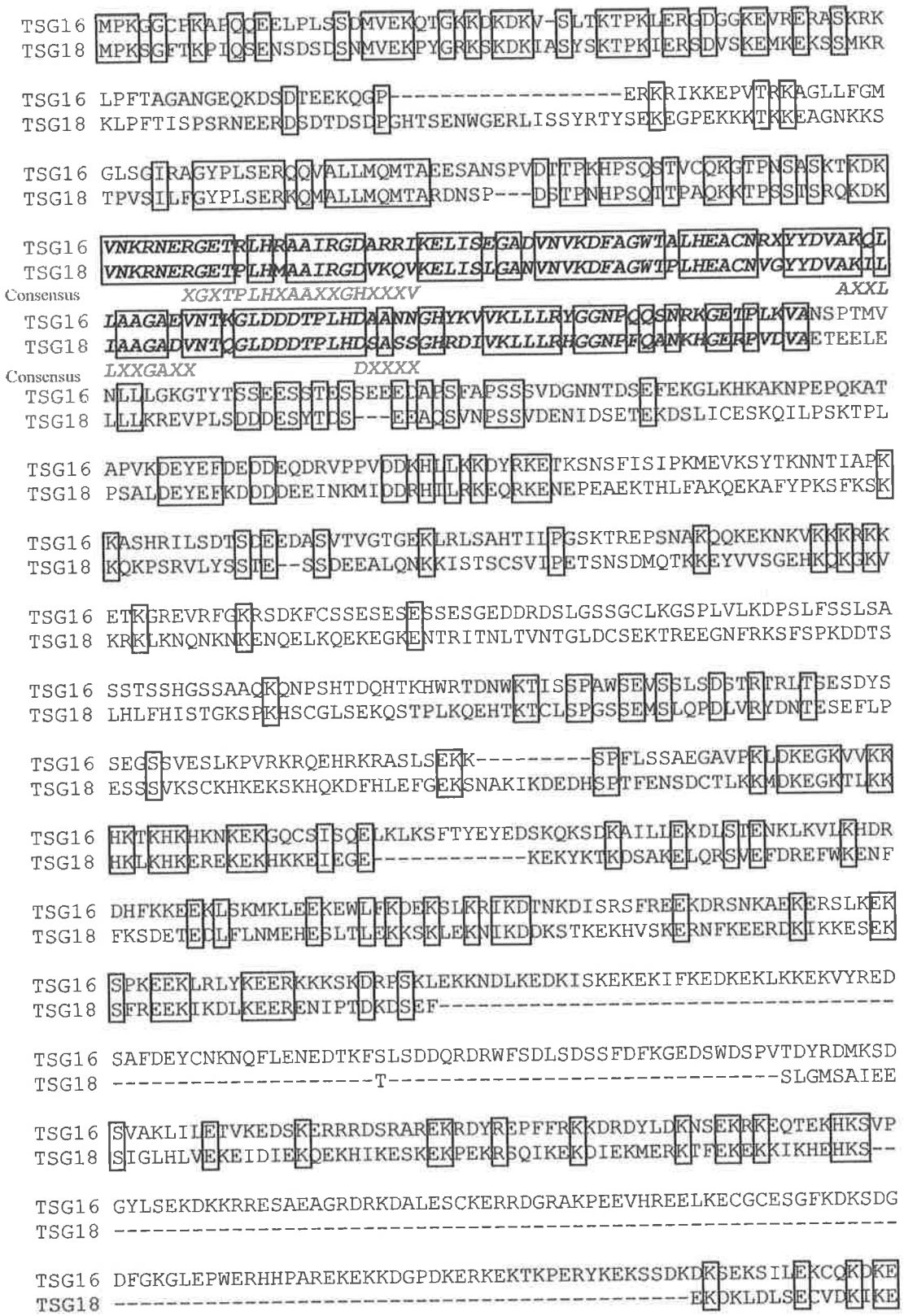


Figure 8: Anti-PIAS1 fails to immunoprecipitate TSG16, and anti-myc fails to immunoprecipitate PIAS1. 293T cells were transfected with TSG16/myc/pLNCX2 and/or PIAS1/pTarget. 24 hours post transfection cells were treated with INFγ and lysed. Recombinant proteins from total cell lysates (500 μg) were immunoprecipitated with either anti-myc (2 μg) or anti-PIAS1 (10 μg). Immunoprecipitates were resolved on NuPAGE gels, blotted and probed with either anti-myc (A) or anti-PIAS1 (B). Anti-PIAS1 and anti-myc immunoprecipitated exogenously expressed PIAS1 and TSG16 respectively, however no corresponding interacting TSG16 or PIAS1 were detected.

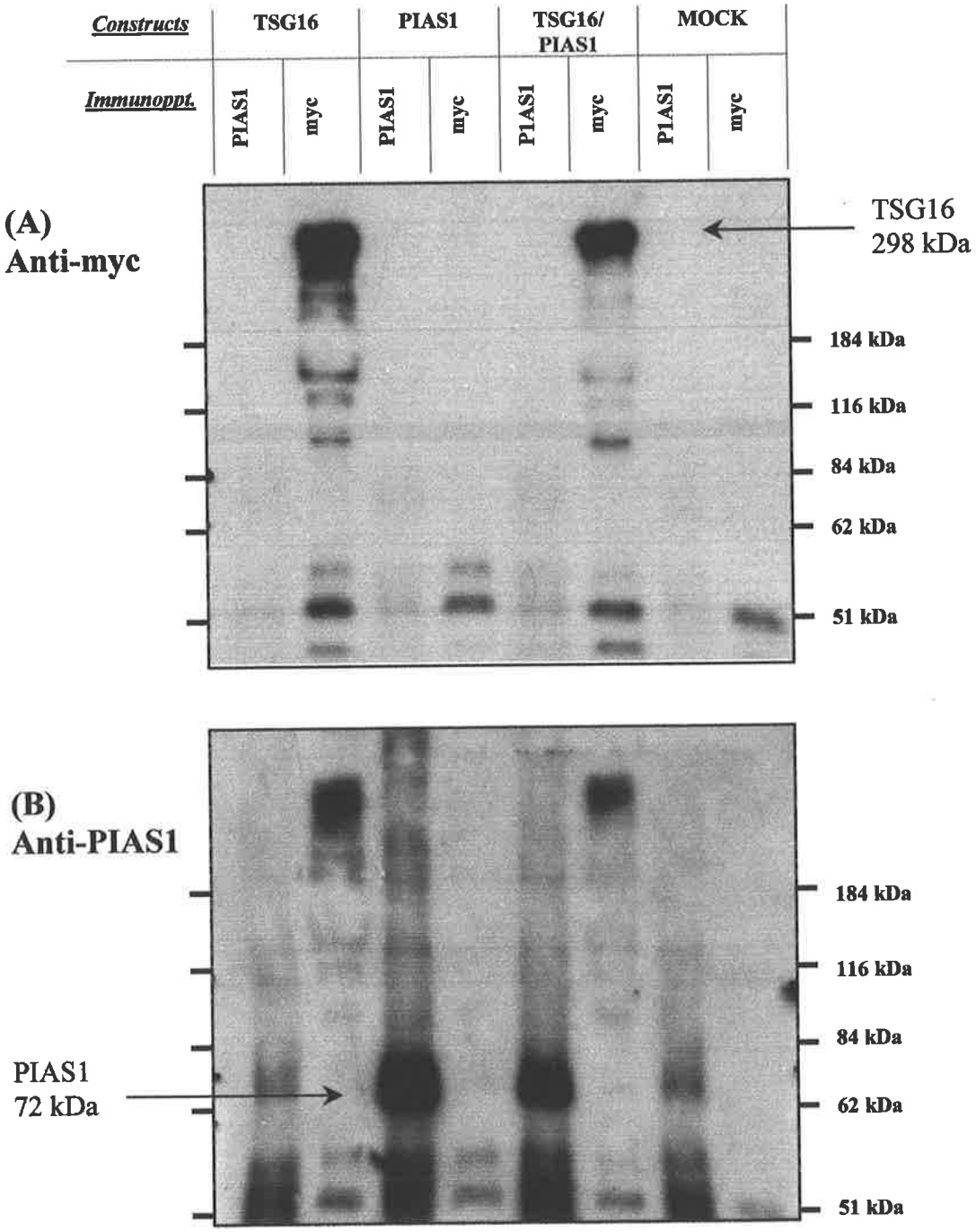


Figure 9: Cytokine stimulated STAT phosphorylation. (**A**) Time course analyses of LIF stimulated phosphorylation of endogenous STAT3. 293T cells were stimulated (minutes), lysed, and 20 µg of total protein was resolved on NuPAGE gels, blotted and probed with either anti-phospho-STAT3 or anti-STAT3. LIF stimulated STAT3 phosphorylation was maintained for at least 60 minutes. (B&C) INFγ stimulated both exogenous and endogenous STAT1 phosphorylation. 293T cells were transfected with various amounts of STAT3/pcDNA3.1 or mock vector, stimulated for 15 minutes and lysed. 20 µg of total protein was resolved on NuPAGE gels, blotted and probed with either anti-phospho-STAT1 (B) or anti-STAT1 (C).

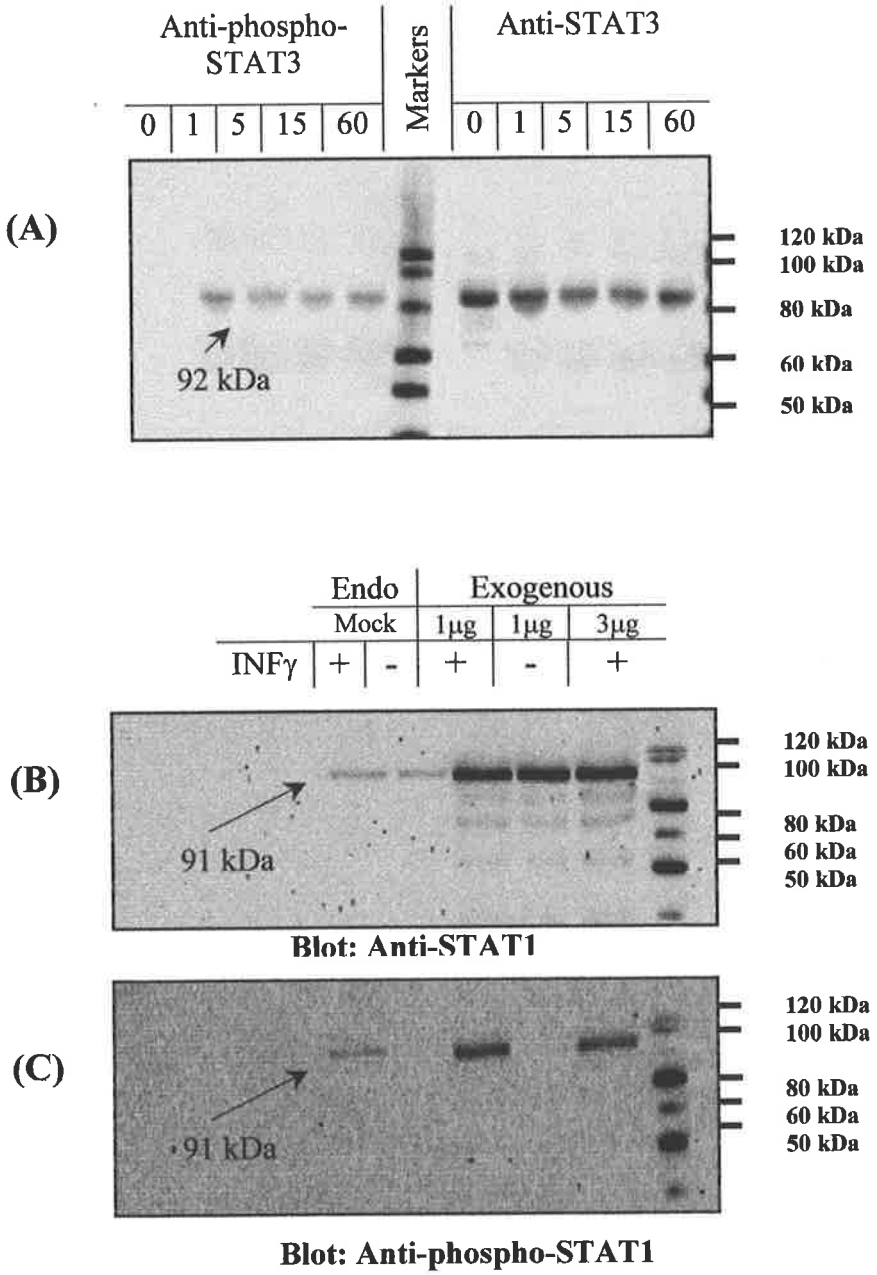


Figure 10: Anti-flag immunoprecipitations. PIAS1 fails to interact with flagged tagged ANK in 293T cells. Cells were transfected with PIAS1/pTarget and ANK/flag/pTarget. 24 hours post transfection cells were treated with INFγ and lysed. Recombinant proteins from total cell lysates (500 μg) were immunoprecipitated with 10 μg of anti-flag. Supernatants were retained and proteins eluted by boiling in 1X loading buffer. Proteins were resolved on NuPAGE gels, blotted and probed with either anti-flag (A) or anti-PIAS1 (B). Anti-flag immunoprecipitated exogenously expressed ANK, however no corresponding interacting PIAS1 was detected. In addition, two non-specific proteins were immunoprecipitated by anti-flag (*). For each sample the immunoprecipitated proteins (**Elute**) and the cell lysate, post immunoprecipitiation (**supernatant**) are presented.

Constructs	Ankyrin/PIAS1			MOCK			Ankyrin/PIAS1		
Immunoppt	Protein-G			Flag			Flag		
Load.	Elute	'	Supernatant	Elute	'	Supernatant	Elute	'	Supernatant

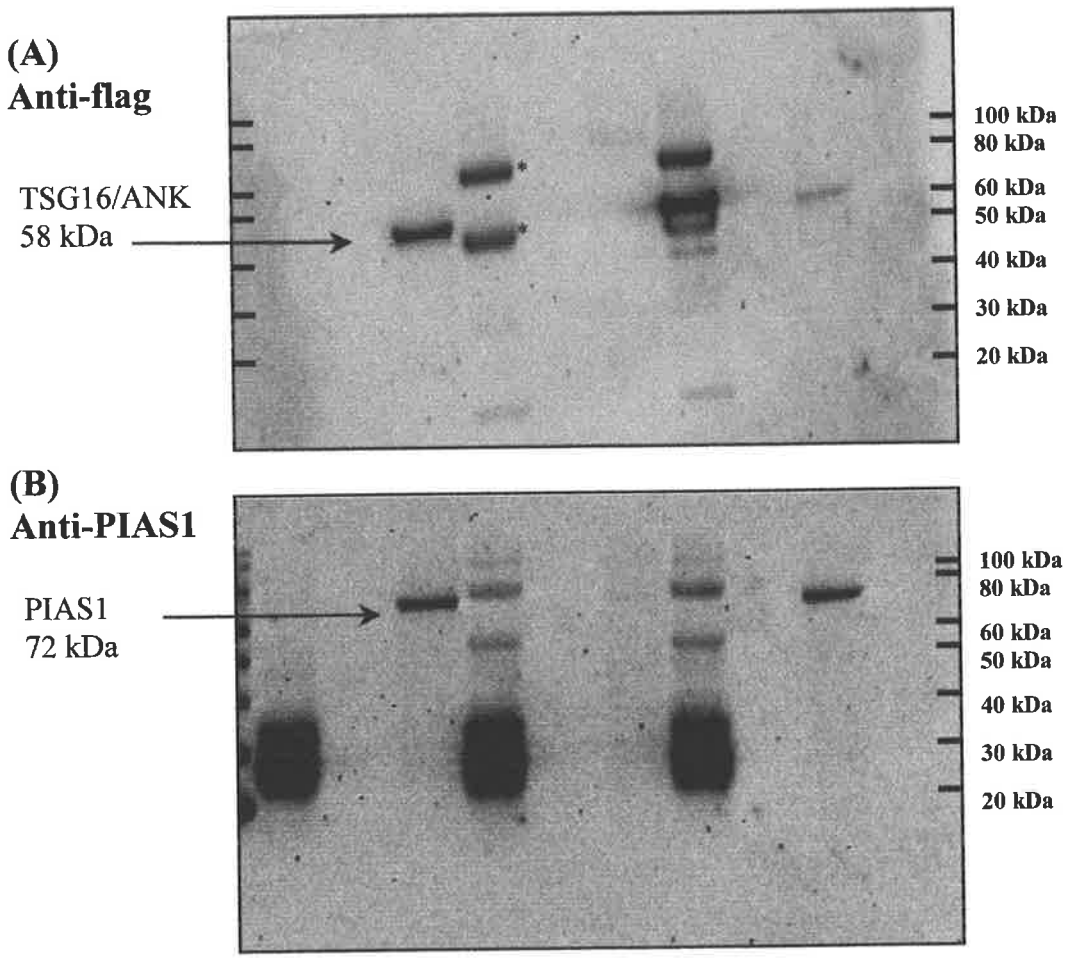


Figure 11: PIAS1 Immunoprecipitation. PIAS1 fails to interact with flagged tagged ANK in 293T cells. Cells were transfected with PIAS1/pTarget and ANK/flag/pTarget. 24 hours post transfection cells were treated with INFγ and lysed. Recombinant proteins from total cell lysates (500 μg) were immunoprecipitated with 10 μg of anti-PIAS1. Supernatants were retained and proteins eluted by boiling in 1X loading buffer. Proteins were resolved on NuPAGE gels, blotted and probed with either anti-PIAS1 (A) or anti-flag (B). Anti-PIAS1 immunoprecipitated exogenously expressed PIAS1, however no corresponding interacting ANK was detected. For each sample the immunoprecipitated proteins (**Elute**) and the cell lysate, post immunoprecipitiation (**supernatant**) are presented

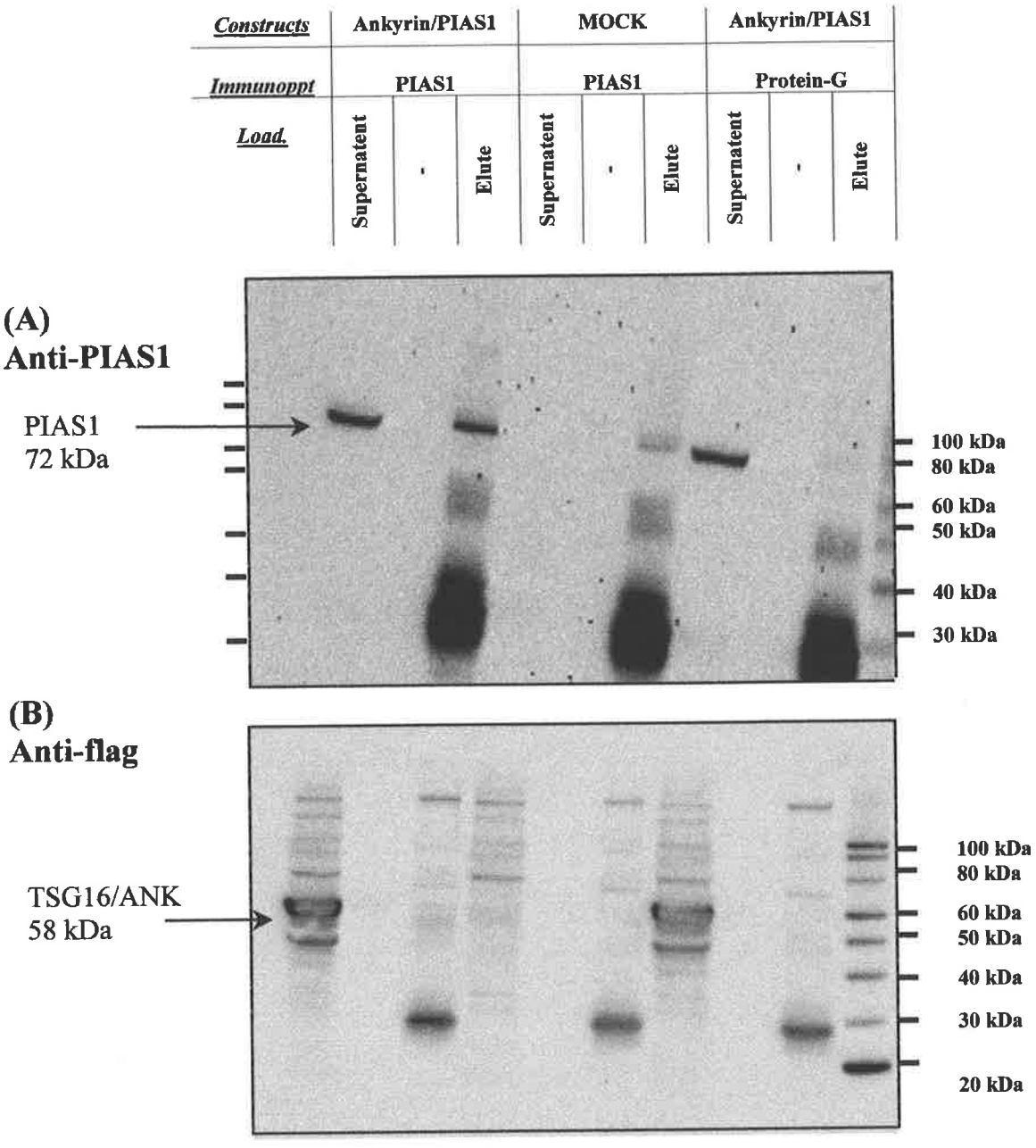


Figure 12:. PIAS3/GST fails to interact with flagged tagged ANK in 293T cells. Glutathione-sepharose immunoprecipitations. Cells were transfected with PIAS3/pDEST27, GST/pDEST27 and/or ANK/flag/pTarget. 24 hours post transfection cells were treated with LIF and lysed. Recombinant proteins from total cell lysates (500 µg) were immunoprecipitated with glutathione-sepharose. Supernatants (**Super**) were retained and proteins eluted with reduced glutathione (5mM-**GSSH**), followed by boiling (**Boil**). Proteins were resolved on NuPAGE gels, blotted and probed with either anti-flag (A) or anti-GST (B). Glutathione-sepharose immunoprecipitated exogenously expressed PIAS3/GST, however no corresponding interacting ANK was detected.

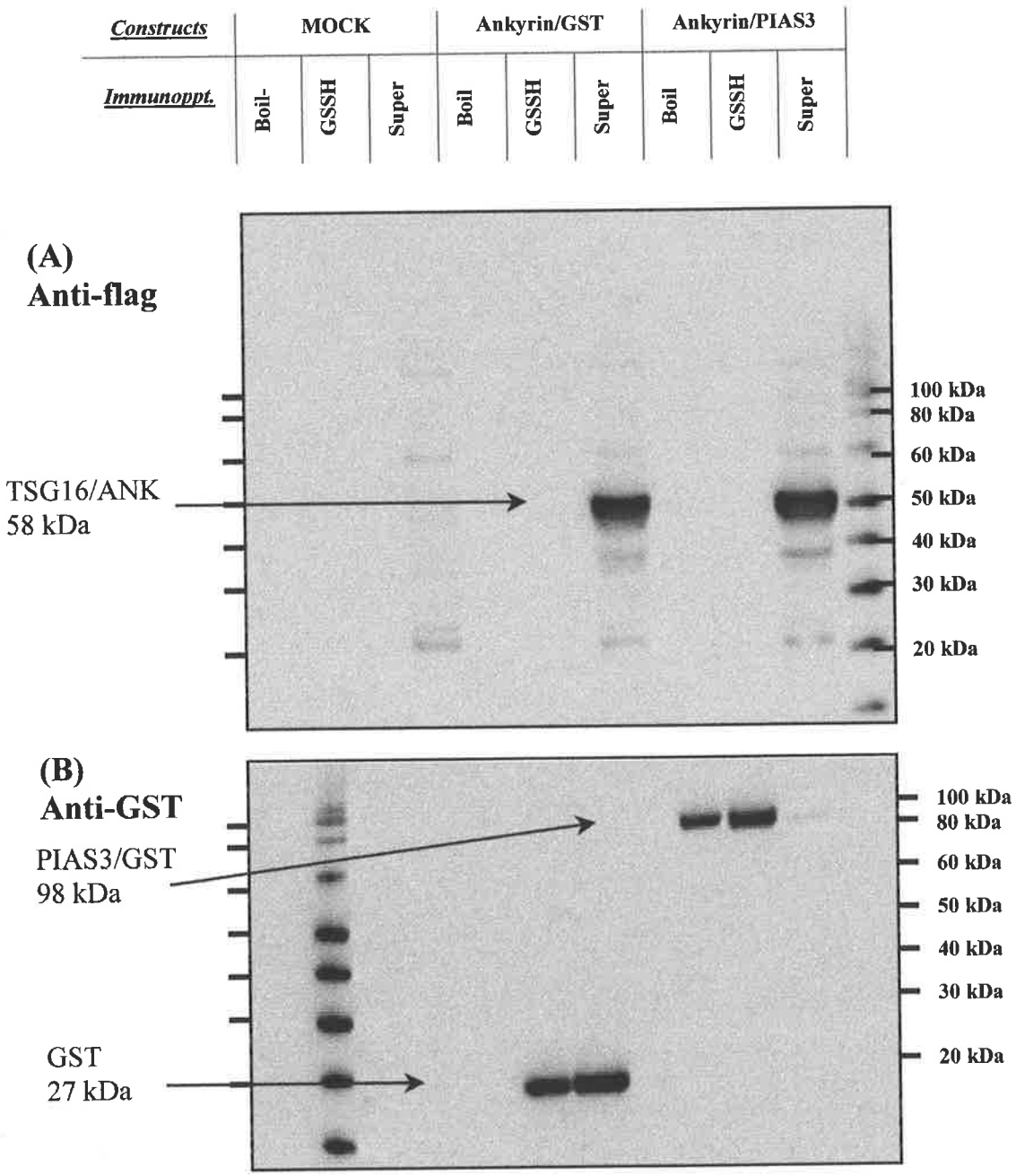


Figure 13: PIAS3/GST fails to interact with flagged tagged ANK in 293T cells. Anti-flag immunoprecipitates. Cells were transfected with PIAS3/pDEST27, GST/pDEST27 and/or ANK/flag/pTarget. 24 hours post transfection cells were treated with LIF and lysed. Recombinant proteins from total cell lysates (500 µg) were immunoprecipitated with anti-flag (10 µg), boiled, resolved on NuPAGE gels, blotted and probed with either anti-flag (A) or anti-GST (B). Anti-flag immunoprecipitated exogenously expressed ANK, however no corresponding interacting PIA3/GST was detected. For each sample the cell lysate (**Lysate**) immunoprecipitated proteins (**Elute**) and the cell lysate, post immunoprecipitiation (**supernatant**) are presented

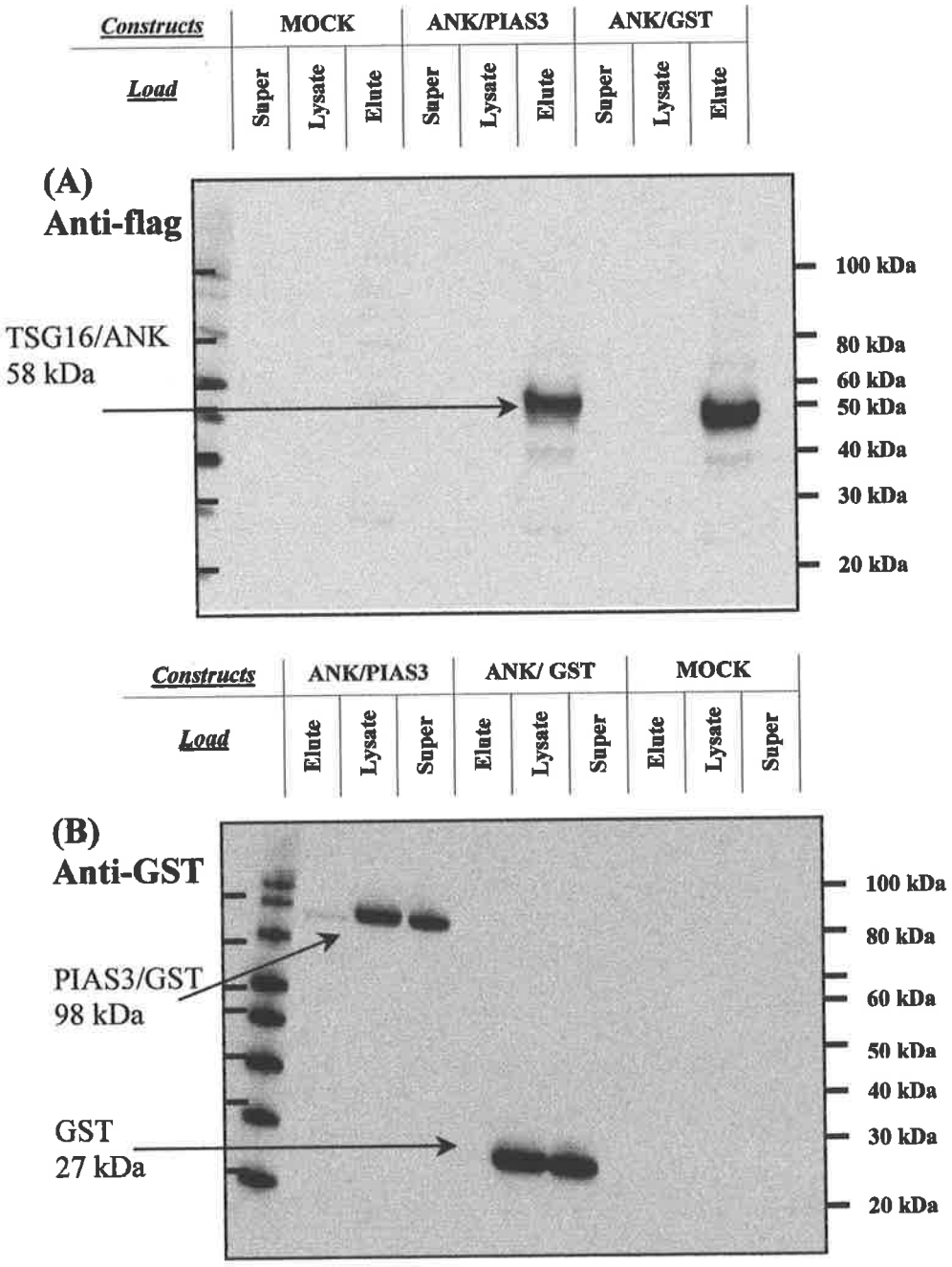
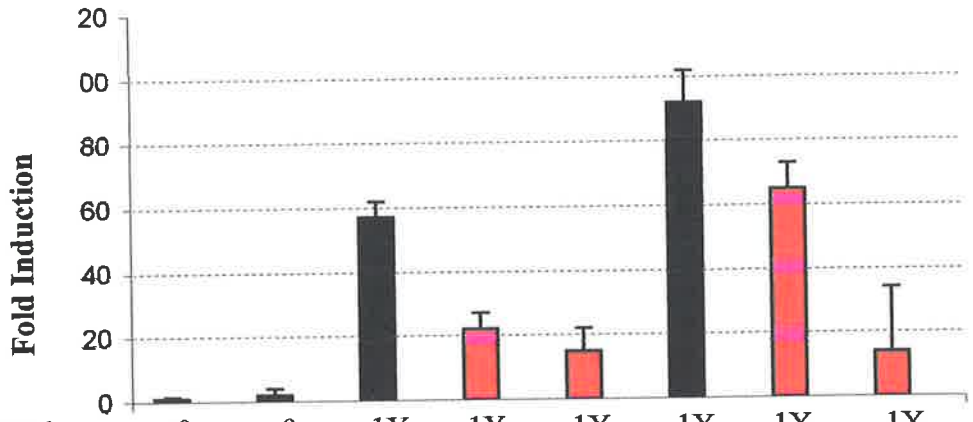


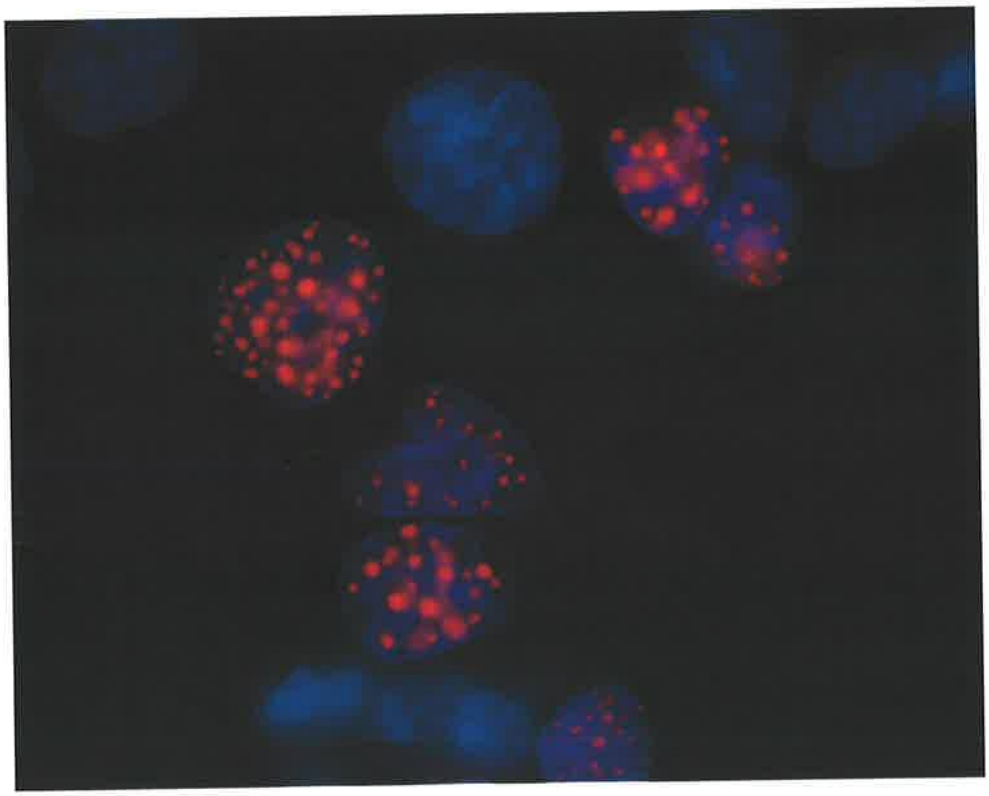
Figure 14: PIAS3 inhibits STAT3 mediated transcription. 293T cells were transfected with a STAT3 reporter constuct (pAPRE-luc or pSTAT3-luc) with combinations of PIAS3/pDEST12.2 and/or STAT3/pcDNA3.1. Cells were stimulated with LIF and the STAT3 transcriptional activity measured and expressed as fold induction over non-stimulated control cells. The pAPRE-luc reporter was stimulated 57 fold by endogenous STAT3 and 95 fold by exogenously expressed STAT3 (black bars). Exogenously expressed PIAS3 reduced endogenous STAT3 by 58% and exogenous STAT3 by 85% (red bars).



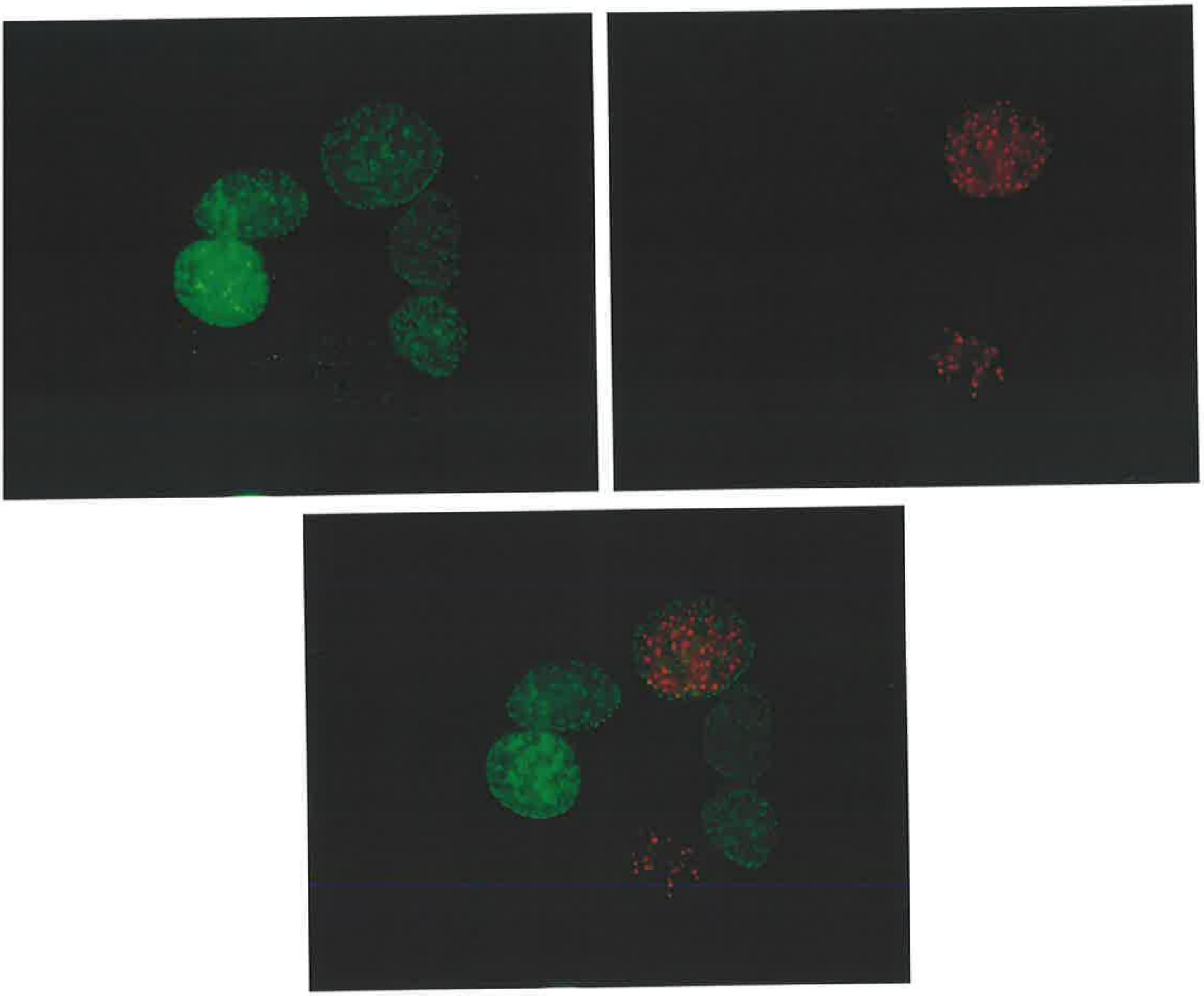
pAPRE-luc	0	0	1X	1X	1X	1X	1X	1X
pSTAT3-luc	1X	5X	0	0	0	0	0	0
PIAS3	0	0	0	1X	2X	0	1X	2X
STAT3	0	0	0	0	0	1X	1X	1X

Figure 15: TSG16 co-localization studies. 293T cells were transiently transfected with combinations of the following expression constructs TSG16/myc/pLNCX2, PIAS1/pDEST27, PIAS3/pDEST27, Ankyrin/flag/pTarget and PML/pDEST27. 24 h post transfection cells were treated with LIF (10ng/ml for 12 h) and recombinant proteins visualized with fluorochrome tagged secondary antibodies. Nuclei were stained with DAPI (blue fluorescence) and all proteins exhibited unique nuclear localizations.

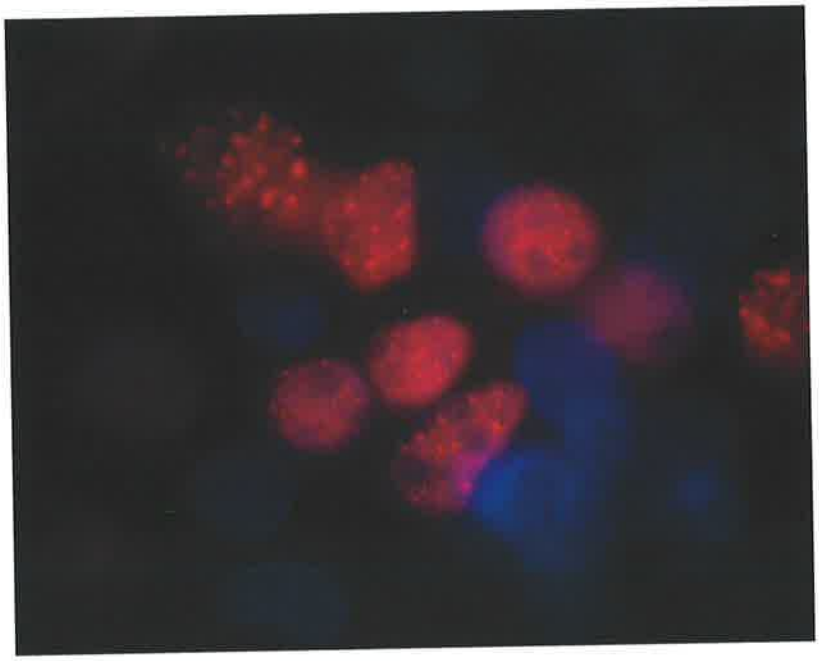
(A) TSG16/myc recombinant proteins were detected CY3 (red) labeled secondary antibodies. TSG16 shows nuclear dot localization.

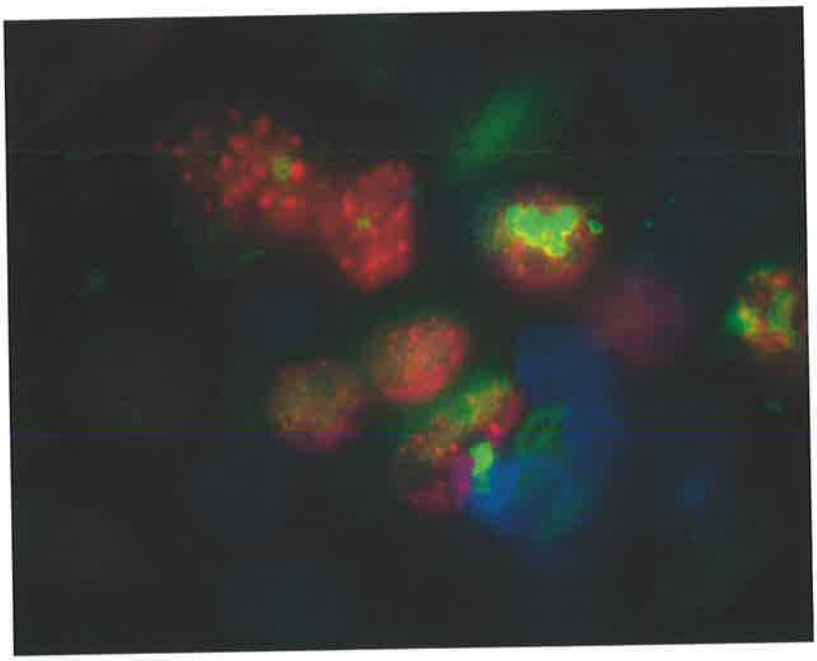


(B) TSG16/myc and PIAS3 co-transfected cells. PIAS3 exogenously expressed proteins were detected FITC (green) labeled secondary antibodies. The PIAS3 nuclear dots are more abundant and smaller than the red TSG16 dot structures. When the images are superimposed the proteins do not co-localize.

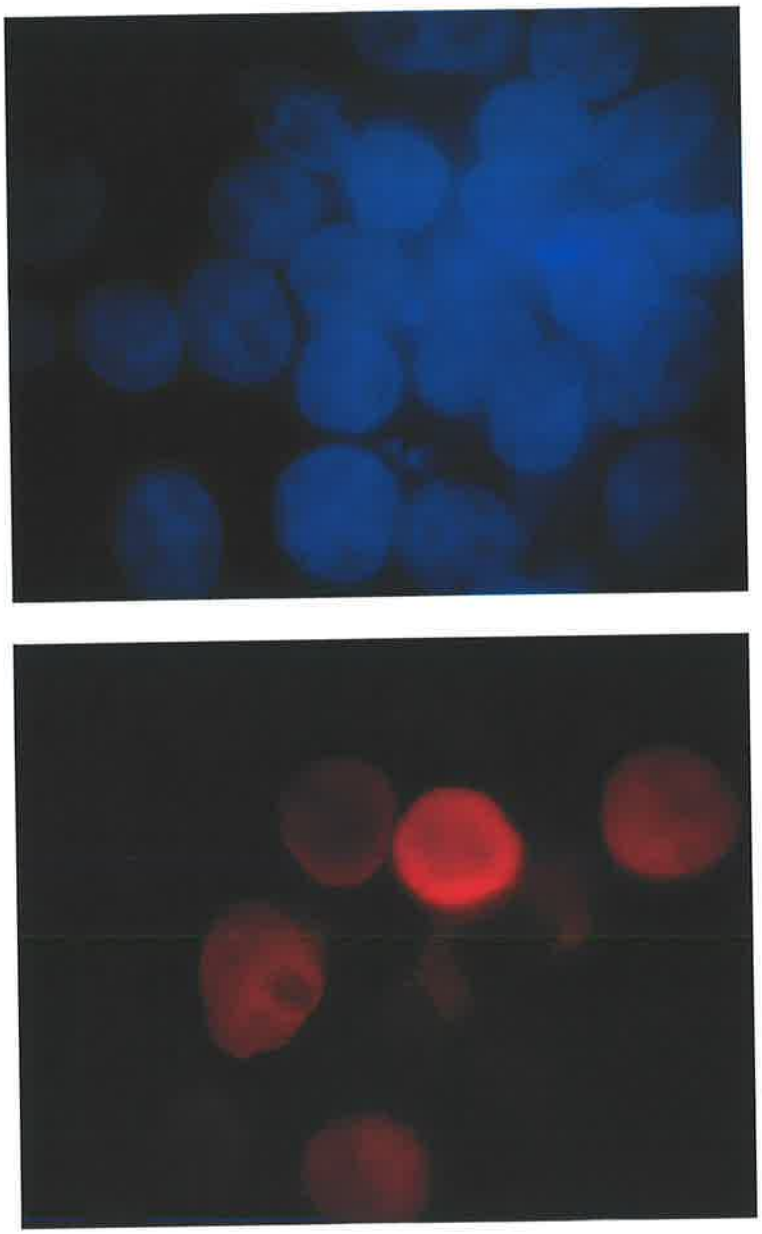


(C) TSG16/myc and PML co-transfected cells. PML exogenously expressed proteins were detected FITC (green) labeled secondary antibodies. The PML nuclear dots are more abundant and smaller than the red TSG16 dot structures. When the images are superimposed the proteins do not co-localize.





(D) Ankyrin/flag/pTarget recombinant proteins were detected CY3 (red) labeled secondary antibodies. Ankyrin/pTarget shows a diffuse nuclear localization.



Tables

Table 1: *(A)* Splice Sites of the TSG16.

Exon	Size (bp)	3' Splice site (intron/exon)	5' Splice site (exon/intron)	Intron size (bp)
1	317	tcctggagcc/AGAGGCCGCC	TCGGAGCACG/**gt**gagaggcg	71932
2	85	cttgtttc**ag**/AATATATACA	TATGGAATCG/**gt**aaatactc	113150
3	146	tcttttga**ag**/CAGACGGTTG	TGGGAAAAAG/**gt**aatgctgc	11588
4	142	cttttttct**ag**/GATAAAGATA	GACACAGAAG/**gt**accagccc	14022
5	171	cgcctcac**ag**/AGAAGCAGGG	AACAGCCCAG/**gt**gagcccgg	184
6	204	tggcctgc**ag**/TGGACACAAC	GACTTCGCAG/**gt**gggcaccc	1952
7	143	ctgtttcc**ag**/GCTGGACGGC	GCACTACAAG/**gt**tggcatcc	2341
8	148	tgtcttgt**ag**/GTGGTGAAGC	AGCTCGACGG/**gt**aagtcacg	389
9	6578	cctccaac**ag**/AGAGCTCAGA	CATCCCCGTG/**gt**gagtgcgg	3878
10	99	ttgcctgt**ag**/ATCGCACCCC	CATCGAGCGG/**gt**aagtgggc	135
11	144	cctccttc**ag**/GAGAAGCTGA	GGAGAGCCAG/**gt**agggcccg	3909
12	93	gtttccac**ag**/GGTGACGAGA	CCGCATGAAG/**gt**cggtgttg	2153
13	1037	tgtcccgc**ag**/ACTTGCCTCC	ATTAAAATGC/tctttgtaca	

(B) RT-PCR primer pairs used for confirmation of TSG16.

Primer Set	Nucleotide sequences (5'-3')	Size (bp)
Exon 1-2	1F CAG CCG TCG CGG AGC CGC GC	
	2R CCA TAG CTG AAA GTC AGT GC	115
Exon 2-3	2F TAT ATA CAG CCC TGC TCT GG	
	3R TGT GGT GCT TTA GGG CAC C	174
Exon 3-4	3F CAG ACG GTT GAT GAT GAA GC	
	4R TGT GTC CGA GTC CTT CTG C	288
Exon 4-5	4F GCG GAA GCT GCC CTT CACC	
	5R GGG CTG TTG GCA GAC TCC	228
Exon 5-7	5F GGC CCT TCT CAT GCA GAT G	
	7R GCG TCG TGC AAA GGC GTG	372
Exon 7-9	7F CTA CTA CGA CGT CGC GAA G	
	9R TGC CGT CGA CTG AAC TGG AAG	314
Exon 9-12	9F GGG AGG TGA TCC AGC AGA C	
	12R CCG CCG GGG CCA GTG AGG	1019

***Table 2*: Oligonucleotide primers for Mutation analysis of *TSG16*.**

Primer Set	Nucleotide sequences (5'-3')	Size (bp)
Exon 2	TAT ATA CAG CCC TGC TCT GG CCA TAG CTG AAA GTC AGT GC	286
Exon 3	GAG TGG ATG AGT TTT GAG CTG CCC CAT CTG GGT GCG GTG	255
Exon 4	AAC AAA ATC TTG AGA GCT TGA G CTC CGG GCT GCC TGT GG	258
Exon 5	GAC ACC ATG GCA TGA AAT CTC AGG CAG AGC CCC TTC CCT G	270
Exon 6	AAG CAG GAC CAC CCC AC A G GGC TCC AGT GGG GCT CT C	300
Exon 7	CAG GAG GGA AGT TCT GAG G CCA CCC AGA TGC CAG GAC	250
Exon 8	GGA ACG GTA GAG GAG TTG TG AAC ACC ACA GGG CAG CTC C	275
Exon 9.1	CTT CCC GCA GAC ACA TCT G TCG TCG TCC TCA TCA AAC TC	265
Exon 9.2	CAA GGC CAA GAA CCC AGA G CTC GTC CGA CGT GTC TGA C	255
Exon 9.3	ACC AAA GAA AGC GTC CCA TC CTC TGA GGA CTC GCT CTC C	267
Exon 9.4	CGG AAA GCG GAG CGA CAA G CTG ATA AAG AAC TGA CCT CTG	293
Exon 9.5	CGG ACA GAC AAT TGG AAA ACC AAC ACT GTC CCT TCT CCT TG	300
Exon 9.6	GTC CCC AAA CTG GAC AAG G TGA TTT TTC ATC TTT AAA GAG CC	297
Exon 9.7	AAG AAG AGA AAC TTA GCA AAA TG	

	TTC TCT TTT GAA ATT TTG TCC TC	286
Exon 9.8	ACC GGC CCT CAA AAT TAG AG GTA GTC TGT CAC TGG CGA G	283
Exon 9.9	GAT TTC AAA GGG GAG GAC AG CGG ACT CTC TCC TCT TCT TG	283
Exon 9.10	GTG TCC CTG GCT ACC TTT C TTT GTC TTC TCC TTC CTT TCC	286
Exon 9.11	ACG AGA GAA GGA GAA GAA GG TCA GAG AAG TCT TCT GAG ATG	284
Exon 9.12	GAA AGC GTC TCT CGA CCA AG TTG TCC TTC TGC CTC TCA GG	309
Exon 9.13	GGA AGT CTT CTG ACA AGC AG ATG TTC TTT GTC CGA CTT CTC	298
Exon 9.14	GAA GCA CGC TGC CGA AGA C GCG CTC TCC CTC GGC TTG	282
Exon 9.15	AGG TCT CTT TCA CGG AGC C CAA TGG TTT TAT CTA GCT CAT C	288
Exon 9.16	TAC AAC ATG AAA GCT GAC ATA G TCT GCT CGT CCC TGT GAT G	287
Exon 9.17	GCG ATG CCA AAC TGA AGG AG GCT TCG CCT TCT CCT TGA G	300
Exon 9.18	GAT CGA GGA GCG CCA CAA G TGG GCC GGC TCT GGT CAG	323
Exon 9.19	CAG GTG CAG ACT CCA AAG AC ACT TGA AGC CAC GGA GAA CC	293
Exon 9.20	CTC CCC CTC CTT TTT CGA C ACA AGC AGG CAA ACT TCT CC	272
Exon 9.21	GCC ACC CTC GAT GGA AGA C GGG AAG GAA CCA GCA GCT C	233
Exon 9.22	TCT CAA GTT TAC AAG CAA AAC C CTT CAG CAG GAG GTC CGA G	299

Exon 9.23	ACC TCT GAA AAC CCT GTG AG GCC TCT GAG GTG GAG ATG G	263
Exon 9.24	GAG GAC GTC AAG GAC GGA G ACC TGG CCG AGG TGA GAC	242
Exon 9.25	ATA GCT TCC TGG ACG GCA G CAG CCA CTA CGG TGG AAA C	255
Exon 9.26	CTC CAG AAG AGA TGC CTC C CTG GTC TCC GCT CCC CAG	284
Exon 9.27	ATC CGG ACT CCA GCG TGG CGG CCG TGG GTT TTG GTT C	264
Exon 9.28	ACA CTG AGG CCT CCC GTG GCG GCG TTT GCG CGG ATG	287
Exon 9.29	CCC CTC CGA GAA GGA GTG CTT CTC CTC GAT CTT CTT CC	295
Exon 9.30	GAT GCC ATC GAG CCC TAC C GCC ATG AGT GGG ACA AGA C	298
Exon 9.31	GAG AAA CAC AGG GAG AAA TGG CTC TAT CAT TCC CGT TGC TC	301
Exon 10	TAT CCT TCG TGC AGT TCC AC GGG ACA CAG CCA CGC TCC	264
Exon 11	GGA GCG TGG CTG TGT CCC GGT CAA AGT GCA GAA TCT ATC	279
Exon 12	CTT GAG AAC GGT CAC TGC AG GTC ACC ACC CAT CAC AGA AC	257
Exon 13	GTC TGG GGC TCT CCC TTC CAG TCC TGG GCA GCC GTG	296

Mammalian Expression Constructs	Description	Primers	Size of ORF (bp)	Size of Protein (kDa)	Source
TSG16/pLNCX2	CMV promoter c-terminal myc-tag	See table 4.	7996	298	Jason Powell
ANK/pTarget	CMV promoter c-terminal flag-tag Non-directional cloning	F-5′ GTG GCC CTT CTC ATG CAG AT R-5′ TGC CGT CGA CTG AAC TGG AAG GTG	1380	58	Hayley Spendlove
PIAS1/pTarget	CMV promoter Non-directional cloning	F-5′ ATG GCG GAC AGT GCG GAA CT F-5′ CAG TCC AAT GAA ATA ATG TCT GG	1956	72	Hayley Spendlove
PIAS3/pDEST12.2	CMV promoter Gateway cloning	F-5′ GGG GAC AAG TTT GTA CAA AAA AGC AGG CTT CAC CAT GGC GGA GCT GGG CGA ATT AAA G F-5′ GGG GAC CAC TTT GTA CAA GAA AGC TGG GTC TCA GTC CAG GGA AAT GAT GTC TG	1860	67	Jason Powell
PIAS3/pDEST27	CMV promoter n-terminal GST-tag Gateway cloning	F-5′ GGG GAC AAG TTT GTA CAA AAA AGC AGG CTT CAC CAT GGC GGA GCT GGG CGA ATT AAA G F-5′ GGG GAC CAC TTT GTA CAA GAA AGC TGG GTC TCA GTC CAG GGA AAT GAT GTC TG	1860	67+GST=98	Jason Powell
STAT1/pcDNA3.1	CMV promoter Blunt ended cloning	F-5′ GTA GCT GCT CCT TTG GTT GAA TCC R-5′ AGG CTG GCT TGA GGT TTG TA	2253	91	Jason Powell
STAT3/pcDNA3.1	CMV promoter Cloned into NotI and HindIII sites	F-5′ TAA ATT AAG CGG CCG CAC CCC GGC C TTG GCG CTG TCT CT R-5′ GAT GAA GCT TGA TGA TCT GGG GTT TGG CTG TGT	2304	92	Jason Powell

Table 3: Summary of constructs used for TSG16 functional studies.

Reporter Constructs	Description			Source
	Promoter	Enhancer	Reporter	
pTA-luc	Minimal TA promoter TATA box of from the thymidine kinase basal promoter of the herpes simplex virus (HSV-TK)	None	Firefly luciferase 61 kDa	BD Biosciences
pSTAT3-TA-luc	Minimal TA promoter. TATA box of from the thymidine kinase basal promoter of the herpes simplex virus (HSV-TK)	3 copies of the STAT3 bindind sequence. STAT13 homodimer specific.	Firefly luciferase 61 kDa	BD Biosciences
pGAS-TA-Luc	Minimal TA promoter TATA box of from the thymidine kinase basal promoter of the herpes simplex virus (HSV-TK)	2 copies of INFγ activation sequence (GAS). STAT1 homodimer specific	Firefly luciferase 61 kDa	BD Biosciences
pAPRE-luc	Minimal junB promoter	4 copies of APRE (acute phase respones element)	Firefly luciferase 61 kDa	Sandra Nicholson
pRL-CMV	CMV promoter	None	*Renilla* luciferase 38 kDa	Promega
β-gal/pcDNA3.1	CMV promoter	None	β-galactosidase	Marina Kochetkova

Construct	*Description of cloning procedure*	*Primer sequence (5'→3')*	*Product size*
TSG16 1-3/pBluescript	Forward primer within exon 3, 3 bp 5' of the start codon with an engineered 5' HindIII site (red). Reverse primer within exon 9, 99 bp 3' of the XmaI site. pBluescript was digested with HindIII and XmaI.	**F1**-AAT ATA AGC TTG ATC GGG TGA GGA GCA GG **R1**-GCT TTT CAG TCT GCT CTT TCC	2802
TSG16 4-6/ pSP72	Forward primer within exon 9, 101 bp 5' of the XmaI site. Reverse primer within exon9, 3 bp 3' of the EcoRI site. pSP72 was digested with XmaI and EcoRI.	**F2**-CAG ACT ACA GGG ACA TGA AG **R2**-GCT TTT CAG TCT GCT CTT TCC	2847
TSG16 1-6/pSP72	TSG16 1-3/pBluescript was digested with HindIII and XmaI to release insert (TSG16 1-3). TSG16 4-6/ pSP72 was digested with XmaI and HindIII (from pSP72 multi-cloning site) for insertion of TSG16 1-3.	-	5458
TSG16 7-8/pSP72	Forward primer within exon 9, 40 bp 5' of the EcoRI site, containing a 5' engineered HindIII site (red). Reverse primer within exon 13, 6 bp 5' of the stop codon, containing a 3' in-frame myc tag (blue), ClaI site (red) and stop codon (underlined). pSP72 was digested with HindIII and Cla1.	**F3**-CGC TCG GTC TCT GTC GAC **R3**-TAT TTA TCG ATT CAG CAG GTC CTC CTC GCT GAT CAG CTT CTG CTC TAC AAA GTC GTC GTT GAC GTC	2620
TSG16 1-8/pSP72	TSG16 7-8/pSP72 was digested with EcoRI and ClaI to release insert (TSG16 7-8). TSG16 1-6/ pSP72 digested with EcoRI and ClaI (from pSP72 multi-cloning site) for insertion of TSG16 7-8.		8078
TSG16 1-8/pLNCX2	TSG16 1-8/pSP72 was digested with HindIII and ClaI to release insert (TSG16 1-8). pLNCX2 was digested with HindIII and ClaI for insertion of TSG16 1-8.		8078

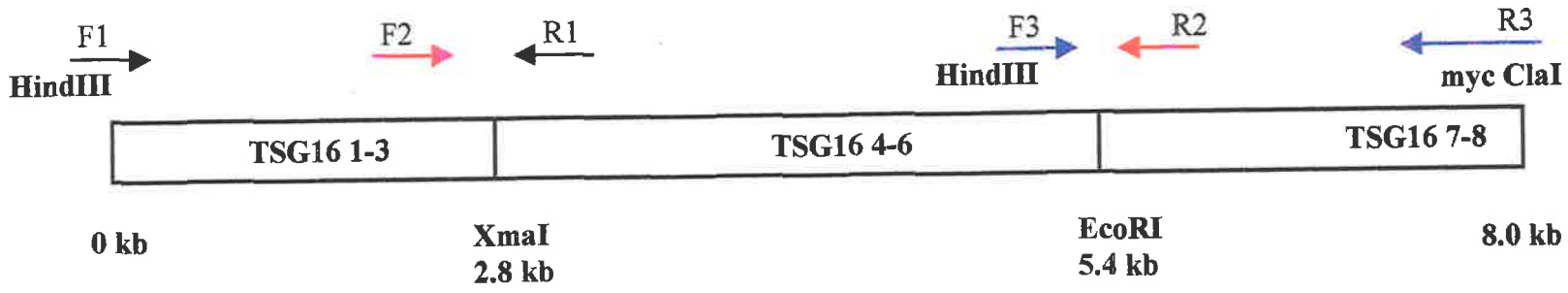


Table 4: Summary of TSG16 cloning procedure.

Table 5: (A) Results of TSG16 SSCP analysis on sporadic breast tumours.

Exon	Position	Tumour ID	Polymorphism						
			Paired Tumour/blood DNA				Normal		Present in the normal population
			Blood		Tumour				
			Nuc.	Amino acid	Nuc.	Amino acid	Nuc.	Amino acid	
8	***1353***	***88/248[1]***	***A(C/T)G***	***Thr/Meth***	***A(T)G***	***Meth***	***A(C)G***	***Thr***	***No (0/165)***
9.10	3375	29/96[1] 12/96[1]	G(C/T)G	Ala/Val	G(T)G	Val	G(C/T)G	Ala/Val	> 5 % (2/50) Yes
9.20	*5802*	*91/250[1]*	*(G/A)CC*	*Ala/Thr*	*(A)CC*	*Thr*	*(G/A)CC*	*Thr/Ala*	*0.6 %* *(1/165)*
9.23	6546	19/96[1] 8/96[1] 757[2] 670[2]	(C/G)CC	Pro/Ala	(G)CC	Ala	(G/C)CC	Ala/Pro	>10% (7/50)
9.27	7290	757[2] 919[2] 573[2]	(C/T)CC	Pro/Ser	(T)CC	Pro/Ser	(T/C)CC	Pro/Ser	4 % (4/100) Yes

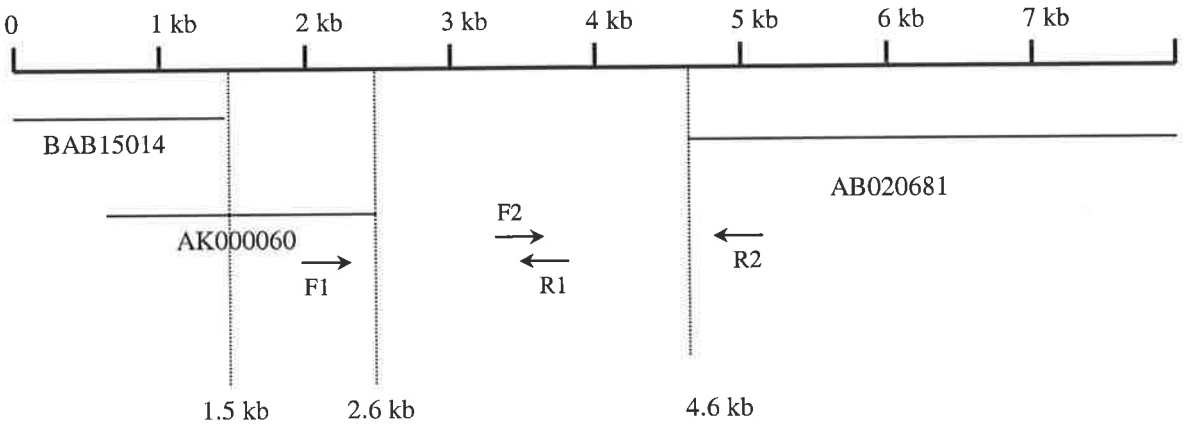
[1]: Flinders tumours
[2]: Lieden tumours

(B) TSG16 polymorphisms detected in KconFab DNA.

Exon	Position	KConFAB family ID	Polymorphism						
			Paired Tumour/blood DNA				Normal		Present in the normal population
			Blood		Tumour				
			Nuc.	Amino acid	Nuc.	Amino acid	Nuc.	Amino acid	
7	-	*0001.00.009* *00.009.0429*	*Not Sequenced*		*Tumour requested*	-	*Band shift absent*	-	*No*
8	-	*0009.00.005*	*Not sequenced*	-	*Tumour requested*	-	*Band sift absent*		*No*
9.13	-	00.004.0274 02.002.0088 98.002.0064	Not sequenced	-	-	-	Band shift present	-	Yes 10 %(10/100)
9.14	-	41 set 1 00.004.0274 02.002.0088 98.002.0064	Not sequenced	-	-	-	Band shift present	-	Yes 11 %(11/100)
9.16	-	0009.00.005	Not sequenced	-	-	-	Band shift present	-	Yes 5 %(5/100)
9.18	-	00.004.0274 02.002.0088 98.002.0064 99.009.0318 98.005.0214 99.006.0407 98.005.0268	Not sequenced	-	-	-	Band shift present	-	Yes 7 %(7/100)
9.20	**5801**	**0015.99.002** **0007.99.007**	**(A/G)CC**	**Thr/Ala**	**Tumour requested**	-	**(A/G)CC**	**Thr**	**Yes 0.5 % (1/200)**
9.21	**5729**	**0019.99.006**	**(C/T)CC**	**Pro/Ser**	**(C/T)CC**	**Pro/Ser**	**C(C)C**	**Pro**	**No (0/200)**
9.27		21 set 2 22 set 2	Not sequenced	-	-	-	Band shift present	-	Yes 11 %(11/100)

Table 6:

(A) Schematic representation of the *in silico* cloning of TSG18



(B) RT-PCR primer pairs used for confirmation of TSG18

Primer Set	Nucleotide sequences (5'-3')	Size (bp)
Exon 7	F1 CTT CTT TGG GTA TGA GTG C	
	R1 AAC CGT TCC TTT TCT TTT TG	973
Exon 7	F2 AAG ATA CCC GAC CTA AAG AAA AGA	
	R2 GCC GGA TCC TGA GCA CA	1267

(C) Splice sites of TSG18

Exon	Size (bp)	3' Splice site (intron/exon)	5' Splice site (exon/intron)	Intron size (bp)
1	163	tgtcctggga/GCGAGGAGGC	CTTGCCTCGG/**gt**ccgtacgggt	45416
2	138	ttctttac**ag**/ATCCAGGATG	TGGAAGAAAG/**gt**atatgatt	13031
3	148	tttttaat**ag**/AGTAAAGACA	TCAGACACAG/gtttgtttca	8777
4	69	cattctgt**ag**/ATTCAGATCC	ACATACTCAG/**gt**aattagat	4113
5	147	ataattcc**ag**/AGAAAGAAGG	AACAGTCCAG/**gt**gataccta	2780
6	201	tttcttac**ag**/ATTCCACACC	GATTTTGCAG/**gt**aagactag	4973
7	143	ctaactttag/GTTGGACACC	GCACAGAGAT/**gt**aagtatga	4951
8	148	attgttgc**ag**/ATAGTAAAGC	AGTTACACAG/**gt**ttgtttca	32211
9	4721	tttattct**ag**/ATTCCGAAGA	TATTCCCACA/**gt**aagtaaca	4858
10	99	ttttaatt**ag**/ATTACACCAC	TATTGAAAGG/**gt**aagaaatg	11635
11	144	attcaact**ag**/GAAAAACTCA	GGACTCTCAG/**gt**aaaatgtt	3881
12	96	tctcttttag/TCTGATGACA	CAAATTAAAG/**gt**atgtatgt	1296
13	2818	ttttaaat**ag**/ACCTGTCTTT	3'UTR	unknown

Table 7: Results from the Y2H screen performed by Ms Hayley Spendlove. The ANK domain was used as bait to screen a human breast cDNA library.

Sequence recovered	Description	Number of clones	% of total clones recovered
PIAS1	- Inhibitor of STAT1 - Sumo E3 ligase	45/113	40
PIASxα	- Regulates steroid reeptors - Sumo E3 ligase	17/113	15
BRD7	- Bromo-domain containing protein	17/113	13
Ubc9	- Sumo E2 conjugating enzyme	11/113	10
CTCL	-**CTCL** tumor antigen se14-3 (Cutaneous T-cell lymphoma associated antigen se14-3) - Protein kinase C binding protein 1 (Rack7)	9/113	8
PIAS3	- Inhibitor of STAT3 - Sumo E3 ligase	8/113	7
Pm-Scl	-A 75 kDa autoantigen involved in polymyosititis/ scleroderma overlap syndrome	3/113	3
FAF1	- FAS associated factor 1 - Enhances the apoptotic activity of FAS	2/113	2
Sp100	-Constituents of nuclear bodies or PODs	1/113	1

Chapter 4

Chapter 4. Characterization of the YAC from 16q24 postulated to carry a senescence gene for breast cancer cells

4.1) Introduction

Functional evidence from micro-cell mediated YAC transfer experiments supported the presence of a breast cancer cellular senescence gene at 16q24 (Reddy *et al*, 1999). The introduction of chromosome 16, or segments of 16q, into human and rat mammary tumour cell lines resulted in the restoration of replicative senescence. Escape from senescence in tumour cells has been postulated to be an early step in the evolution of malignant tumours (O'Brien *et al*, 1986; Medina, 1996; Yeager *et al*, 1998).

Currently these YAC experiments provide the only independent functional data for the presence of a tumour suppressor at 16q24.3. Using similar functional complementation assays, cellular senescence genes have been identified on at least ten different chromosomes (reviewed in Oshimura and Barrett, 1997). However, so far only one gene, MORF4, which restores senescence in immortal Hela cells, has been cloned based on such complementation data (Bertram *et al*, 1999).

YAC transfer experiments into the immortal human (MCF-7) and rat (LA7) mammary tumour cell lines allowed the construction of a YAC physical map of the genomic interval believed to carry the 16q24 cellular senescence gene (Reddy *et al*, 1999, 2000). YAC clones were retrofitted to incorporate the *neo* selectable marker and transferred into immortal human mammary cells in a two-step process, spheroplast fusion with the mouse cell line A9, followed by microcell mediated delivery to mammary cells. This two-step process avoided the problems associated with the inefficient delivery of YAC DNA into human cells. The introduction of non-desirable mouse background DNA was an unavoidable consequence of the two-step process. Microcell hybrids from both MCF-7 and LA7 cells exhibited enlarged, flat, vacuolated morphology and underwent cell cycle arrest. These alterations were indicative of cellular senescence. Furthermore, the hybrids underwent reversion to immortal growth concordantly with the loss of the donor YAC DNA. YAC clones from other genomic regions had no effect on the growth characteristics of recipient cells. This same YAC clone was reported to restore replicative senescence in immortal cell lines derived from hepatoma, prostate, ovarian tumours and SV40 transformed cells (RS Athwal, unpublished results), implying this causative gene may be common to many different tumours.

Six YAC clones spanning a 3-7 cM region were isolated and analysed for the presence or absence of 40 DNA markers mapped to 16q24.3 (Reddy *et al*, 1999, 2000). High-resolution pulse field mapping enabled YAC clone overlaps to be established and the candidate region was reduced to a single 360 kb YAC clone, 792E1. The markers D16S498 and D16S476 initially defined the genomic interval encompassed by YAC 792E1 (Reddy *et al*, 2000). The availability of the previously described sequenced physical map of 16q24.3 (Chapter 2 and Powell *et al,* 2002) enabled refined positioning of YAC 792E1, to between the markers D16S3123 and D16S476.

This chapter describes identification and characterization of the two candidate genes, *FBXO31* and Hs.118944 present in YAC 792E1. The genomic structure and expression profile of these uncharacterized Unigene EST clusters is initially presented, and subsequently these candidates were screened for potential tumour suppressor function by three independent methodologies, mutational analysis in breast tumour DNA, expression analysis in breast cancer cell lines exhibiting LOH (2.3.3.3) and cell based assays for functional complementation.

4.2 Materials and Methods.

4.2.1) YAC 792E1 mapping. YAC 792E1 was obtained from Research Genetics Inc. The available sequenced physical map allowed selection of markers for PCR based mapping (2.2.1.1). BAC, PAC and YAC DNA were used as templates in standard PCR reactions and preliminary analysis confirmed the presence of three genes (data not shown). Many Unigene clusters were rejected from analysis as they failed to fulfill the criteria indicative of "real" genes (2.2.2.2). RT-PCR products were generated for (2.2.2.5) *FBXO31* (5' CCGGCGGGAGGCAGGAGGAGT and 5' GCGGCGGTAGGTCAGGCAGTTGTC), Hs.112744 (5' CCACGCGGGTCCATTAGAAGC and 5' GGGAAGCCACCTCGGACTGA) and Hs.118944 (5' GCGGGGTGCGTCGGGCTCTG and 5' AGGGGTGCTGCTTTCGGGGTCTCA), and used to probe the *Eag I* restricted BAC and PAC contig (2.2.1.7). All three genes, *FBXO31,* Hs.112744 and Hs.118944, hybridized to a 45 kb *Eag I* restriction fragment, confirming their presence in this genomic interval (Fig 2). Multiplex PCR experiments using 5' and 3' specific Hs.112744

primers showed that Hs.112744 lies within the genomic region present in our BAC/PAC contig but is deleted in YAC 792E1 (Ms Jo Crawford, personal communication).

4.2.2) Transcript characterization. For both Hs.118944 and *FBXO31,* full-length cDNA clones were pieced together from overlapping ESTs, and their corresponding genomic structures identified by alignment with the available genomic sequence. RT-PCR with primer pairs designed to cross intron/exon boundaries confirmed the transcription status of both Hs.118944 and *FBXO31* (2.2.2.5).

4.2.3) Northern blot analysis. Human multiple tissue Northern blot membranes (BD Biosciences) were probed with RT-PCR products (2.2.2.5) generated with the primer combination 5' TGCGAAGCTGCTTCACCGAT and 5' GGCCGTACATGCACTCCACTG for *FBXO31* and 5' ATGTGTACGTCA and 5' AGTGCAGTACACGATG for Hs.1188944 (3.2.2). After the cDNA probe had been labeled, QIAquick cleaned and pre-reassociated (2.1.1.8), they were hybridized in 10 ml of ExpressHyb solution containing denatured salmon sperm DNA (100 μg/ml) overnight at 65°C. The following day membranes were washed three times in 2XSSC, 1% SDS for 10 min at room temperature. Methods are described in detail in 3.2.2.

4.2.4) Single-stranded conformation analysis and sequencing. SSCA was performed on the entire open reading frames and the flanking intronic regions of both Hs.118944 and *FBXO31,* as essentially described in section 3.2.3. Hex-labelled primers were designed in exon flanking intronic regions (Tables 3 and 6) and used to screen paired breast tumour DNA and normal DNA from peripheral blood lymphocytes of the same patient (n=46). The breast tumour panel exhibiting 16q LOH has been previously described (3.2.3). Products were resolved on 4% acrylamide gels using the GelScan 2000 (Corbett Research). PCR products showing conformational change were re-amplified with unlabelled primers and sequenced using BigDye terminator chemistry (Perkin-Elmer: 2.2.1.12). Olivia McKenzie and myself performed the screening

4.2.5) Plasmid constructs. All constructs were amplified from normal breast RNA (BD Biosciences) using Superscript RNase H- reverse transcriptase (Invitrogen) according to the manufacture's protocol (2.2.2.5). Full-length transcripts were cloned into the retroviral expression vector pLNCX2, with 3' in-frame myc-tags. Directional cloning of all isoforms was achieved with forward primers containing *BamHI* restriction sites and reverse primers containing *Hind III* restriction sites. The myc-tag was engineered into the reverse primer. Products were amplified, digested with *Hind III* and *BamHI* (2.2.1.5), and ligated (2.2.1.9, 2.2.1.11) into the restricted pLNCX2 vector. All clones were sequenced verified (2.2.1.12). Primer sequences used for amplification are presented in Table 2. Dr Marina Kochetkova cloned the *FBX031*/pLNCX2 construct.

4.2.6) Cell culture and retroviral infection. The breast cancer cell lines MCF-7 and SKBR3, HEK 293T cells, and the immortalized breast epithelial cell line MCF-12A, were maintained and infected as previously described (3.2.5 & 3.2.6). 24 h post-infection cell assays were conducted.

4.2.7) Cell growth suppression assay. To assay monolayer colony formation, 3 x 10^3 of infected tumour cells were plated in 6-well dishes in the presence of 500 ng/ml G418 (Sigma). After 2 weeks of selection, cells were fixed in 3.7% formaldehyde in PBS, stained with Giemsa, and dried for subsequent quantification. The number of colonies visible to each well without magnification was determined. The Anchorage independent growth assay of colony formation has been previously described (Wang *et al*, 1993). Dr Marina Kochetkova performed the growth suppression assays.

4.3 Results.

4.3.1) YAC 792E1 mapping. The availability of the physical map at 16q24.3 enabled YAC 792E1 to be precisely located. Markers from the 16q24.3 interval were selected for PCR analyses of YAC template DNA. YAC 792E1 spans approximately 360 kb of genomic DNA and is positioned between markers D16S3123 and D16S476 (Fig 1). The initial marker order presented by Reddy *et al* (2000) was shown to be inconsistent with the order established from the BAC/PAC physical map (Chapter 2, Fig 6 & 7). Considering the BAC/PAC physical map was sequenced verified this marker order was accepted as the consensus.

BLAST analysis of the NCBI non-redundant, Unigene, and EST databases (3) using the genomic sequence of the region as query identified three plausible genes *FBXO31*, Hs.112744 and Hs.118944. Several additional Unigene clusters and singleton ESTs were rejected as they failed to fulfill criteria indicative of "real" genes and were interpreted as database contaminants (see section 2.2.2.2). Southern blot analyses of RT-PCR products generated from *FBXO31*, Hs.112744 and Hs.118944 against the PAC/BAC contig (2.2.1.7) confirmed the presence of all three genes (Fig 2). All three genes hybridized to a 45 kb *Eag I* restriction fragment, present in 11 overlapping BAC clones including the subsequently sequenced BAC RP11-534N20.

PCR analyses from the YAC template was performed to ensure that all three genes within the BAC/PAC contig were contained within the YAC, 792E1 (Ms Jo Crawford, personal communication). Multiplex PCR with 5' and 3' specific Hs.112744 primers showed this transcript lies completely within a genomic region presented in the BAC/BAC contig, but was deleted in YAC 792E1 (Fig 1 & chapter 2, Fig 6). The exact size of this YAC interstitial deletion remains undefined but was shown to encompass the entire Hs.112744 transcript. *FBXO31* and Hs.118944 specific primers amplified from both YAC and BAC template DNA. Thus Hs.112744 was omitted as a candidate cellular senescence gene and the remaining two candidates, *FBXO31* and Hs.118944, were further examined. It is possible that the YAC 792E1 clone examined in these studies has undergone some chromosomal rearrangement, and therefore different to that of the original study by Reddy *et al* (2000).

4.3.2) Hs.118944. Hs.118944, contained within the YAC 792E1, is predominately a 1.5 kb testis specific transcript (Fig 3b and 4a). Northern blot analysis with full-length cDNA of 1,435 bp, failed to detect any expression in all other tissues examined. However, examination of the EST database suggests this gene may have a wider tissue expression profile with sequences generated from heart, liver and brain cDNA libraries.

The genomic structure of Hs.118944 was established by the alignment of full-length Hs.118944 with the available genomic sequence at 16q24.3. Hs.118944 consists of 7 exons ranging in size from 37 bp to 187 bp and spans a genomic interval of approximately 15 kb. Exon 1 contained the two different start codons while exon 7 contained two different stop codons and the 3' UTR. All splice junctions conformed to the consensus sequence (Shapiro and Senapathy, 1987) and were subsequently confirmed by RT-PCR (data not shown). The intron and exon boundaries are presented in Table 1.

Four alternative Hs.118944 isoforms were identified. Primer pairs designed to amplify across the exon 1/2 boundary and subsequent sequence analysis revealed exon 2 utilizes 3 alternative splice donor sites, extending this exon by 4 bp (isoform B) or 42 bp, resulting in two different open reading frames (Fig 4a & b). The +42 bp isoform maintains the same open reading frame as isoform A and extends it by 14 amino acids. An alternatively spliced isoform excluding exons 3 and 5 was also detected, producing a third and fourth alternate open reading frames (isoform C and D) (Fig 4b). The four different open reading frames exhibit no known mouse or human protein homologies. Extensive database searches also failed to identify any homologies. RT-PCR analysis was performed with physically adjacent ESTs and *in silico* predicted coding sequences in an effort to extend the transcript and uncover an alternative homologous open reading frame. No adjoining exons were identified (data not shown). Isoform A exhibited a 720 bp open reading frame encoding a protein with a predicted molecular weight 21kDa. Similarly, isoform B has an open reading frame of 675 bp encoding a 25 kDa protein and isoforms C and D have open reading frames of 415 and 432 bp encoding proteins with predicted molecular weights of 15 and 17 kDa, respectively (Fig 4). The primer pairs used to amplify each isoform of Hs.118944 had *BamHI* and *Hind III* restriction sites engineered at their 5' respective ends. A 3' in-frame myc-tag was incorporated into all the reverse primers for recombinant protein detection. Isoforms A, B, C and D were RT-PCR

amplified from total testis RNA, digested with *BamHI* and *Hind III*, and cloned into the retroviral expression vector (Table 2).

The resulting recombinant proteins appeared to be rapidly degraded, as western blot analysis only detected very low levels of protein for short periods of time after expression (data not shown). Immunofluorescence staining also failed to detect the recombinant proteins. The lack of expression was not due to the vector backbone as this same vehicle was used for other different open reading frames and furthermore, mRNA levels were readily detectable in transfected cells (data not shown). Retroviral infection and transient transfection both failed to express the recombinant proteins. As a result of this protein instability, subsequent cell based screens for tumour suppressor function were aborted.

The open reading frame and flanking intronic regions were screened for disease specific mutations in 45 paired breast tumour and normal DNA samples from the same patient. In addition, Hs.118944 was screened in a panel of breast cancer cell lines exhibiting 16q LOH. Fluorescent primers were used to amplify the gene in 7 overlapping amplicons. The primer pairs, corresponding amplicon sizes and locations are summarized in Table 3. No disease specific mutations were identified.

In summary, Hs.118944 exhibits no expression variability in breast cancer cell lines (2.3.3.3), no mutations in breast tumours and appears to be a testis predominant transcript, and therefore was excluded as a candidate cellular senescence gene.

4.3.3) FBXO31. *FBXO31* was originally represented in the NCBI database (3) as the Unigene cluster Hs.7970. A multiple tissue northern blot was probed with a 224 bp RT-PCR product generated from a forward primer spanning exon 2/3 in conjunction with an exon 4 reverse primer. *FBXO31* corresponds to a 3.6 kb transcript expressed in all tissues examined including mammary gland, testis, ovary, uterus, prostate, stomach, bladder, spinal cord, brain, pancreas, and thyroid (Fig 3a). Furthermore, homologous ESTs have been generated from a range of different tissue cDNA libraries, supporting a ubiquitous expression profile.

3612 bp of cDNA sequence has been generated and 9 exons identified (Fig 5a). Exon 1 contained the start codon while exon 7 contained the stop codon and the 3' UTR.

The intron and exon boundaries are presented in Table 4. An alternative spliced isoform was also detected where exon 2 utilized an alternative splice donor site, resulting in an extra 87 bp of coding sequence, maintaining the same open reading frame. *FBXO31* has an open reading frame of 1617 bp encoding a 78 kDa protein (Fig 5). SSCA was performed on a panel of breast cancer cell lines, and on paired breast tumour and normal DNA isolated from the blood lymphocytes of the same patient (Ms Olivia McKenzie, personal communication). The gene was screened in 12 amplicons and the corresponding primer pairs are given in Table 5. No disease specific mutations were identified.

Subsequent *in silico* analysis of the predicted protein sequence identified a putative F-box domain. F-box containing proteins function as ubiquitin protein ligases and the ubiquitin-proteasome pathway is often the target of cancer related deregulation (Spataro *et al*, 1998). F-box proteins are an expanding family of eukaryotic proteins characterized by an amino terminal motif (F-box) of approximately 50 amino acids and are substrate-recognition components of SCF (Skp1, Cullin, F-box protein, Rbx1/Roc1/Hrt1) ubiquitin-protein ligases. Comparison of *FBXO31* with the F-box domain of functionally demonstrated F-box proteins (Fbx1, Fbx2, Fbw1a, Fbw1b and Fbl1) showed *FBXO31* matched the F-box consensus more closely than recognized F-box proteins from each of the three classes (Table 6). Subsequently, co-immunoprecipitation experiments with Skp1 have confirmed that *FBXO31* functions as a F-box protein (Dr Marina Kochetkova, personal communication). The variable carboxy terminal domain of F-box proteins interact in a substrate specific manner with proteins targeted for ubiquitination and subsequent degradation. The ubiquitin-dependant proteasome degradation pathway is an important mechanism for regulating protein abundance in eukaryotes.

Previously we have shown *FBXO31* to display uniform down regulation across a panel of breast cancer cell lines exhibiting 16q LOH, when compared to levels of expression in normal breast tissue (2.3.3.3). Combining this data with the restoration of cellular senescence observed by Reddy *et al* (1999, 2000), it was hypothesized that *FBXO31* functions as the 16q24.3 tumour suppressor gene. Despite *FBXO31* exhibiting no disease specific mutations, it was still considered a candidate as gene inactivation can be achieved by alternative mechanisms such as promoter hypermethylation. The full-

length *FBXO31* transcript was amplified with primer pairs containing 3' restriction sites, directionally cloned into the retroviral expression vector pLNCX2, and used to infect the mammary tumour cell lines MCF-7 and SKBR3. The MCF-7 cell line was chosen as this cell line was used in the original report, showing YAC 792E1 induced cellular senescence (Reddy *et al,* 2000). After selection for two weeks in G418, infected cells were plated on plastic petri dishes or suspended in soft agar for assessment of anchorage-independent growth (Fig 6: Dr Marina Kochetkova performed the growth suppression assays). Data collected from both assays indicated that there was no detectable change in growth characteristics of cells expressing *FBXO31* relative to vector alone. Thus, *FBXO31* was excluded as a candidate cellular senescence gene. This same cell-based assay was used for the initial identification of the putative tumour suppressor *CBFA2T3* (Kochetkova *et al,* 2002) and therefore this construct was included as a positive control. *CBFA2T3* overexpression reduced colony growth on plastic and in soft agar. Escape from senescence in tumour cells has been postulated to be an early step in the evolution of malignant tumours (O'Brien *et al*, 1986; Medina, 1996; Yeager *et al*, 1998).

4.4) Discussion.

YAC 792E1 has been mapped and characterized, which has previously been shown to induce cellular senescence in breast cancer cell lines (Reddy *et al*, 2000). YAC 792E1 was shown to contain two genes, Hs.118944 and *FBXO31,* and subsequent analysis lead to their exclusion as tumour suppressor candidates. Hs.118944 was rejected on the basis of being a testis predominant transcript, exhibiting uniform expression in breast cancer cell lines, and finally, failed to exhibit mutations in breast tumour DNA.

Recombinant Hs.118944 proteins were prone to rapid degradation suggesting that either their endogenous function is tightly regulated, and/or the proteins function at very low physiological concentrations. The lack of expression was not due to the vector backbone as this same vehicle was used for other different open reading frames and furthermore, mRNA levels were readily detectable in transfected cells. As a result of this

protein instability, subsequent cell based screens for tumour suppressor function were aborted.

The *FBXO31* gene exhibited no disease specific mutations but did display reduced expression in breast cancer cell lines. *FBXO31* expression was reduced in eleven cell lines examined, including the non-tumourgenic cell line MCF-12A, when compared to expression levels in normal breast tissue. The known tumour suppressors SYK and CDKN2A, both exhibited large expression variability, with some cell lines exhibiting reduced expression while others exhibiting up regulation. It is therefore speculated that the *FBXO31* generalized down regulation may have been a consequence of cell culture and/or cell immortalization. Future analyses should examine expression levels in other non-tumorigenic cell lines. Retro-viral mediated gene delivery was used to over-express *FBXO31* in the breast cancer cell lines MCF-7 and SKBR3. These cell lines exhibited unchanged growth characteristics, when assessed for anchorage-independent growth on plastic and in soft agar. In conclusion, *FBXO31* failed to exhibit characteristics that would be consistent with a tumour suppressive function.

YAC 792E1 is chimeric and known to contain parts of chromosome 11 (Reddy *et al*, 2000). Our results indicate that the cellular senescence gene is absent from the 16q proportion of YAC 792E1 and future experiments should aim to characterize the chromosome 11 portion of this YAC. Given that chromosome 11 does not restore senescence in MCF-7 or LA7 cells (Negrini *et al*, 1994), it is conceivable that the senescence maybe a consequence of the rearranged YAC expressing unknown chimeric proteins. Accordingly, YAC 792E1 has been shown to exhibit a 16q interstitial deletion of an unknown size.

Functional complementation studies, often in combination with LOH data, have been used to successfully localize cancer related genes to a small chromosomal region, from which a gene associated with cancer has been identified (Seraj *et al*, 2000; Kuramochi *et al*, 2001, Ricketts *et al*, 2002). However there are many cases where an initial localization by functional complementation of a cellular senescence gene, tumour suppressor gene or metastasis suppressor gene, has not resulted in the subsequent identification of the causative gene (Uejima et al, 1995, 1998; Robertson et al, 1996, 1999; Miele et al, 1996, 2000). Although many of these genes may be identified at some

time in the future, the possibility that some of the reported cellular responses occur as a result of the process itself should not be overlooked.

A recent study correlating gene expression profiling in breast cancer with clinical outcome identified *FBXO31* as an important diagnostic marker (van't Veer *et al* 2002). Microarray analysis of sporadic breast tumour RNA revealed variable expression in different tumour types and this variability was associated with an elevated risk of disease metastasis.

Figures

Figure 1: Combined genetic, physical and transcript map of YAC clone 792E1. The YAC location is shown relative to the somatic cell hybrid breakpoint CY2/CY3. BAC and PAC clones are represented as horizontal bars with their corresponding clone name indicated. Dashed lines show STS markers and candidate transcripts are presented above the YAC. Hs.118944 is represented as AK026130

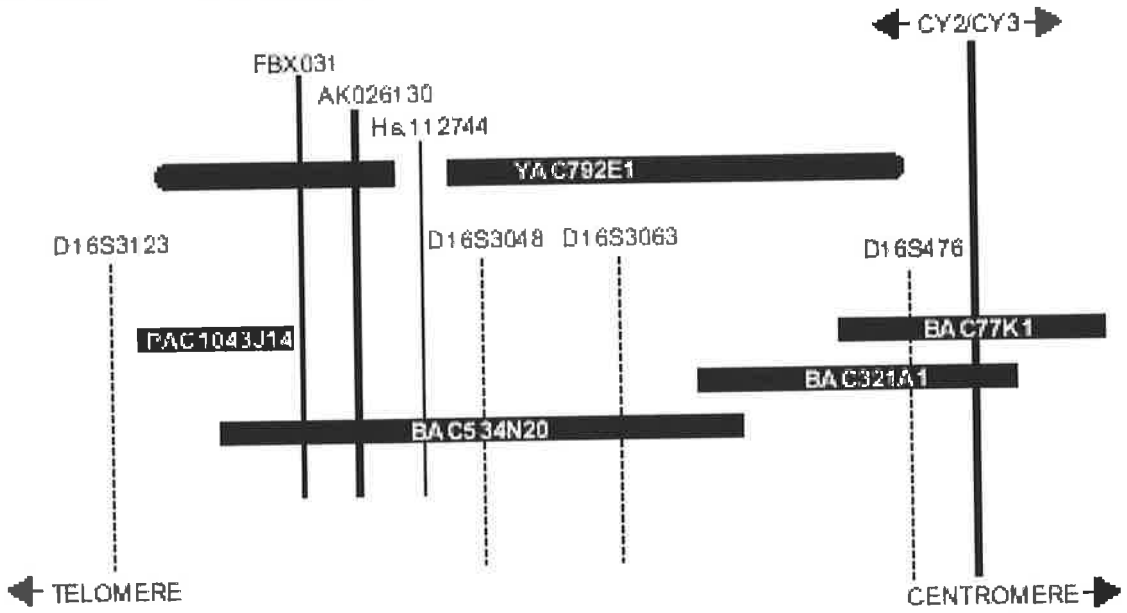
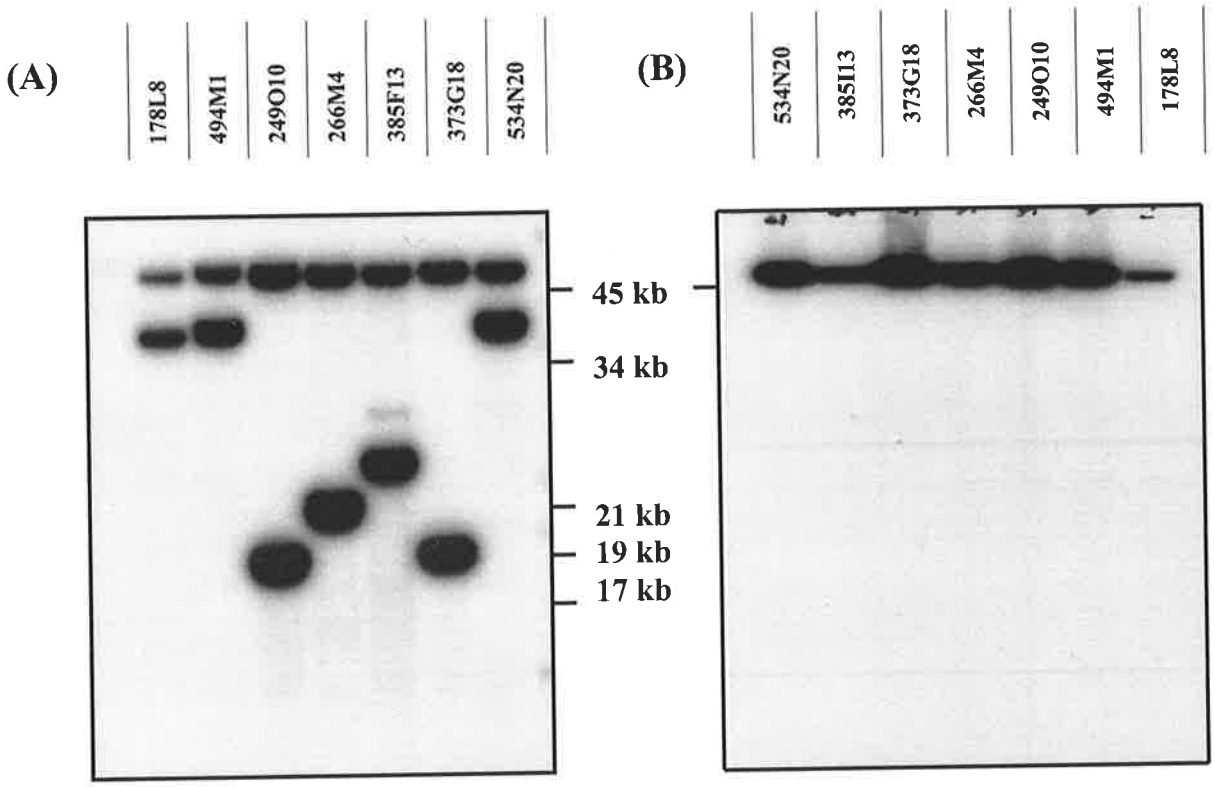


Figure 2: Southern blot analysis of Hs.118944 (A) and *FBXO31* (B) against the PAC and BAC contig at 16q24.3. (A) The 240 bp Hs.119844 RT-PCR product encompassed the 45 kb Eag1 restriction fragment, and several adjacent smaller Eag1 restriction fragments. (B) The 149 bp *FBXO31* RT-PCR product, spanning the exon 2/3 boundary hybridized to the 45 kb Eag1 restriction fragment.



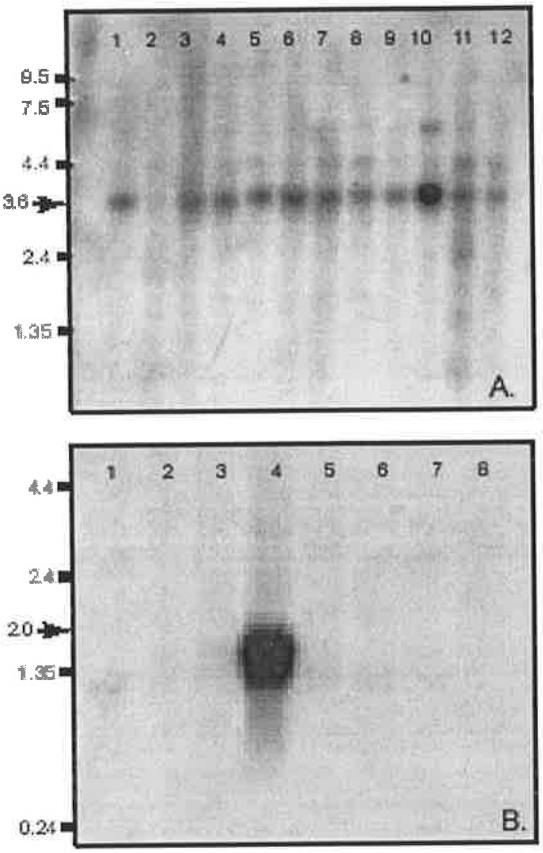


Figure 3: Expression of the candidate transcripts in normal adult tissues. A. Northern blots from (Clontech) containing poly(A)$^+$ mRNA (2µg each lane) were hybridised to : **A)** *FBXO31* cDNA and **B)** Hs.118944 cDNA. Blot A: 1-mammary gland, 2-bone marrow, 3-testis, 4-ovary, 5-uterus, 6-prostate, 7-stomach, 8-bladder, 9-spinal cord, 10 -brain,11-pancreas, 12-thyroid; Blot B: 1-spleen, 2-thymus, 3-prostate, 4-testies, 5-ovary, 6-small intestine, 7-colon, 8-peripheral blood lymphocytes). *FBXO31* corresponds to a 3.5 kb transcript and Hs.118944 corresponds to a 1.5 kb transcript predominantly expressed in testis

Table 1: Splice Sites of the Hs.118944.

Exon	Size (bp)	3' Splice site (intron/exon)	5' Splice site (exon/intron)	Intron size (bp)
1a	103	5'UTR	AGGATGGCAG/**gc**aagtgagt	1267
1b	107	5'UTR	TGGCAGGCAA/**gt**gagtcctg	1263
1c	145	5'UTR	CACCTCGCCT/**gt**gagttccc	1225
2	52	tcttct**ag**/GAATAGCACA	TGGGAAAAAG/**gt**aatgctgc	675
3	126	tcttactc**ag**/ATGCACCCTG	CCGAAAGCAG/**gt**aggtgact	3676
4	147	tctctgtt**ag**/AGTCAAAAGA	TCAGCAGCAG/**gt**aaggagga	816
5	37	tgtcttttt**ag**/ATAGTCAGGA	GGCATCCAAG/**gt**aactccag	4391
6	187	ctttgcct**ag**/CACCCCTGCT	TGGCCATTCG/**gt**gagcaggt	2643
7	79	tgctttcc**ag**/CCAGTGTTTT	3'UTR	

Figure 4: Nucleotide and deduced amino acid sequence of four Hs.118944 isoforms. Isoforms A and C share the same start codon (underlined and bold), as do isoforms B and D (underlined red). Isoforms C and D share the same stop codon (underlined and bold), as do isoforms A and B (underlined red). The first nucleotide of each exon is indicated (underlined/bold). The alternative exon 2 splice donor is highlighted in red

A) Nucleotide sequence.

GGCCACGCGTCGACTAGTACGGGGGGGGGGGGCGACGGGCGGGCGCGCG 50

GGCGCTGTTGGGCTCCTCTGAGCGGTTGAGGTTGGGGCTGGGGTTGAGCT 100

CGGGTTGGGTTCCTCCTGGGGCAGGAGGGAAAGGCGGGAGCCAGGGCGTC 150

AGGGCCTGCGGACG**ATG**TGTGTGTGAAAGGGCGCGTCAGCCATAAGAAGC 200

CAT**ATG**CGTGCGAGCCGGTCCCCACCGTCCCCGCGGCGTTGTCACCATCA 250

TCATGAGGCCACTGGAGCAGCCTCAGGCGCTGCTGCCGGGGGGCCGGGCG 300

CGGGGTGCGTCGGGCTCTGCAGGTTGGCACTCACACCCAGCGCGCAGGAT 350

GGCAG**G**CAA**G**AATAGCACATTTCAAACCTACAAGAAAGAAGTGTGCCTCC 400

CCCGTCATTCG**A**TGCACCCTGGCCCCTGGGCCATCTGCTGTGAATGCCAG 450

ACCAGATTCGGGGGCCGCCTGCCTGTGTCCAGGGTGGAAGCAGCACTGCC 500

TTACTGGGTCCCTCTGTCCCTGAGACCCCGAAAGCAG**A**GTCAAAAGACGG 550

TTCAATTTCCTATCCCCCAAACCGCCAAGATGTGCCCCTGCCTATGCCAC 600

CGCTTTGGGGGCCGCCTCCCGATGCCTAGGGACCAGGCAGTGATGCCCTA 650

CTGGGTGCCCCAGGTCCTGAGGTCTCAGCAGCAG**A**TAGTCAGGAGGCAGC 700

AGAGTTTGAAAGGCATCCAA**C**ACCCCTGCTGGATGCATGCTGCTGGCAC 750

AACTGCTGGCGGATCTGCGGTGATGAGTGCCTGTTGTCCAAGTTCCAGCA 800

GCTCCAGGCCCCCTACCAGGACCAGCTACCGGCTCCTGCAGCGCGTCTGC 850

TGCCCCTCGGCCTCC**TGA**CCCTCCTCCAGGCCATCCCGAGGGTCATCATG 900

GCCATTC**C**CAGTGTTTTGGGGTT**TGA**AGTTGGAATCTTCAGCTACTGTC 950

AAGAACAGCCACAAAAATGTGTCACGATCAAGATCTTTGAGAGTCCACCA 1000

ATCAGGAGGCGTCTGTGACAGTCGCTGTCTTCTCAGAACAGAATCCACAC 1050

CCAGGATTCAACCCAAATGATTTCTCATCAGGTGATTCTTGGTTGTAGCA 1100

AAGTTCATGTGAATGTGGGTGAGTTTCTGTTATGAATGTGGTCAATAAAT 1150

GTTATTTGTGAAACTCTAAACTGGCCACCTCTTAAATGATGACCTCTGAT 1200

TTCCCTTCATTAAGGGGGGACTTGTCCCCATGTCCGTGTCCCTGGGGGAG 1250

GGCAGCTATTTCCGAGTGTCTCTTCAGGGACAGCAGACCCACTGGATGAG 1300

GGCCCGAAAATGGTTTCAGCACCTGTGTCAGGTGGGTCTCCCTTTAAGTA 1350

GCATAACATTGTATATGTGAATTGAATTTTAAAATCTGAGATCTCTTTTA 1400

TATGTCATAAAATTCACTCTTTTAGAGTAAAAAAA 1435

B) Amino acid sequence

Isoform A

```
MRASRSPPSPRRCHHHHEATGAASGAAAGGPGAGCVGLCRLALTPSAQDG     50
RNSTFQTYKKEVCLPRHSMHPGPWAICCECQTRFGGRLPVSRVEAALPYW     100
VPLSLRPRKQSQKTVQFPIPQTAKMCTCLCHRFGARLPMPRDQAVMPYWV     150
PQVLRSQQQIVRRQQSLKGIQAPLLDACCWHNCWRICGDECLLSKFQQLQ     200
APYQDQLPAPAARLLPLGLLTLLQAIPRVIMAIRQCFGV.               239
```

Isoform B

```
MRPLEQPQALLPGGRARGASGSAGWHSHPARRMAGKNSTFQTYKKEVCLP     50
RHSMHPGPWAICCECQTRFGGRLPVSRVEAALPYWVPLSLRPRKQSQKTV     100
QFPIPQTAKMCPCLCHRFGGRLPMPRDQAVMPYWVPQVLRSQQQIVRRQQ     150
SLKGIQAPLLDACCWHNCWRICGDECLLSKFQQLQAPYQDQLPAPAARLL     200
PLGLLTLLQAIPRVIMAIRQCFGV.                              224
```

Isoform C

```
MRASRSPPSPRRCHHHHEATGAASGAAAGGPGAGCVGLCRLALTPSAQDG     50
RNSTFQTYKKEVCLPRHSMHPGPWAICCECQTRFGGRLPVSRVEAALPYW     100
VPLSLRPRKQHPCWMHAAGTTAGGSAVMSACCPSSSSSRPPTRTSYRLLQ     150
RVCCPSAS.                                              158
```

Isoform D

```
MRPLEQPQALLPGGRARGASGSAGWHSHPARRMAGKNSTFQTYKKEVCLP     50
RHSMHPGPWAICCECQTRFGGRLPVSRVEAALPYWVPLSLRPRKQHPCWM     100
HAAGTTAGGSAVMSACCPSSSSSRPPTRTSYRLLQRVCCPSAS.           143
```

Figure 5: Nucleotide (A) and deduced amino acid sequence of *FBXO31 (B)*. Start and stop codons are shown (bold and underlined). The first nucleotide of each exon is indicated (underlined/bold). The F-box domain is in red.

A) *FBXO31* nucleotide sequence

ATGGCGGTGTGTGCTCGCCTTTGCGGCGTGGGCCCGTCGCGCGGATGTCG 50

GCGCCGCCAGCAGCGCCGGGGCCCGGCCGAGACGGCGGCGGCCGACAGCG 100

AGCCGGACACAGACCCCGAGGAGGAGCGCATCGAGGCTAGCGCCGGGGTC 150

GGGGGCGGCTTGTGCGCGGGCCCCTCGCCGCCGCCCCCGCGCTGCTCGCT 200

GCTGGAGCTGCCGCCCGAGCTGCTGGTGGAGATCTTCGCGTCGCTGCCGG 250

GCACGGACCTACCCAGCTTGGCCCAGGTCTGCACGAAGTTCCGGCGCATC 300

CTCCACACCGACACCATCTGGAGGAGGCGTTGCCGTGAGGAGTATGGTGT 350

TTGCGAAAACTTGCGGAAGCTGGAGATCACAGGCGTGTCTTGTCGGGACG 400

TCTATGCGAAGCTGCTTCACCGATATAGACACATTTTGGGATTGTGGCAG 450

CCAGATATCGGGCCATACGGAGGACTGCTGAACGTGGTGGTGGACGGCCT 500

GTTCATCATCGGGTGGATGTACCTGCCTCCCCATGACCCCCACGTCGATG 550

ACCCTATGAGATTCAAGCCTCTGTTCAGGATCCACCTGATGGAGAGGAAG 600

GCTGCCACAGTGGAGTGCATGTACGGCCACAAAGGGCCCCACCACGGCCA 650

CATCCAGATTGTGAAGAAGGATGAGTTCTCCACCAAGTGCAACCAGACGG 700

ACCACCACAGGATGTCCGGCGGGAGGCAGGAGGAGTTTCGGACGTGGCTG 750

AGGGAGGAATGGGGGCGCACGCTGGAGGACATCTTCCACGAGCACATGCA 800

GGAGCTCATCCTGATGAAGTTCATCTACACCAGTCAGTACGACAACTGCC 850

TGACCTACCGCCGCATCTACCTGCCGCCCAGCCGCCCCGACGACCTCATC 900

AAGCCTGGCCTCTTCAAAGGTACCTATGGCAGCCACGGCCTGGAGATTGT 950

GATGCTCAGCTTCCACGGCCGGCGTGCCAGGGGCACCAAGATCACGGGCG 1000

ACCCCAACATCCCCGCTGGGCAGCAGACAGTGGAGATCGACCTGAGGCAT 1050

CGGATCCAGCTGCCCGACCTCGAGAACCAGCGCAACTTCAATGAGCTCTC 1100

CCGCATCGTCCTGGAGGTGCGCGAGAGGGTGCGCCAGGAGCAGCAGGAAG 1150

GCGGGCACGAGGCGGGCGAGGGTCGTGGCCGGCAGGGCCCCCGGGAGTCC 1200

CAGCCAAGCCCTGCCCAGCCCAGGGCAGAGGCGCCCAGCAAGGGCCCAGA 1250

TGGGACACCTGGTGAGGATGGTGGCGAGCCTGGGGATGCCGTAGCTGCGG 1300

CCGAGCAGCCTGCCCAGTGTGGGCAGGGGCAGCCGTTCGTGCTGCCCGTG 1350

GGCGTGAGCTCCAGGAATGAGGACTACCCCCGAACCTGCAGGATGTGTTT 1400

TTATGGCACAGGCCTCATCGCGGGCCACGGCTTCACCAGCCCTGAACGCA 1450

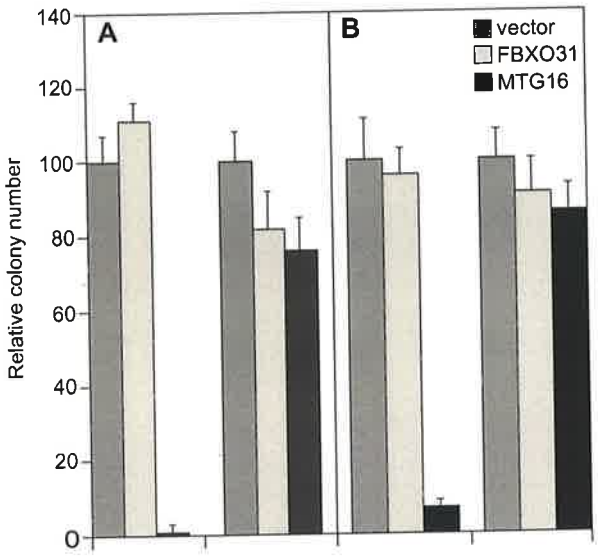
CCCCCGGGGTCTTCATCCTCTTCGATGAGGACCGCTTCGGGTTCGTCTGG 1500

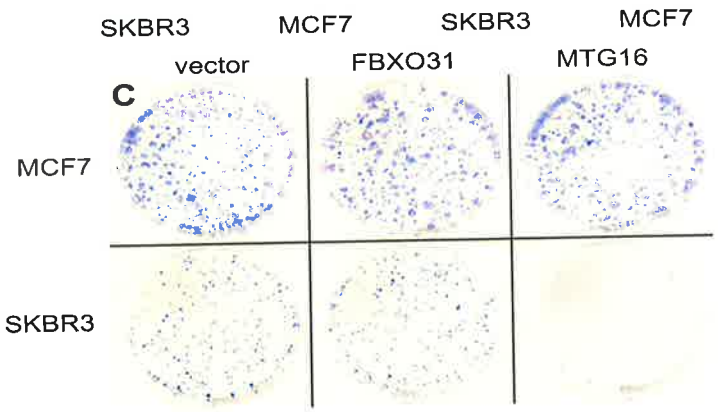
CTGGAGCTGAAATCCTTCAGCCTGTACAGCCGGGTCCAGGCCACCTTCCG 1550
GAACGCAGATGCGCCGTCCCCACAGGCCTTCGATGAGATGCTCAAGAACA 1600
TTCAGTCCCTCACCTCC**TGA**CCGGCCACATCCTTGCCGCCACATCCCGGG 1650
TGGCTCTGGGGCTCTGAACTCTGACCTGTGAATAGAAGCAGCATGCACTT 1700
TGGAAATCCGGCCTTTTGACCAGAACGCACACCTCGTCGGGGGGCCCAGT 1750
CCAGCCACCCCCCAGCACTTTATGTAGAGAGTGTGACATAGACCTGCATA 1800
TTTGTCAGTGCCATGATGGAAGAAGCTGAGCATGTCTTACCAAAAACAGA 1850
GAGAACGAGCCTGAATACAGCAGATGTAGGGGACAGCCGTGGGACCGCGT 1900
GAGAATTGAAGCGGTGGGGTTCCCGCACCCTGGGCTGGCTGGTGGTTTTC 1950
TCGGGAAGCAGGACCCTCCTGACTGGTGCTCTTCCTGTGAGCGGATAGAG 2000
TGATAGACTGGGTCGTGTGTGAGACGCATGTGCTCCACCCCACTCCTTTT 2050
GGGGGAAGCCAGGCAACAGTGGCCTCTGGGAGGGGGTCAGGAAGAGGCGA 2100
ACAGCTCAGGCAGCGCAGGTGTGATGGGCACAGTACGCAGAGCAAGCTCG 2150
GGAAGTTGGTAGGATCTCAGGCTTGGGGCCGGGACTCTGGAGTGAATCCC 2200
CATTTCTCTACCGGCTTGCTTGGAGTTTGGACAGAAGCATTTCACCTCTG 2250
ATCTCAGCTTCCCCACCTGTGGAGTGGGTTTAGTGACCTGAGTCACTAGG 2300
GAATGTCACCTGAATGCACAGCCCAGCCCATGCACCTGCCCCAGCCCCTC 2350
CAGCTTTGGAGCCAAGGCCATCGTTCCAGCCACTTGACTGTCCTCGACGG 2400
CCTGTTCCAGACAGGGCGTTTGTTTTGTCCATGCCTTCCTCCCTGCACGC 2450
ACACGGCGTCAAAACCAAGCTGCCGGCCACTGTCTCCAGAACGCAAGGCT 2500
CCAGGCCCGTGTGTCTGAAGCAGTGAGTGGTCCACACAGGTGCCAGGAGT 2550
GCCCATATGAGATGACGAGGAAACCCCTTTGCAGGTGAGGGGACAGCTTT 2600
CTAGAAAAGCCACACCTGCATCTGGGGACACACTTTGGAAAGTGGGACCC 2650
TCCAGCCTGGAGACCCCATGGACTGATGCCTCCACTGCTGTGTGCCCCAT 2700
GTTGTGTTAACACCTGCGTGTGGGGACCCCATCTGAGGTCTTGGCTGAGG 2750
TTGGCATCTCCTGAAGAACAGAGAGCACGGTGTCCAGAGCTGGCCCTTCC 2800
CCCAGCCCACAGCCAGCTCCGTGCCCGAGTGGGCGTCCCCAGCGAGCCTT 2850
CCCTCTCTGCCGCTTGTCCTTGTGTCTGGGCTGCTCCAAGTCCTTGTGCT 2900
GGGCACCCTGGACACGTCCTGCTGGTGAGGGACCTCGGGAAGGTGACAGT 2950
CTGTGTGCCTTGGTGTGGAGACCAACCTGAGGATGTCCTGGGAAATGTTT 3000
TCCTGATGAATTTCTCCTTGACTGGCCTTTAAAGAACATAAGAATTCCCA 3050
TTGCCCAGCCTCAGTGCATTTGGCAAATGCTTACTTTGCTTCCCAGAGTC 3100
AGAGAATTGGCAAAGGTTCCTAAATGGTAATCTGGCCGGCCTGGGAGAAA 3150
GACTCACGAGAAAAGCCAGTGGAGAAAGCGCCCTTCCAGGGCGGCAGCAG 3200
CGGGAGCCACGCAGACCCCGAGGCGCACCTGCTGGCTCTTGTGTGTGGCC 3250
CCAGTTTCTAGCGGCTTTTGCAGCATTAGCCTACAAGCTTTGTCACTCCC 3300
TGCCCTCTGTGGTGGTCACTGTTTTTCTCTCTTGCCAAATGAGGCAGTCT 3350
CTGAGTGACGGTGACTGTGGCCTTGAAGCCTGGAGGACTGTTGGGCATGT 3400

AGACTGGCACCTTGAAGATTCACCATTGTTTAAATAAAATCAAGCAAATG 3450
CTTTTTTACCAAGAGCCCGAGCCTCGCTCTAAGGGACGCAGTCCTAGAGG 3500
CGTGCCCTTTGGGGCTTGAAGAGCACACTGTGGGACGCACGTGCTTCTGA 3550
TTAAAGGAATCTCAGATCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3600
AAAAAAAAAAAA 3612

B) *FBXO31* amino acid sequence

MAVCARLCGVGPSRGCRRRQQRRGPAETAAADSEPDTDPEEERIEASAGV 50
GGGLCAGPSPPPPRCSLLELPPELLVEIFASLPGTDLPSLAQVCTKFRRI 100
LHTDTIWRRRCREEYGVCENLRKLEITGVSCRDVYAKLLHRYRHILGLWQ 150
PDIGPYGGLLNVVVDGLFIIGWMYLPPHDPHVDDPMRFKPLFRIHLMERK 200
AATVECMYGHKGPHHGHIQIVKKDEFSTKCNQTDHHRMSGGRQEEFRTWL 250
REEWGRTLEDIFHEHMQELILMKFIYTSQYDNCLTYRRIYLPPSRPDDLI 300
KPGLFKGTYGSHGLEIVMLSFHGRRARGTKITGDPNIPAGQQTVEIDLRH 350
RIQLPDLENQRNFNELSRIVLEVRERVRQEQQEGGHEAGEGRGRQGPRES 400
QPSPAQPRAEAPSKGPDGTPGEDGGEPGDAVAAAEQPAQCGQGQPFVLPV 450
GVSSRNEDYPRTCRMCFYGTGLIAGHGFTSPERTPGVFILFDEDRFGFVW 500
LELKSFSLYSRVQATFRNADAPSPQAFDEMLKNIQSLTS. 539





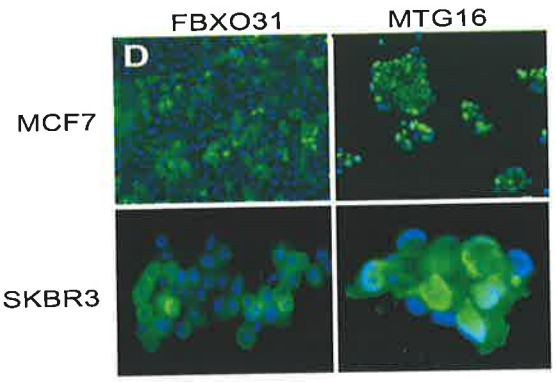


Figure 6: The coding region of the FBXO31 cDNA was amplified and inserted into pLNCX2. Retroviral infections and cell growth suppression assays were as previously described in Kochetkova et al (2002). Briefly, MCF7 and SK-BR-3 breast cancer cell lines were infected with recombinant retroviruses expressing Myc-tagged FBXO31, Myc-tagged CBFA2T3 as a positive control and Neo only as a negative control. Infected cells were assessed in the presence of G418 for changes in cell morphology and colony growth on plastic and in soft agar. A. Cell colonies were grown for two weeks, fixed, stained with Giemsa and counted. B. Infected cells were suspended in soft agar and colonies greater than 50 cells were scored after three weeks. C. Representative plates for MCF7 and SK-BR-3 cell lines infected with FBXO31, CBFA2T3 and empty vector. D. Cells expressing myc-tagged proteins were visualized using Myc monoclonal antibody and fluorescent microscopy. Green fluorescence signal (FITC) indicate infected cells, blue fluorescence signal (DAPI) - uninfected cells. The results shown are representative of at least three independent experiments and the values are mean and SD of triplicate samples.

Tables

***Table 2*: Oligonucleotide primers for *FBXO31* and Hs.118944 isoform amplification.**

CONSTRUCT	PRIMER SET	ORF size
Hs.118944.A/pLNCX2	F/BamH1 5' *TAA TAT* GGA TCC CAT ***ATG*** CGT GCG AGC CGG	720 bp
	R/HindIII 5' *TAA TAA GCT T**TC** **ACA GCA GGT CCT CCT CGC TGA TCA GCT TCT GCT** CAA CCC CAA AAC ACT GGC GAA TGG C	
Hs.118944.B/pLNCX2 (+4 bp)	F/BamH1 5' *AGT TAT* GGA TCC ATC ***ATG*** AGG CCA CTG GAG	675 bp
	R/HindIII 5' *TAA TAA GCT T**TC** **ACA GCA GGT CCT CCT CGC TGA TCA GCT TCT GCT** CAA CCC CAA AAC ACT GGC GAA TGG C	
Hs.118944.C/pLNCX2 (no exons 4 and 5)	F/BamH1 5' *TAA TAT* GGA TCC CAT ***ATG*** CGT GCG AGC CGG	477 bp
	R/HindIII 5' *TAA TAA GCT T**TC** **ACA GCA GGT CCT CCT CGC TGA TCA GCT TCT GCT** CGG AGG CCG AGG GGC AGC AGA CGC G	
Hs.118944.D/pLNCX2 (+4 bp no exons 4/5)	F/BamH1 5' *AGT TAT* GGA TCC ATC ***ATG*** AGG CCA CTG GAG	432 bp
	R/HindIII 5' *TAA TAA GCT T**TC** **ACA GCA GGT CCT CCT CGC TGA TCA GCT TCT GCT** CGG AGG CCG AGG GGC AGC AGA CGC G	
***FBXO31*/pLNCX2** (-87 bp)	F/BamH1 5' *AGT TAT* GGA TCC ATC ***ATG*** GCG GTG TGT GCT	1617 bp
	R/HindIII 5' *TAA TAA GCT T**TC** **ACA GCA GGT CCT CCT CGC TGA TCA GCT TCT GCT** CCC GGT CAG GAG GTG AGG GAC TGA A	

Primer Sequences		
	Bold	myc tag
	Italic	buffer sequence for efficient restriction
	Underlined	restriction site
	Normal	coding region
	Bold/*Italic*	in-frame start or stop codons

***Table 3*: Oligonucleotide primers for Mutation analysis of Hs.118944.**

Primer Set	Nucleotide sequences (5'-3')	Size (bp)
Exon 1	GGC GCG TCA GCC ATA AGA AG ACA GGC GAG GTG TCG GGT AG	224
Exon 2	TAT TAA GAG TCA GTT CTG AGA AG CAG GGC CAG TGG CTG AGG A	186
Exon 3	GTG CTT GAG AGC CTC CTG TCT CTG AGG GGT GAC TGG TCA ACT	270
Exon 4	TCC ATC CCC TTC TGA ATC TTC T CCA CCT CAC TCT TCA GTT CTC	235
Exon 5	CCA GGC TGA CTG CTT GCC T GTG ACC AGG GAT GGC TGT TG	241
Exon 6	GGA GAA CAC GCC CTG TCC TC CCC TCA GGT GGA CGT CAC CA	289
Exon 7	CCT AGT AAC TGT CTC TCT TT TCC CCC AGG GAC ACG GAC AT	370

***Table 4*: Splice Sites of the *FBXO31*.**

Exon	Size (bp)	3' Splice site (intron/exon)	5' Splice site (exon/intron)	Intron size (bp)
1	343	5'UTR	TGCCGTGAGG/**gt**gagcgcgc	23042
2	72	cttgttac**ag**/AGTATGGTGT	TATGCGAAGC/**gt**gagtgaat	1797
2.5	87	gtctgttc**ag**/GTATAAACCC	TACACCTGCC/**gt**atgtacct	11160
3	77	cctcctgt**ag**/TGCTTCACCG	GAACGTGGTG/**gt**aagtcccg	3408
4	168	cctcctgt**ag**/GTGGACGGCC	CCACATCCAG/**gt**gtgtgcag	646
5	75	aacactga**ag**/ATTGTGAAGA	GAGGCAGGAG/**gt**gagcccac	6612
6	110	cttttgga**ag**/GAGTTTCGGA	GTCAGTACGA/**gt**gagtgcgg	697
7	154	ctccccac**ag**/CAACTGCCTG	CAAGATCACG/**gt**gagtggcg	1017
8	401	tgctccac**ag**/GGCGACCCCA	GCAGGATGTG/**gt**aaggatg	2375
9	2174	ttctgctc**ag**/TTTTTATGGC	3'UTR	

Table 5: Composite alignment of mammalian F-box proteins. Amino acid sequences of the novel F-box motifs from the three sub classes: Fbx (FBXO31, Cyclin F and NFB42), Fbl (Skp2), and Fbw (β-TRCP1 and β -TRCP2). Identical amino acids are shown in grey and underlined.

```
Consensus          LP-----EILL-IF-YL-----L-----VCK-W--V-------LW
hFbxo31            LPP----ELLVEIFASLPGT--LPSLAQVCTKFRRILH--TDTIW
hFbx1 (Cyclin F)   LPE----DVLFHILKWLS-VEDILAVRAVHSQLKDLVDN-HASVW
HFbx2 (NFB42)      LPE----DVLFHILKWLSAAELVQACRLVCLRWKELVD--CAPLW
hFbl1 (Skp2)       LPD----ELLLCIFSCL-CLPELLKVSCVCKRWYRLAS--DESLW
hFbw1a (B-TRCP1)   LPARCLDHIAENILSYLC-AKSLCAAELVCKEWYRVTS--DCMLW
hFbw1b (B-TRCP2)   LPEQCLDHIAENILSYLD-ARSLCAAELVCKEWQRVIS--ECMLW
```

***Table 6*: Oligonucleotide primers for Mutation analysis of *FBXO31*.**

Primer Set	Nucleotide sequences (5'-3')	Size (bp)
Exon 1a	GCG CTG GAG CGT GCG CAC A AGC TCG GGC GGC AGC TCC A	269
Exon 1b	GGT CGG GGG CGG CTT GTG GCC TCC ACC TGG CAG GGA	252
Exon 2	CTG TCG CGT TAT GAG TTG TTG GTA CAA AGT TAA TCA TGG ATG GT	168
Exon 2.5	AGG CAT TGG GTC GTA TTC AC AGA AGC CAA AGC TCG CAG GA	198
Exon 3	GGC ACG CTG GGT CTA ACA C CCT GCC CGT GCA CAG ACC T	167
Exon 4	CTC ATG GAC CTT TGC CCA TCT GTC TGC AGC TGA GAA TAG CAC	290
Exon 5	GTG ATG GAC TCT GTT CCT CAC AGG TCC GCA CCA TAT GAA CAC	170
Exon 6	CAC AGC CTC CTG TCA TAT GGA ACC CCA GCA CCG AGC AGG A	187
Exon 7	GGC GTT CTC AGT CCT GCC T CCC TGA CTC CAC AGC CCA C	284
Exon 8a	CTG GCC TGA GCC CTG CTG A ACC CTC TCG CGC ACC TCC A	171
Exon 8b	CAA TGA GCT CTC CCG CAT C CCA TGC TGT CCC ACC TTC A	354
Exon 9	AGA ATG CTG TAC GTG GCG TG AGG AGG TGA GGG ACT GAA TG	292

Conclusion

Conclusion

In conclusion, the construction of a sequenced physical map has permitted the identification of 104 candidate genes from the 16q24.3 breast cancer LOH interval. Seven of theses genes have been previously screened for mutation in breast tumour DNA: SPG7, BBC1, CPNE7, CDK10, FANCA, GAS11, and C16ORF3. None of these genes were found to harbor any mutations in sporadic breast cancer DNA samples. Quantitative mRNA expression analysis in breast cancer cell lines served as an initial screen for potential tumour suppressors. Of the 104 genes studied, 3 exhibited considerable expression variability across the panel of breast cancer cell lines, *CBFA2T3, CYBA* and Hs.7970. The notion of these genes being breast cancer tumour suppressors is compatible with what is known about their respective function.

Additional genes presented as candidates based on functional evidence. For example, microcell mediated chromosome transfer experiments identified a 16q24 YAC clone capable of inducing cellular senescence in breast cancer cell lines (Reddy *et al*, 2000). Subsequent experiments examined these genes in detail for potential tumour suppressor function.

Additional genes presented as candidates based on their homologies to other known human or mouse proteins. For example, TSG16 exhibited significant homology to the ankyrin (ANK) repeat motif of the BRCA1-associated RING domain protein 1 (BARD1). Functional analyses were performed to determine the biological function of TSG16. A homologous gene on chromosome 18, termed TSG18, was also identified and characterized, and together with TSG16, form a novel protein family.

References

REFERENCES

1. Aaronson, D. S. and Horvath, C. M. (2002). A road map for those who don't know JAK-STAT. *Science.* **296**: 1653-1655.

2. 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**: 3389-402.

3. Armes, J. E. and Venter, D. J. (2002). The pathology of inherited breast cancer. *Pathology.* **34**(4): 309-14.

4. Arnold, R. S., Shi, J., Murad, E., Whalen, A. M., Sun, C. Q., Polavarapu, R., Parthasarathy, S., Petros, J. A. and Lambeth, J. D. (2001). Hydrogen peroxide mediates the cell growth and transformation caused by the mitogenic oxidase Nox1. *Proc Natl Acad Sci U S A.* **98**: 5550-5.

5. Aubele, M., Cummings, M., Walsch, A., Zitzelsberger, H., Nahrig, J., Hofler, H. and Werner M. (2000). Heterogeneous chromosomal aberrations in intraductal breast lesions adjacent to invasive carcinoma. *Anal Cell Pathol.* **20**(1): 17-24.

6. Baeuerle, P. A. and Baltimore, D. (1996). NF-kB: Ten years after. *Cell.* **87**: 13-20.

7. Baylin, S. B., Herman, J. G., Graff, J. R., Vertino, P. M. and Issa, J. P. (1998). Alterations in DNA methylation: a fundamental aspect of neoplasis. *Adv Can Res.* **72**: 141-96.

8. Bertram, M. J., Berube, N. G., Hang-Swanson, X., Ran, Q., Leung, J. K., Bryce, S., Spurgers, K., Bick, R. J., Baldini, A., Ning, Y., Clark, L. J., Parkinson, E. K., Barrett, J. C., Smith, J. R. and Pereira-Smith, O. M. (1999). Identification of a gene that

reverses the immortal phenotype of a subset of cells and is a member of a novel family of transcription factor-like genes. *Mol Cell Biol.* **19**(2): 1479-1485.

9. Bertwistle, D. and Ashworth, A. (1998). Functions of the BRCA1 and BRCA2 genes. *Curr Opin Genet Dev.* **8**(1): 14-20.

10. Berx, G., Cletonjansen, A. M., Nollet, F., Deleeuw, W. J. F., Vandevijver, M. J., Cornelisse, C. and Vanroy, F. (1995). E-cadherin is a tumour invasion suppressor gene in human lobular breast cancers. *EMBO Journal.* **14**(24): 6107-15.

11. Birot, A., Duret, L., Bartholin, L., Santalucia, B., Tigaud, I., Magaud, J. and Rouault, J. (2000). Identification and molecular analysis of *BANP*. *Gene.* **253**: 189-196.

12. Bisogna, M., Calvano, J. E., Ho, G. H., Orlow, I., Cordon-Cardo, C., Borgen, P. I. and Van Zee, K. J. (2001). Molecular analysis of the INK4A and INK4B gene loci in human breast cancer cell lines and primary carcinomas. *Cancer Genet Cytogenet.***125**: 131-8.

13. Borden, K. L. (2000). RING domains: master builders of molecular scaffolds. *J Mol Biol.* **295**(5): 1103-12.

14. Bos, J. L. (1989). ras oncogenes in human cancer: a review. *Cancer Res.* **49**(17): 4682-9.

15. Bowcock, A. M. (1999). Breast Cancer: Molecular Genetics, Pathogenesis, and Therapeutics, Humana Press, New Jersey.

16. Brehm, A., Miska, E. A., McCance, D. J., Reid, J. L., Bannister, A. J. and Kouzarides, T. (1998). Retinoblastoma protein recruits histone deacetylase to repress transcription. *Nature.* **391**(6667): 597-601.

17. Brenner, A. J. and Aldaz, C. M. (1997). The genetics of sporadic breast cancer. *Prog Clin Biol Res.* **396**: 63-82.

18. Bromberg, J. (2002). Stat proteins and oncogenesis. *J Clin Invest.* **109**: 1143-48.

19. Brown, N. S. and Bicknell, R. (2001). Hypoxia and oxidative stress in breast cancer. Oxidative stress: its effects on the growth, metastatic potential and response to therapy of breast cancer. *Breast Cancer Res.* **3**: 323-7.

20. Brzovic, P. S., Meza, J. E., King, M. and Klevi, R. E. (2001). BRCA1 RING domain cancer-predisposing mutations: structural consequences and effects of protein-protein interactions. *J Biol Chem.* **276**(18): 14537-40.

21. Burdon, R. H., Gill, V. and Alliangana, D. (1996). Hydrogen peroxide in relation to proliferation and apoptosis in BHK-21 hamster fibroblasts. *Free Radic Res.* **24**: 81-93.

22. Bustin, S.A. (2000). Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J Mol Endocrinol.* **25**: 169-93.

23. Callen, D. F., Crawford, J., Derwas, C., Clenton-Jansen, A. M., Cornelisse, C. J. and Baker, E. (2002). Defining regions of loss of heterozygosity of 16q in breast cancer cell lines. *Cancer Genet Cytogenet.* **133**: 76-82.

24. 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**: 503-11.

25. Callen, D. F., Hildebrand, C. E. and Reeders, S. (1992). Report of the Second International Workshop on Human Chromosome 16 Mapping. *Cytogenet Cell Genet.* **60**(3-4): 158-67.

26. Carrera, C. and Payne S. (1999). Ductal carcinoma *in situ* (DCIS) of the breast: the need for psychosocial research. *Psycho-oncology.* **8**(6): 538-45.

27. Carrier, F., Georgel, P. T., Pourquier, P., Blake, M., Kontny, H. U., Antinore, M. J., Gariboldi, M., Myers, T. G., Weinstein, J. N., Pommier, Y. and Fornace, A. J. Jr. (1999). Gadd45, a p53-responsive stress protein, modifies DNA accessibility on damaged chromatin. *Mol Cell Biol.* **19**(3): 1673-85.

28. 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**(5937): 779-84.

29. Chehab, N. H., Malikzay, A., Appel, M. and Halazonetis, T. D. (2000). Chk2/hCds1 functions as a DNA damage checkpoint in G(1) by stabilizing p53. *Genes Dev.* **14**(3): 278-88.

30. 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 tumour suppressor gene region. *Cancer Res.* **56**: 5605-9.

31. Chen, Y., Chen, C. F., Riley, D. J., Allred, D. C., Chen, P. L., Von Hoff, D., Osborne, C. K. and Lee, W. H. (1995). Aberrant subcellular localization of BRCA1 in breast cancer. *Science.* **270**(5237): 789-91.

32. Chen, Y., Lee, W. and Chew, H. K. (1999). Emerging roles of BRCA1 in transcriptional regulation and DNA repair. *J Cell Physiol.* **181**: 385-92.

33. Chenevix-Trench, G., Spurdle, A. B., Gatei, M., Kelly, H., Marsh, A., Chen, X., Donn, K., Cummings, M., Nyholt, D., Jenkins, M. A., Scott, C., Pupo, G. M., Dork, T., Bendix, R., Kirk, J., Tucker, K., McCredie, M. R., Hopper, J. L., Sambrook, J., Mann, G. J. and Khanna, K. K. (2002). Dominant negative ATM mutations in breast cancer families. *J Natl Cancer Inst.* **94**(3): 205-15.

34. Cho, Y., Gorina, S., Jeffrey, P. D. and Pavletich, N. P. (1994). Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. *Science.* **265**(5170): 346-55.

35. Chomczynski, P. and Sacchi, N. (1987). Single-step method of RNA isolation by acid quanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem.* **162**(1): 156-9.

36. Chou, Y. H., Chung, K. C., Jeng, L. B., Chen, T. C. and Liaw, Y. F. (1998). Frequent allelic loss on chromosomes 4q and 16q associated with human hepatocellular carcinoma in Taiwan. *Cancer Lett.* **123**: 1-6.

37. Chung, C. D., Liao, J., Lui, B., Roa, X., Jay, Berta, P. and Shuai, K. (1997). Specific inhibition of STAT3 signal transduction by PIAS3. *Science.* **278**: 1803-1805.

38. Cleton-Jansen, A. M., Callen, D. F., Seshadri, R., Goldup, S., Mccallum, B., Crawford, J., Powell J. A., Settasatian, C., van Beerendonk, H., Moerland, E. W., Smit, V. T., Harris, W. H., Millis, R., Morgan, N. V., Barnes, D., Mathew, C. G. and Cornelisse, C. J. (2001). Loss of heterozygosity mapping at chromosome arm 16q in 712 breast tumours reveals factors that influence delineation of candidate regions. *Cancer Res.* **61**: 1171-7.

39. Cleton-Jansen, A. M., Moerland, E. W., Pronk, J. C., van Berkel, C. G., 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 anaemia A gene in breast tumours with loss of heterozygosity at 16q24.3. *Br J Cancer.* **79**: 1049-1052.

40. Cleton-Jansen, A. M. (2002). E-cadherin and loss of heterozygosity at chromosome 16q in breast carcinogenesis. Different genetic pathways to ductal and lobular breast cancer? *Breast Cancer Res.* **4**(1): 5-8.

41. Coopman, P. J., Do, M. T., Barth, M., Bowden, E. T., Hayes, A. J., Basyuk, E., Blancato, J. K., Vezza, P. R., McLeskey, S. W., Mangeat, P. H. and Mueller, S. C. (2000). The SYK tyrosine kinase suppresses malignant growth of human breast cancer cells. *Nature.* **406**: 742-7.

42. Crawford, J., Ianzano, L., Savino, M., Whitmore, S., 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 PISSLRE gene: structure, exon skipping, and exclusion as tumour suppressor in breast cancer. *Genomics.* **56**: 90-7.

43. Cressman, V. L., Backlund, D. C., Hicks, E. M., Gowen, L. C., Godfrey V. and Koller, B. H. (1999). Mammary tumor formation in p53- and BRCA1-deficient mice. *Cell Growth & Differentiation.* **10**(1): 1-10.

44. Crouch, F. and Weber, B. L. (1996). Mutations and polymorphisms in the familial early-onset breast cancer gene. *Hum Mutat.* **8**: 8-18.

45. Darnell, J. E. Jr. (1997). STATs and gene regulation. *Science.* **277**: 1630-5.

46. 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-Kb/Rel and nuclear co-regulators. *Oncogene.* **18**: 3316-23.

47. Demeter, J. G., Waterman, N. G. and Verdi, G. D. (1990). Familial male breast carcinoma. *Cancer.* **65**(10): 2342-53.

48. Desterro, J. M., Rodriguez, M. S., Hay, R. T. (1998). SUMO-1 modification of IkappaBalpha inhibits NF-kappaB activation. *Mol Cell.* **2**: 233-39.

49. Devilee, P. and Cornelisse, C. J. (1994). Somatic genetic changes in human breast cancer. *Biochimica et Biophysica Acta.* **1198**: 113-30.

50. Devilee, P., Cleton-Jansen, A. M. and Cornelisse, C. J. (2001). Ever since Knudson. *Trends Genet.* **17**(10): 569-573.

51. Devilee, P., van den Broek, M., Mannens, M., Slater, R., Cornelisse, C. L., Westerveld, A. and Khan, P. M. (1991). Differences of patterns of allelic loss between two common types of adult cancer, breast and colon carcinoma, and Wilm's tumour of childhood. *Int J Cancer.* **47**: 817-21.

52. DiMario, J. X. (2002). Activation and repression of growth factor receptor gene transcription. *Int J Mol Med.* **10**(1): 65-71.

53. Doucas V. (2000). The promyelocytic (PML) nuclear compartment and transcription control. *Biochem Pharmacol.* **60**(8): 1197-201.

54. Downward, J. (2003). Targeting RAS signalling pathways in cancer therapy. *Nat Rev Cancer.* **3**(1): 11-22.

55. Easton D. F., Bishop, D. T., Ford, D. and Crockford, G. P. (1993). Genetic linkage analysis in familial breast and ovarian cancer: results from 214 families. The Breast Cancer Linkage Consortium. *Am J Hum Genet.* 52(4): 678-701.

56. Ewing, B and Green, P. (2000). Analysis of expressed sequence tags indicates 35,000 human genes. *Nat Genet.* **25**(2): 23223-4.

57. Ewing, B. and Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**: 186-94.

58. Ewing, B., Hillier, L., Wendl, M.C., and Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* **8**: 175-85.

59. Fearon, E. R. and Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. *Cell.* **61**(5) :759-67.

60. Fleming, M. A., Potter, J. D., Ramirez, C. J., Ostrander, G. K. and Ostrander, E. A. (2003). Understanding missense mutations in the BRCA1 gene: an evolutionary approach *Proc Natl Acad Sci U S A.* **100**(3): 1151-6.

61. Freemont, P. S. (2000). RING of destruction. *Curr Biol.* **10**(2): 84-97.

62. Futreal, P. A., Liu, Q., Shattuck-Eidens, D., Cochran, C., Harshman, K., Tavtigian, S., Bennett, L. M., Haugen-Strano, A., Swensen, J. and Miki, Y. (1994). BRCA1 mutations in primary breast and ovarian carcinomas. *Science.* **266**(5182): 120-2.

63. Gamou, T., Kitamura, E., Hosoda, F., Shimizu, K., Shinohara, K., Hayashi, Y., Nagase, T., Yokoyama, Y. and Ohki M. (1998). The partner gene of AML1 in t(16;21) myeloid malignancies is a novel member of the CBFA2T1(ETO) family. *Blood.* **91**: 4028-37.

64. Gatei, M., Shkedy, D., Khanna, K. K., Uziel, T., Shiloh, Y., Pandita, T. K., Lavin, M. F. and Rotman, G. (2001). Ataxia-telangiectasia: chronic activation of damage-responsive functions is reduced by alpha-lipoic acid. *Oncogene.* **20**(3): 289-94.

65. 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**: 146-52.

66. Giles, R., D. F. Callen, G. R. Sutherland and M.-C. King (1992). Preliminary test of linkage of chromosome 16q markers to breast cancer in families with older onset disease. abstract in "Report of the Second International Workshop on Human Chromosome 16 Mapping". *Cytogenet Cell Genet* **60**: 174.

67. Gong, G., DeVries, S., Chew, K. L., Cha, I., Ljung, B. M. and Waldman, F. M. (2001). Genetic changes in paired atypical and usual ductal hyperplasia of the breast by comparative genomic hybridization. *Clin Cancer Res.* **7**: 2410-4.

68. Gonzalez, I. L., Sylvester, J. E. (1997). Incognito rRNA and rDNA in databases and libraries. *Genome Res.* **7**: 65-70.

69. Gorgoulis, V. G., Koutroumbi, E. N., Kotsinas, A., Zacharatos, P., Markopoulos, C., Giannikos, L., Kyriakou, V., Voulgaris, Z., Gogas, I. and Kittas, C. (1998). Alterations of p16-pRb pathway and chromosome locus 9p21-22 in sporadic invasive breast carcinomas. *Mol Med.* **4**(12): 807-22.

70. Greene, A. S. (2002). Application of physiological genomics to the microcirculation. *Microcirculation.* **9**(1): 3-12.

71. Griendling. K. K. and Ushio-Fukai, M. (2000). Reactive oxygen species as mediators of angiotensin II signaling. *Regulatory Peptides.* **91**: 21-7.

72. Guiford, P. (1999). E-Cadherin downregulation in cancer: fuel on the fire. *Mol Med Today.* **5**(4): 172-7.

73. Gupta, A., Butts, B., Kwei, K. A., Dvorakova, K., Stratton, S. P., Briehl, M. M. and Bowden, G. T. (2001). Attenuation of catalase activity in the malignant phenotype plays a functional role in an in vitro model for tumour progression. *Cancer Lett.* **173**: 115-25.

74. Hakem, R., de la Pompa, J. L. and Mak, T. W. (1998). Developmental studies of Brca1 and Brca2 knock-out mice. *J Mammary Gland Biol Neoplasia.* **3**(4): 431-45.

75. Hall, J. M., 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**: 1684-9.

76. Harbour, J. W. and Dean, D. C. (2000). Rb function in cell-cycle regulation and apoptosis. *Nat Cell Biol.* 2(4): E65-7.

77. Harris, H., Miller, O. J., Klein, G., Worst, P. and Tachibana, T. (1969). Suppression of malignancy by cell fusion. *Nature.* **223**(204): 363-8.

78. Hashizume, R., Fukuda, M., Maeda, I., Nishikawa, H., Oyake, D., Yabuki, Y., Ogata, H. and Ohta, T. (2001). The RING heterodimer BRCA1-BARD1 is a ubiquitin ligase inactivated by a breast cancer derivied mutation. *J Biol Chem.* **276**(18): 14537-40.

79. Hedenfalk, I. A., Ringner, M., Trent, J. M. and Borg, A. (2002). Gene expression in inherited breast cancer. *Adv Cancer Res.* **84**: 1-34.

80. Hildebrand, D., Tiefenbach, J., Heinzel, T., Grez, M. and Maurer, A.B. (2001). Multiple regions of ETO cooperate in transcriptional repression. *J Biol Chem.* **276**: 9889-95.

81. Houvras, Y., Benezra, M., Zhang, H. B., Manfredi, J. J., Weber, B. L. and Licht J. D. (2000). BRCA1 physically and functionally interacts with ATF1. *J Biol Chem.* **275**(46): 36230-7.

82. Huiping, C., Sigurgeirsdottir, J. R., Jonasson, J. G., Eiriksdottir, G., Johannsdottir, J. T., Egilsson, V. and Ingvarsson, S. (1999). Chromosome alterations and E-cadherin gene mutations in human lobular breast cancer *Br J Cancer.* **81**(7): 1103-10.

83. Hupp, T. R. (1999). Regulation of p53 protein function through alterations in protein-folding pathways. *Cell Mol Life Sci.* **55**: 88-95.

84. Ingvarsson, S. (1999). Molecular genetics of breast cancer progression. *Semin Cancer Biol.* 9(4): 277-88.

85. International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. *Nature.* **409**: 860-921.

86. Jacobson, M. D. (1996). Reactive oxygen species and programmed cell death. *Trends Biochem Sci.* **21**: 83-6.

87. 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 BRAC1 protteins in discrete nuclear domains. *Proc Natl Acad Sci USA.* **94**(22): 12075-80.

88. Katsanis, N., Worley, K. C. and Lupski, J. R. (2001). An evaluation of the draft human genome sequence. *Nat Genet.* **29**: 88-91.

89. Khanna, K. K. and Jackson, S. P. (2001). DNA double-strand breaks: signalling, repair and the cancer connection. *Nat Genet.* **27**(3): 247-254.

90. Kim, D. W., Sovak, M. A., Zanieski, G., Nont, G., Romieu-Mourez, R., Lau, A. W., Hafer, L. J., Yaswen, P., Stampfer, M., Rogers, A. E., Russo, J. and Sonesshein, G. E. (2000). Activation of NF-kB/Rel occurs early during neoplastic transformation of mammary cells. *Carcinogenesis.* **21**(5): 871-9.

91. Kirschbaum, M. H. and Yarden, Y. (2000). The ErbB/HER family of receptor tyrosine kinases: A potential target for chemoprevention of epithelial neoplasms. *J Cell Biochem.* **34**: 52-60.

92. Kittiniyom, K., Gorse, K. M., Dalbegue, F., Lichy, J. H., Taubenburger, J. K. and Newsham, I. F. (2001). Allelic loss on chromosome band 18p11.3 occurs early and reveals heterogeneity in breast cancer progression. *Breast Cancer Res.* **3**(3): 192-8.

93. Kleiman, F. E. and Manley, J. L. (1999). Functional interaction f BRCA1-associated BARD1 with Polyadenylation Factor Cst-50. *Science.* **285**: 1576-9.

94. Knudson, A. G. Jr. (1971). Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A.* **68**(4): 820-3.

95. Knudson, A. G. Jr. and Strong, L. C. (1972). Mutation and cancer: neuroblastoma and pheochromocytoma. *Am J Hum Genet.* **24**(5): 514-32.

96. Kochetkova, M., McKenzie, O. L. D., Bais, A. J., Martin, J. M., Secker, G. A., Seshadri, R., Powell, J. A., Hinze, S. J., Gardner, A. E., Spendlove, H. E., O'Callaghan, N. J., Cleton-Jansen, A., Cornelisse, C., Whitmore, S. A., Crawford, J., Kremmidiotis, G., Sutherland, G. R. and Callen, D. F. (2002). CBFA2T3 (MTG16) is a putative breast tumour suppressor from the breast cancer loss of heterozygosity region at 16q24.3. *Cancer Res.* **62**: 4599-604.

97. Kotaja, N., Karvonen, U., Janne, O. A. and Palvimo, J. J. (2002). PIAS proteins modulate transcription factors by functioning as SUMO-1 ligases. *Mol Cell Biol.* **22**(14): 5222-34.

98. Kozu, T., Miyoshi. H., Shimizu ,K., Maseki, N., Kaneko, Y., Asou, H., Kamada, N. and Ohki, M. (1993). Junctions of AML1/CBFA2T1(ETO) fusion are constant in t(8-

21) acute myeloid leukemia detected by reverse transcription polymerase chain reaction. *Blood.* **82**: 1270-6.

99. Kuramochi, M., Fukuhara, H., Nobukuni, T., Kanbe, T., Maruyama, T., Ghosh, H. P., Pletcher, M., Isomura, M., Onizuka, M., Kitamura, T., Sekiya, T., Reeves, R. H. and Murakami, Y. (2001). TSLC1 is a tumor-suppressor gene in human non-small-cell lung cancer. *Nat Genet.* **7**(4): 427-30.

100. Kurose, K., Hoshaw-Woodard, S., Adeyinka, A., Lemeshow, S., Watson, P.H. and Eng, C. (2001). Genetic model of multi-step breast carcinogenesis involving the epithelium and stroma: clues to tumour-microenvironment interactions. *Hum Mol Genet.* 10(18): 1907-13.

101. Lammie, G. A., Fantl, V., Smith, R., Schuuring, E., Brookes, S., Michalides, R., Dickson, C., Arnold, A. and Peters, G. (1991). D11S287, a putative oncogene on chromosome 11q13, is amplified and expressed in squamous cell and mammary carcinomas and linked to BCL-1. *Oncogene.* **6**(3): 439-44.

102. Levy, D. E. and Darnell, J. E. Jr. (2002). Stats: transcriptional control and biological impact. *Nat. Rev. Mol. Cell. Biol.* **3**(9): 651-3.

103. Levy, D. E. and Lee, C. (2002). What does Stat3 do? *J Clin Invest.* **109**: 1143-8.

104. Li, Z., Wang, L., Zhang, L., Zhang, B., Yu, Y., Zeng, Z., Zhou, M., Huang, W., Chen, Z., Chen, S. and Li, G. (2001). Chromosomal aberrations analysed by comparative hybridization in nasopharyngeal carcinoma. *Zhonghua Yi Xue Za Zhi.* **18**(5): 338-42.

105. Liao, J., Fu, Y. and Shuai, K. (2000). Distinct roles of the NH2- and COOH-terminal domains of the protein inhibitor of activated signal transducer and activator

of transcription (STAT) 1 (PIAS1) in cytokine-induced PIAS1-STAT1 interaction. *Proc Natl Acad Sci USA.* **97**(10): 5267-72.

106. Lim, D. S. and Hasty, P. (1996). A mutation in mouse rad51 results in an early embryonic lethal that is suppressed by a mutation in p53. *Mol Cell Biol.* **16**: 7133-43.

107. Liu, B., Liao, J., Rao, X., Kunshner, S. A., Chung, C. D., Chang, D. D. and Shuai, A. K. (1998). Inhibition of Stat1-mediated gene activation by PIAS1. *Proc Natl Acad Sci USA.* **95**: 10626-31.

108. Liu, X., Robinson, G. W., Wagner, K. U., Garrett, L., Wynshaw-Boris, A. and Hennighausen, L. (1997). Stat5a is mandatory for adult mammary gland development and lactogenesis. *Genes Dev.* **11**: 179-86.

109. Loftus, B. J., Kim, U., Sneddon, V. S., Kalush, F., Brandon, R., Fuhrmann, J., Mason, T., Crosby, M., Barnstead, M., Cronin, L., Mays, A. D., Cao, Y., Xu, R. X., Kand, H., Mitchell, S., Eichler, E. E., Harris, P. C., Venter, C. J. and Adams, M. D. (1999). Gemone duplications and other features in 12 Mb of DNA sequence from human chromosome 16p and 16q. *Genomics.* **60**: 295-308.

110. Lohmann, D. R. (1999). RB1 gene mutations in retinoblastoma. *Hum Mutat.* **14**(4): 283-8.

111. Lorick, K. L., Jensen, J. P., Fang, S., Ong A. M., Hatakeyama, S., and Weissman, A. M. (1999). RING fingers mediate ubiquitin-conjugated enzyme (E2)-dependent ubiquitination. *Proc Natl Acad Sci USA.* **96**(20): 11364-9.

112. Lovering, R., Hanson, I, M., Borden, K. L., Martin, S., O'Reilly, N. J., Rahman, G. I., Pappin, D., Trowsdale, J. and Freemont, P. S. (1993). Identification and preliminary characterization of a protein motif related to the zinc finger. *Proc Natl Acad Sci USA.* **90**(6): 2112-6.

113. Lux, S. E., and John, K. M. (1990). Analysis of cDNA human erythrocyte ankyrin indicates a repeated structure with homology to tissue-differentiation and cell-cycle control proteins. *Nature*. **344**: 36-42.

114. Mancini, D., Singh, S., Ainsworth, P. and Rodenhiser, D. (1997). Constitutively methylated CpG dinucleotides as mutation hot spots in the retinoblastoma gene (RB1). *Am J Hum Genet.* **61**(1): 80-7.

115. Mattsson, K., Pokrovskaja, K., Kiss, C., Klein, G. and Szekely, L. (2001). Proteins associated with the promyelocytic leukemia gene product (PML)-containing nuclear body move to the nucleolus upon inhibition of proteasome-dependent protein degradation. *Proc Natl Acad Sci U S A*. **98**: 4913-7.

116. Medina, D. (1996). Preneoplasia in mammary tumorigenesis. *Cancer Treat Res.* **83**: 37-69.

117. Menard, S., Balsari, A., Casalini, P., Tagliabue, E., Campiglio, M., Bufalino, R. and Cascinelli, N. (2002). HER-2-positive breast carcinomas as a particular subset with peculiar clinical behaviors. *Clin Cancer Res*. **8**(2): 520-25.

118. Mendelsohn, J., Howely, P., Israel, M. A. and Liotta, L. A. (2001). The molecular basis of cancer. 2nd Edit, Saunders Company, USA.

119. Meraz, M. A., White, J. M., Sheehan, K. C., Bach, E. A., Rodig, S. J., Dighe A. S., Kaplan, D. H., Riley, J. K., Greenlund, A. C., Campbell, D., Carver-Moore, K., DuBois, R. N., Clark, R., Aguet, M. and Schreiber, R. D. (1996). Targeted disruption of the Stat1 gene in mice reveals unexpected physiologic specificity in the JNK-STAT signaling pathway. *Cell.* **84**: 431-42.

120. Miele, M. E, Jewett, M. D., Goldberg, S. F., Hyatt, D. L., Morelli, C., Gualandi, F., Rimessi, P., Hicks, D. J., Weissman, B. E., Barbanti-Brodano, G. and Welch, D. R. (2000). A human melanoma metastasis-suppressor locus maps to 6q16.3-q23. *Int J Cancer.* **86**(4): 524-8.

121. Miele, M. E., Robertson, G., Lee, J. H., Coleman, A., McGary, C. T., Fisher. P. B., Lugo, T. G. and Welch, D. R. (1996). Metastasis suppressed, but tumorigenicity and local invasiveness unaffected, in the human melanoma cell line MelJuSo after introduction of human chromosomes 1 or 6. *Mol Carcinog.* **15**(4): 284-99.

122. Miki, Y., Swensen, J., Shattuckeidens, D., Futreal, P. A., Harshman, K., Tavtigian, S., Liu, Q. Y., Cochran, C., Bennett, L. M., Ding, W., Bell, R., Rosenthal, J., Hussey, C., Tran, T., Mcclure, M., Frye, C., Hattier, T., Phelps, R., Haugenstrano, A., Katcher, H., Yakumo, K., Gholami, Z., Shaffer, D., Stone, S., Bayer, S. and Skolnick, M. H. (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science.* **266**(5182): 66-71.

123. Minamoto, T., Mai, M. and Ronai, Z. (2000). K-ras mutation: early detection in molecular diagnosis and risk assessment of colorectal, pancreas, and lung cancers-a review. *Cancer Detect Prev.* **24**(1): 1-12.

124. Moerland, E., Breuning, M.H., Cornelisse, C.J. and Cleton-Jansen, A.M. (1997). Exclusion of BBC1 and CMAR as candidate breast tumour-suppressor genes. *Br J Cancer.* **76**: 1550-3.

125. Moll, U. M., Riou, G., Levine, A. J. (1992). Two distinct mechanisms alter p53 in breast cancer: mutation and nuclear exclusion. *Proc Natl Acad Sci U S A.* **89**(15): 7262-6.

126. Mori, Y. M., Matsunaga, T., Abe, S. Fukushige, K. Miura, M. Sunamura, K. Shiiba, M. Sato, T. Nukiwa[3] and A. Horii. (1999). Chromosome band 16q24 is frequently deleted in human gastric cancer. *Br J Cancer* **80**: 556-62.

127. Mueller and Young, (1998). Emery's Elements of Medical Genetics, 10th ed. Churchill Livingston, Hong Kong.

128. Nakajima, K., Yamanaka, Y., Nakae, K., Kojima, H., Ichiba, M., Kiuchi, N., Kitaoka, T. Fukada, T., Hibi, M. and Hirano, T. (1996). central role for Stat3 in IL-6-induced regulation of growth and differentiation in M1 leukemia cells. *EMBO J.* **15**(14): 3651-8.

129. Narod, S. A. Modifiers of risk of hereditary breast and ovarian cancer. (2002). *Nat Rev Cancer.* **2**(2): 113-23.

130. Nathanson, K. L. and Weber, B. L. (2001). "Other" breast cancer susceptibility genes: searching for more holy grail. *Hum Mol Genet.* **10**(7): 715-20.

131. Negrini, M., Sabbioni, S., Possati, L., Rattan, S., Corallini, A., Barbanti-Brodano, G. and Croce, C. M. (1994). Suppression of tumorigenicity of breast cancer cells by microcell-mediated chromosome transfer: studies on chromosomes 6 and 11. *Cancer Res.* **54**(5): 1331-6.

132. O'Brien, W., Stenman, G. and Sager, R. (1986). Suppression of tumor growth by senescence in virally transformed human fibroblasts. *Proc Natl Acad Sci U S A.* **83**: 8659-65.

133. Oesterreich, S. and Fuqua, S. A. (2002). Tumor suppressor genes in breast cancer. *Endocr Relat Cancer.* **6**(3): 405-19.

134. Oshimura, M., Barrett, J. C. (1997). Multiple pathways to cellular senescence: role of telomerase repressors. *Eur J Cancer*. **33**(5): 710-5.

135. Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J. J. and de Jong, P. J. (1998). An improved approach for construction of bacterial artificial chromosome libraries *Genomics*. **52**(1): 1-8.

136. Osorio, A., de la Hoya, M., Rodriguez-Lopez, R., Martinez-Ramirez, A., Cazorla, A., Granizo, J. J., Esteller, M., Rivas, C., Caldes, T. and Benitez, J. (2002) Loss of heterozygosity analysis at the BRCA loci in tumor samples from patients with familial breast cancer. *Int J Cancer*. **99**(2): 305-9.

137. Ouchi, T., Lee S. W., Ouchi, M., Aaronson, S. A. and Horvath, C. M. (2000). Collaboration of signal transducer and activator of transcription 1 (STAT1) and BRCA1 in differential regulation of IFN-gamma target genes. *Proc Natl Acad Sci U S A*. **97**(10): 5208-13.

138. Pang, J. C., Dong, Z., Zhang, R., Liu, Y., Zhou, L. F., Chan, B. W., Poon, W. S. and Ng, H. K. (2003). Mutation analysis of DMBT1 in glioblastoma, medulloblastoma and oligodendroglial tumors. *Int J Cancer*. **105**(1): 76-81.

139. Parant, J. M. and Lozano, G. (2003). Disrupting TP53 in mouse models of human cancers. *Hum Mutat*. **21**(3): 321-6.

140. Polyak, K. (2001). On the birth of breast cancer. *Biochim Biophys Acta*. **1552**(1): 1-13.

141. Porter, D. A., Krop, I. E., Nasser, S., Sgroi, D., Kaelin, C. M., Marks, J. R., Riggins, G. and Polyak, K. (2001). A SAGE (serial analysis of gene expression) view of breast tumor progression. *Cancer Res*. **61**(15): 5697-702.

142. Powell, J. A., Gardner, A. E., Bais, A. J., Hinze, S. J., Baker, E., Whitmore, S., Crawford, J., Kochetkova, M., Spendlove, H. E., Doggett, N. A., Sutherland, G. R., Callen, D. F. and Kremmidiotis. G. (2002). Sequencing, transcript identification, and quantitative gene expression profiling in the breast cancer loss of heterozygosity region 16q24.3 reveal three potential tumor-suppressor genes. *Genomics.* **80**(3): 303-10.

143. Prendergast, G. C., Khosravi-Far, R., Solski, P. A., Kurzawa, H., Lebowitz, P. F. and Der, C. J. (1995). Critical role of Rho in cell transformation by oncogenic Ras. *Oncogene.* **10**(12): 2289-96.

144. Rae, J., Noack, D., Heyworth, P. G., Ellis, B. A., Curnutte, J. T. and Cross, A .R. (2000). Molecular analysis of 9 new families with chronic granulomatous disease caused by mutations in CYBA, the gene encoding p22(phox). *Blood.* **96**: 1106-12.

145. Rahman, N., Abidi, F., Ford, D., Arbour, L., Rapley, E., Tonin, P., Barton, D., Batcup, G., Berry, J., Cotter, F., Davison, V., Gerrard, M., Gray, E., Grundy, R., Hanafy, M., King, D., Lewis, I., Ridolfi Luethy, A., Madlensky, L., Mann, J., O'Meara, A., Oakhill, T., Skolnick, M., Strong, L. and Stratton, M. R. (1998). Confirmation of FWT1 as a Wilms' tumour susceptibility gene and phenotypic characteristics of Wilms' tumour attributable to FWT1. *Hum Genet.* **103**(5): 547-56.

146. Rajan, J., V., Wang, M., Marquis, S. T. and Chodosh, L. A. (1996). Brca2 is coordinately regulated with Brca1 during proliferation and differentiation in mammary epithelial cells. *Proc Natl Acad Sci USA.* **93**(23): 13078-83.

147. Rajavel, K. S. and Neufeld, E. F. (2001). Nonsense-mediated decay of human HEXA mRNA. *Mol Cell Biol.* **21**: 5512-9.

148. Rechsteiner, M., and Rogers, S. W. (1996). PEST sequences and regulation by proteolysis. *Trends Biochem Sci.* **21**: 267-71.

149. Reddel, R. R. (2000). The role of senescence and immortalization in carcinogenesis. *Carcinogenesis*. **21**(3): 477-84.

150. Reddy, D. E., Sandhu, A. K., DeReil, J. K., Athwal, R. S. and Kaur, G. P. (1999). Identification of a gene at 16q24.3 that restores cellular senescence in immortal mammary tumour cells. *Oncogene*. **18**: 5100-7.

151. Reddy, D. E., Keck, C. L., Popescu, N., Athwal, R. S. and Kaur, G. P. (2000). Identification of a YAC from 16q24 carrying a senescence gene for breast cancer cells. *Oncogene*. **19**: 217-22.

152. Ricketts, S. L., Garcia, N. F., Betz, B. L. and Coleman, W. B. (2002). Identification of candidate liver tumor suppressor genes from human 11p11.2-p12. *Genes Chromosomes Cancer*. **33**(1): 47-59.

153. Robertson, G. P., Huang, H. J. and Cavenee, W. K. (1999). Identification and validation of tumor suppressor genes. *Mol Cell Biol Res Commun*. **2**(1): 1-10.

154. Robertson, G., Coleman, A. and Lugo, T. G. (1996). A malignant melanoma tumor suppressor on human chromosome 11. *Cancer Res.* **56**(19): 4487-92.

155. Rödel, B., Tavassoli, K., Karsunky, H., Schmidt, T., Bachmann, M., Schaper, F., Heinrich, P., Shuai, K., Elsasser, H. P. and Moroy, T. (2000). The zinc finger protein Gfi-1 can enhance STAT3 signaling by interacting with the STAT3 inhibitor PIAS3. *EMBO J.* **19**(21): 5845-55.

156. Roest Crollius, H., Jaillon, O., Bernot, A., Dasilva, C., Bouneau, L., Fischer, C., Fizames, C., Wincker, P., Brottier, P., Quetier, F., Saurin, W. and Weissenbach, J. (2000). Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence. *Nat Genet*. **25**(2): 235-8.

157. Russo, J., Yang, X., Hu, Y. F., Bove, B. A., Huang. Y., Silva, I. D., Tahin, Q., Wu, Y., Higgy, N., Zekri, A. and Russo, I. H. (1998). Biological and molecular basis of human breast cancer. *Front Biosci.* **3**: D944-60.

158. Saccone, S., De Sario, A., Della Valle, G. and Bernardi, G. (1992). The highest gene concentrations in the human genome are in telomeric bands of metaphase chromosomes. *Proc Natl Acad Sci U S A.* **89**: 4913-7.

159. Sachdev, S., Hoffman, A. and Hannink, M. (1998). Nuclear localization of IkB ankyrin repeats de ne a novel class of *cis*-acting nuclear import sequences. *Mol Cell Biol.* **18**(5): 2524-34.

160. Saito, H., Inazawa, J., Saito, S., Kasumi, F., Koi, S., Sagae, S., Kudo, R., Saito, J., Noda, K. and Nakamura, Y. (1993). Detailed deletion mapping of chromosome 17q in ovarian and breast cancers: 2-cM region on 17q21.3 often and commonly deleted in tumors. *Cancer Res.* **53**(14): 3382-5.

161. Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). Molecular cloning: a laboratory manual. Second Edition. Cold Spring Harbour Laboratory Press, New York.

162. Sato, M., Mori, Y., Sakurada, A., Fukushige, S., Ishikawa, Y., Tsuchiya, E., Saito, Y., Nukiwa, T., Fujimura, S. and Horii, A. (1998). Identification of a 910-kb region of common allelic loss in chromosome bands 16q24.1-q24.2 in human lung cancer. *Genes Chromosomes Cancer.* **22**: 1-8.

163. 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 cancers with loss of heterozygosity at 16q24.3. *Genomics*. **61**: 219-26.

164. Scheffzek, K., Ahmadian, M. R., Kabsch, W., Wiesmuller, L., Lautwein, A., Schmitz, F. and Wittinghofer, A. (1997). The Ras-RasGAP complex: structural basis for GTPase activation and its loss in oncogenic Ras mutants. *Science.* **277**(5324): 333-8.

165. Schmidt, D. and Muller, S. (2002). Members of the PIAS family act as SUMO ligases for c-Jun and p53 and repress p53 activity. *Proc Natl Acad Sci U S A.* **99**: 2872-7.

166. Schneider, G. and Schmid, R. M. (2003). Genetic alterations in pancreatic carcinoma. *Mol Cancer.* **2**(1): 1-15.

167. Schultz, D. C., Vanderveer, L., Berman, D. B., Hamilton, T. C., Wong, A. J. and Godwin, A. K. (1996). Identification of two candidate tumor suppressor genes on chromosome 17p13.3. *Cancer Res.* **56**(9): 1997-2002.

168. Scita, G., Tenca, P., Frittoli, E., Tocchetti, A., Innocenti, M., Giardina, G. and Di Fiore, P. P. (2000). Signaling from Ras to Rac and beyond: not just a matter of GEFs. *EMBO J.* **19**(11): 2393-8.

169. Scully, R., Chen, J., Plug, A., Xiao, Y., Weaver, D., Feunteun, J., Ashley, T. and Livingston, D. M. (1997). Association of BRCA1 with Rad51 in mitotic and meiotic cells. *Cell.* **88**(2): 265-75.

170. Sedwick, S. G. and Smerdon, S. J. (1999). The ankyrin repeat: a diversity of interactions on a common structural framework. *Trends Biochem Sci.* **24**(8): 311-6.

171. Seraj, M. J., Samant, R. S., Verderame, M. F. and Welch, D. R. (2000). Functional evidence for a novel human breast carcinoma metastasis suppressor, BRMS1, encoded at chromosome 11q13. *Cancer Res.* **60**(11): 2764-9.

172. Settasatian, C., Whitmore, S. A., Crawford, J., Bilton, R. L., Cleton-Jansen, A. M., Sutherland, G. R. and Callen, D. F. (1999). Genomic structure and expression analysis of the spastic paraplegia gene, SPG7. *Hum Genet.* **105**: 139-44.

173. Shao, J. Y., Huang, X. M., Huang, L. X., Wu, Q. L., Xia, J. C., Wang, H. Y., Feng, Q. S., Ren, Z. F., Ernberg, I., Hu, L. F. and Zeng, Y. X. (2001). Loss of heterozygosity and its correlation with clinical outcome and Epstein-Barr virus infection in nasopharyngeal carcinoma. *Anticancer Res.* **21**(4B): 3021-9.

174. Shao, J. Y., Zeng, W. F. and Zeng, Y. X. (2002). Molecular genetic progression on nasphyaryngeal carcinoma. *Ai Zheng.* **21**(1): 1-10.

175. Shapiro, M. B. and Senapathy, P. (1987). RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acid Res.* **15**: 7155-74.

176. Shen, C. Y., Yu, J. C., Lo, Y. L., Kuo, C. H., Yue, C. T., Jou, Y. S., Huang, C. S., Lung, J. C. and Wu, C. W. (2000). Genome-wide search for loss of heterozygosity using laser capture microdissected tissue of breast carcinoma: an implication for mutator phenotype and breast cancer pathogenesis. *Cancer Res.* **60**(14): 3884-92.

177. Shinohara A., Ogawa H., Ogawa T. (1992). Rad51 protein involved in repair and recombination in *S. cerevisiae* is a Rec-A-like protein. *Cell.* **69**: 457-70.

178. Smith, P. D., Crossland, S., Parker, G., Osin, P., Brooks, L., Waller, J., Philp, E., Crompton, M. R., Gusterson, B. A., Allday, M. J. and Crook, T. (1999). Novel p53

mutants selected in BRCA-associated tumours which dissociate transformation suppression from other wild-type p53 functions. *Oncogene*. **18**(15): 2451-9.

179. Smith-Sorensen B, Hovig E. (1999). CDKN2A (p16INK4A) somatic and germline mutations. *Hum Mutat.* **7**(4): 294-303.

180. Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol.* **98**(3): 503-17.

181. Sorescu, D., Szocs, K. and Griendling, K. K. (2001). NAD(P)H oxidases and their relevance to atherosclerosis. *Trends Cardiovasc Med.* **11**: 124-31.

182. Sovak, M. A., Bellas, R. E., Kim, D. W., Zanieski, G. J., Rogers, A. E., Traish, A. M. and Sonenshein, G. E. (1997). *J Clin Invest.* **100**(12): 2952-60.

183. Spataro, V., Norburt, C. and Harris, A. (1998). The ubiquitin-proteasome pathway in cancer. *Br J Cancer* **77**: 448-55.

184. Staden, R., Beal, K.F., and Bonfield, J.K. (2000). The Staden package, 1998. *Methods Mol Biol.* **132**: 115-30.

185. Stanbridge, E. J. (1976). Suppression of malignancy in human cells. *Nature.* **260**(5546): 17-20.

186. Stirzaker, C., Millar, D. S., Paul, C. L., Warnecke, P. M., Harrison, J., Vincent, P. C., Frommer, M. and Clark, S. J. (1997). Extensive DNA methylation spanning the Rb promoter in retinoblastoma tumors. *Cancer Res.* **57**(11): 2229-37.

187. Storey, A., Thomas, M., Kalita, A., Harwood, C., Gardiol, D., Mantovani, F., Breuer, J., Leigh, I. M., Matlashewski, G. and Banks, L. (1998). Role of a p53

polymorphism in the development of human papillomavirus-associated cancer. *Nature*. **393**(6682): 229-34.

188. Suzuki, A., Delapompa, J. L., Hakem, R., Elia, A., Yoshida, R., Mo, R., Nishina, H., Chuang, T. Wakeham, A. Itie, A., Koo, W., Billia, P., Ho, A., Fukumoto, M., Hui, C. C. and Mak, T. W. (1997) BRCA2 is required for embryonic cellular proliferation in the mouse. *Genes & Development*. **11**(10): 1242-52.

189. 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 Chromosomes Cancer*. **17**: 225-33.

190. Takeichi, M. Cadherin cell adhesion receptors as a morphogenetic regulator. (1991). *Science*. **251**(5000): 1451-5.

191. Teglund, S., MaKay, C., Schuetz, E., van Deursen, J. M., Stravopodis, D., Wang, D., Brown, M., Bodner, S., Grosveld, G. and Ihle, J. N. (1998). Stat5a and Stat5b proteins have essential and nonessential, or redundant, roles in cytokine responses. *Cell*. **93**: 841-50.

192. Thai, T. H., Du, F., Tsan, J. T., Jin, Y., Phung, A., Spilllman, M. A., Massa, H. F., Muller, C. Y., Ashfaq, R., Mathis, J. M., Miller, D. S., Trask, B. J., Baer, R. and Bowcock, A. M. (1998). Mutations in BRCA1-associated RING domain (BARD1) gene in primary breast, ovarian and uterine cancers. *Hum Mol Genet*. **7**(2): 195-202.

193. Tonin, P. N. (2000). Genes implicated in hereditary breast cancer syndromes. *Semin Surg Oncol*. **18**(4): 281-6.

194. 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**: 513-7.

195. Turkson, J. and Jove, R. (2000). STAT proteins: novel molecular targets for cancer drug discovery. *Oncogene.* **19**: 6613-26.

196. Udy, G., Towers, R. Snell, R., Wilkins, R., Park, S., Ram, P., Waxman, D. and Davey, H. (1997). Requirement of Stat5b for sexual dimorphism of body growth rates and liver gene expression. *Proc Natl Acad Sci USA.* **94**: 7239-44.

197. Uejima, H., Mitsuya, K, Kugoh, H., Horikawa, I. And Oshimura, M. (1995). Normal human chromosome 2 induces cellular senescence in the human cervical carcinoma cell line SiHa. *Genes Chromosomes Cancer.* **14**(2): 120-7.

198. Uejima, H., Shinohara, T., Nakayama, Y., Kugoh, H. and Oshimura, M. (1998). Mapping a novel cellular-senescence gene to human chromosome 2q37 by irradiation microcell-mediated chromosome transfer. *Mol Carcinog.* **22**(1): 34-45.

199. van 't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M. J., Witteveen, A. T., Schreiber, G. J., Kerkhoven, R. M., Roberts, C., Linsley, P. S., Bernards, R. and Friend, S, H. (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature.* **415**(6871): 530-6.

200. Venter J. C., Adams M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A. *et al.* (2001). The sequence of the human genome. *Science.* **291**: 1304-51.

201. Verma, I. M., Stevenson, J. K., Schwarz, E. M., Antwerp, D. V. and Miyamoto, S. (1995). Rel/NF-kB/IkB family: intimate tales of association and dissociation. *Genes and Development.* **9**: 2723-35.

202. Visser, M., Sijmons, C., Bras, J., Arceci, R. J., Godfried, M., Valentijn, L. J., Voute, P. A. and Baas, F. (1997). Allelotype of pediatric rhabdomyosarcoma. *Oncogene.* **15**: 1309-14.

203. Wagner, K. D., Wagner, N. and Schedl, A. (2003). The complex life of WT1. *J Cell Sci.* **116**(9): 1653-8.

204. Wang, I., Hwang, L. Y., Bowcock, A. M. and Baer, R. (1993). Tumour suppressor activity of *RB* and *p53* genes in human breast carcinoma cells. *Oncogene.* **8**: 279-88.

205. Wang, X., Gleich, L., Pavelic, Z. P., Li, Y. Q., Gale, N., Hunt, S., Gluckman, J. L. and Stambrook, P. J. (1999). Cervical metastases of head and neck squamous cell carcinoma correlate with loss of heterozygosity on chromosome 16q. *Int J Oncol.* **14**(3): 557-61.

206. Weintraub S. J, Chow, K. N., Luo, R. X., Zhang, S. H., He, S. and Dean, D. C. (1995). Mechanism of active transcriptional repression by the retinoblastoma protein. *Nature.* **375**(6534): 812-5.

207. Westra, W. H., Baas, I. O., Hruban, R. H., Askin, F. B., Wilson, K., Offerhaus, G. J. and Slebos, R. J. (1996). K-ras oncogene activation in atypical alveolar hyperplasias of the human lung. *Cancer Res.* **56**(9): 2224-8.

208. 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*. **50**: 1-8.

209. 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*. **52**: 325-31.

210. Wieland, I., Arden, K. C., Michels, D., Klein-Hitpass, L., Bohm, M., Viars, C. S. and Weidle, U. H. (1999). Isolation of DICE1: a gene frequently affected by LOH and downregulated in lung carcinomas. *Oncogene*. **18**: 4530-7.

211. Wistuba, I. I., Behrens, C., Virmani, A. K., Mele, G., Milchgrub, S., Girard, L., Fondon, J. W. 3rd., Garner, H. R., McKay, B., Latif, F., Lerman, M. I., Lam, S., Gazdar, A. F. and Minna, J. D. (2000). High resolution chromosome 3p allelotyping of human lung cancer and preneoplastic/preinvasive bronchial epithelium reveals multiple, discontinuous sites of 3p allele loss and three regions of frequent breakpoints. *Cancer Res.* **60**(7): 1949-60.

212. Wooster, R., Neuhausen, S. L., Mangion, J., Quirk, Y., Ford, D., Collins, N., Nguyen, K., Seal, S., Tran, T., Averill, D., *et al.* (1994). Localisation of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. *Science*. **265**: 2088-90.

213. Wright, F. A., Lemon, W. J., Zhao, W. D., Sears, R., Zhuo, D., Wang, J. P., Yang, H. Y., Baer, T., Stredney, D., Spitzner, J., Stutz, A., Krahe, R. and Yuan, B. (2001). A

draft annotation and overview of the human genome. *Genome Biol.* **2**(7): RESEARCH0025. Epub 2001 Jul 04.

214. Wu, L. C., Wang, Z. W., Tsan, J. T., Spillman, M. A., Phung, A., Xu, X. L., Yang, C. W., Hwang, L. Y., Bowcock, A. M. and Baer, R. (1996). Identification of a RING protein that can interact *in vivo* with the BRCA1 gene product. *Nat Genet.* **14**: 430-9.

215. Yakicier, M. C., Legoix, P., Vaury, C., Gressin, L., Tubacher, E., Capron, F., Bayer, J., Degott, C., Balabaud, C. and Zucman-Rossi, J. (2001). Identification of homozygous deletions at chromosome 16q23 in aflatoxin B1 exposed hepatocellular carcinoma. *Oncogene.* **20**(37): 5232-8.

216. Yan, J., Fang, Y., Liang, Q., Huang, Y. and Zeng, Y. (2001). Novel chromosomal alterations detected in primary nasphyaryngeal carcinoma by comparative genomic hybridization. *Chin Med J.* **114**(4): 418-21.

217. Yarden, R. I. and Brody, L. C. (1999) BRCA1 interacts with components of the histone deacetylase complex. *Proc Natl Acad Sci U S A.* **96**(9): 4983-8.

218. Yeager, T. R., DeVries, S., Jarrard, D. F., Kao, C., Nakada, S. Y., Moon, T. D., Bruskewitz, R., Stadler, W. M., Meisner, L. F., Gilchrist, K. W., Newton, M. A., Waldman, F. M. and Reznikoff, C. A. (1998). Overcoming cellular senescence in human cancer pathogenesis. *Genes Dev.* **12**(2): 163-74.

219. Yuan, B. Z., Zhou, X., Zimonjic, D. B., Durkin, M. E., Popescu, N. C. (2003). Amplification and Overexpression of the EMS 1 Oncogene, a Possible Prognostic Marker, in Human Hepatocellular Carcinoma. *J Mol Diagn.* **5**(1): 48-53.

220. Yuan, Y., Mendez, R., Sahin, A. and Dai, J. L. (2001). Hypermethylation leads to silencing of the SYK gene in human breast cancer. *Cancer Res.* **61**: 5558-61.

221. Yuille, M. A. and Coignet, L. J. (1998). The ataxia telangiectasia gene in familial and sporadic cancer. *Recent Results Cancer Res.* **154**: 156-73.

222. Zhong, Q., Chen, C. F., Li, S., Chen, Y. M., Wang, C. C., Xiao, J., Chen, P. L., Sharp, Z. D. and Lee, W. H. (1999). Association of BRCA1 with the hRad50-hMre11-p95 complex and the DNA damage response. *Science.* **285**(5428): 747-50.

223. Zhong S, Muller S, Ronchetti S, Freemont PS, Dejean A, Pandolfi PP.(2000). Role of SUMO-1-modified PML in nuclear body formation. *Blood.* **95**(9): 2748-52.

Appendix #1

Appendix 1:

1.1) Suppliers

• Ajax Chemicals	Australia
• American Type Culture Collection (ATTC)	Virginia, USA
• Amersham Phamacia Biotech	Buckinghamshire, England
• BAC/PAC Recourses	Rosewell Park Cancer Institute, New York, USA
• BDH Lab Supplies	Poole, England
• BD Biosciences	California, USA
• BioRad Laboratories	California, USA
• Genome Systems	St. Louis, Missouri, USA
• Invitrogen Coorporation	California, USA
• MBI Fermentas	Western Australia, Australia
• NEN Life Science Products	Massachusetts, USA
• New England Bioloabs	Beverly, USA
• Perkin Elmer	California, USA
• Pierce	California, USA
• Pharmaica Biotech	Uppsala, Sweden
• Progen	California, USA
• Promega	WI, USA
• Qiagen	California, USA
• Roche	Germany
• Santa Cruz Biotechnology Inc.	California, USA
• Sigma Chemical Company	St Louis, MO, USA
• Silenus	Victoria, Australia
• Stratagene	California, USA
• Vector Laboratories	California, USA

1.2) Chemicals and compounds

• Bactoagar	Sigma
• Bactotryptone	Sigma
• Brij-35	BDH Lab Supplies
• BSA (bovine serum albumin-pentax fraction V)	Sigma
• Chloroform	BDH Lab Supplies
• DAPI (diamidino phenylindole dihydrochloride)	Sigma
• DEPC (diethylpyrocarbonate)	BDH Lab Supplies
• Deoxy-nucleotide triphosphates (dNTPs)	Pharmacia Biotech
• Dextran sulphate	Pharmacia Biotech
• *N,N*-dimethyl formamide	Sigma
• DMEM (Dulbecco's modified Eagles medium)	Invitrogen
• DMSO (dimethylsulphoxide)	Sigma
• DTT (dithiothreitol)	Sigma
• EDTA (ethylenediaminetetracetic acid)	Ajax
• EGTA	Ajax
• Ethanol (99.5 % v/v)	BDH Lab Supplies
• FCS (fetal calf serum)	Invitrogen
• Ficoll (Type 400)	Sigma
• Formaldehyde	Ajax
• Formamide	Sigma
• Glucose	Ajax
• Glutamine	Ajax
• Glycerol	Ajax
• Human placental DNA	Sigma
• IPTG (isopropylthio-β-D-galactosidase)	Sigma
• Isopropanol	Ajax
• *N*-lauroylsarcosine (Sarkosyl)	Sigma
• Nylon membrane	NEN Life Science Products
• Lipofectamine 2000	Invitrogen

• Magnesium chloride ($MgCl_2.6H_2O$)	Ajax
• Magnesium sulphate ($MgSO_4.7H_2O$)	Ajax
• Maltose	BDH Lab Supplies
• β-Me (β-mercaptoethanol)	BDH Lab Supplies
• Mixed bed resin (20-50 mesh)	BioRad, USA
• Paraffin oil	Ajax
• PEG (polyethylene glycol) 3350	Sigma
• Phenol	Sigma
• PMSF	Sigma
• Polybrene	Sigma
• Polyvinypyrolidone (PVP-40)	Sigma
• Potassium dihydrogen orthophosphate (KH_2PO_4)	Ajax
• Propidium iodide	Sigma
• Salmon sperm DNA	Calbiochem
• Sodium acetate	Ajax
• Sodium biocarbonate	Ajax
• Sodium chloride	Ajax
• tri-Sodium citrate	Ajax
• Sodium dihydrogen orthophosphate ($NaH_2PO4.2H_20$)	Ajax
• SDS (sodium dodecyl sulphate)	BDH Lab Supplies
• Sodium floride (NaF)	Sigma
• Sodium vanadate (Na_2VO_3)	Sigma
• Slide mounting solution	Vector Labaratories
• di-Sodium hydrogen orthophosphate ($Na_2HPO_4.7H2O$)	Ajax
• Sodium hydroxide	Ajax
• Sucrose	Ajax
• Tris-base	Invitrogen
• Tris-HCl	Invitrogen

- Triton X-100 — Ajax
- Yeast extract — Invitrogen

1.3) Cytokines.

- Interferon α (INFα) — Sigma
- Interferon γ (INFγ) — Sigma
- Leukemia inhibitory factor (LIF), human — Sigma

1.4) Antibodies and related reagents.

1.4.1) Primary antibodies.

- Flag — Sigma
- GST — Amersham Phamacia Biotech
- myc (9E10) — Santa Cruz Biotehnology
- myc-agarose (9E10) — Santa Cruz Biotehnology
- PIAS1 (?) — Santa Cruz Biotehnology
- PIAS3 (?) — Santa Cruz Biotehnology
- phospho-STAT1 — Cell signaling
- phospho-STAT3 (B-7) — Santa Cruz Biotehnology
- PML — Santa Cruz Biotehnology
- STAT1 — Cell signaling
- STAT3 (C-20) — Santa Cruz Biotehnology

1.4.2) Secondary antibodies.

- Goat conjugated horse radish peroxidase (HRP) — Silenus, Australia
- Goat conjugated FITC — Silenus, Australia
- mouse conjugated horse radish peroxidase (HRP) — Silenus, Australia
- mouse conjugated CY3 — Silenus, Australia
- mouse conjugated FITC — Silenus, Australia
- Rabbit conjugated HRP — Silenus, Australia
- Rabbit conjugated FITC — Silenus, Australia

1.4.3) Other reagents.

- GST-sepharose — Amersham Phamacia Biotech
- Protein-G — Amersham Phamacia Biotech

1.5) Enzymes.

All restriction enzymes were purchased from either New England Biolabs or Progen. 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.

- CIAP (calf intestinal alkaline phosphatase) — New England Biolabs
- *E.Coli* DNA polymerase I (Klenow fragment) — New England Biolabs
- Protease K — MBI Fermentas
- RNaseH — Promega
- RNAsin — Promega
- Superscript RNase H^- reverse transcriptase — Invitrogen
- T4 DNA Ligase — Progen
- *Taq* DNA polymerase — Invitrogen
- X-gal (5-Bromo-4-chloro-3-indoyl-β-D galactosidase) — Progen

1.6) Electrophoresis.

- Acrylamide (40%) — BioRad
- Agarose — Progen
- Ammonium persulphate (APS) — BDH
- Bromophenol blue — BDH
- Enhanced chemiluminescence (ECL) — Amersham
- Ethidium bromide (EtBr) — Sigma
- Genescreen Plus nylon membrane — Dupont
- Nitrocellulose — Amersham
- TEMED (tetramethylethylenediamine) — Progen

1.7) Antibiotics.

- Ampicillin — Sigma
- Chloramphenicol — Sigma
- Kanamycin — Progen
- Tetracycline — Sigma
- G418 — Sigma
- Penicillin — Invitrogen
- Streptomycin — Invitrogen

1.8) Bacterial Strains.

All constructs were transformed into commercially purchased *E. coli* XL-1 Blue MRF' competent cells (Stratagene), or *E.coli* DH5α competent cells (Invitrogen).

1.9) Media.

All liquid media was prepared with double distilled water and was sterilized by autoclaving. In the cases where liquid media was prepared for the pouring of plates, antibiotics were added once the autoclaved media had cooled to approximately 50°C.

1.9.1) Bacterial media.

- L-Broth (Luria-Bertaini Broth): 1% (w/v) Bactryptone; 0.5 % (w/v) yeast extract; 1 % (w/v) yeast NaCl; pH to 7.5 with NaOH.
- L-Agar: L-Broth; 1.5% (w/v) Bactoagar.
- L-Ampicillin: L-Broth; 1.5% (w/v) Bactoagar; 100 μg/ml ampicillin.
- L-Chloramphenicol: L-Broth; 1.5% (w/v) Bactoagar; 34 μg/ml chloramphenicol.
- L-Kanamycin: L-Broth; 1.5% (w/v) Bactoagar; 50 μg/ml kanamycin.
- L-Tetracycline: L-Broth; 1.5% (w/v) Bactoagar; 34 μg/ml tetracycline.

1.9.2) Cell culture media.

- DMEM — Invitrogen
- OPTI-MEM — Invitrogen

- RPMI — Invitrogen
- FCS (fetal calf serum) — Invitrogen
- L-glutamate — Invitrogen

1.10) DNA vectors.

• pSP72	cloning vector	Invitrogen
• pBluescript	cloning vector	Invitrogen
• pcDNA3.1	mammalian expression vector	Invitrogen
• pTarget	mammalian expression vector	Promega
• pDONR	Gateway donor vector	Invitrogen
• pDEST12.2	mammalian expression vector	Invitrogen
• pDEST27	mammalian expression vector N-terminal GST tag	Invitrogen
• pLNCX2	retroviral expression vector	Stratagene
• pVPack-VSV-G	retroviral vector	Stratagene
• pVPack-GP	retroviral vector	Stratagene

1.11) Miscellaneous materials and kits.

- [α-^{32}P]dCTP — Perkin Elmer
- BAC human genomic library, release II — Genome Systems
- BAC human genomic library RPCI-II (segment 3) — BAC/PAC Resources
- BigDye Terminator Cycle sequencing kit — Perkin Elmer
- β-Galactosidase assay kit — Stratagene
- Dyal-Luciferase Reporter Assay System — Promega
- dNTPs — Invitrogen
- ExpressHYB solution — BD Biosciences
- halfTERM Dye Terminator sequencing reagent — GenPack
- Labeling mix-dCTP — Pharmacia
- Multiple tissue Northern blots — BD Biosciences
- NuPAGE gel electrophoresis — Invitrogen

- Oligo-dT_{12-18} — Invitrogen
- PAC human genomic library, release 1 — Genome Systems
- PAC human genomic library RPCI-4 — BAC/PAC Resources
- pGEM-T cloning kit — Promega
- Protease inhibitors — Roche
- pTarget cloning kit — Promega
- Qiagen Plasmid mini kit — Qiagen
- QIAquick purification kit — Qiagen
- 5' RACE kit — Invitrogen
- Random hexamers — Perkin Elmer
- 1X Reporter Lysis Buffer — Promega
- TRIzol — Invitrogen
- 3MM Whatman paper — Lab Supply

1.12) Markers.

1.12.1 Protein

- Prestained protein ladder — Invitrogen
- Magic marker — Invitrogen

1.12.2 DNA

- SPP1 (EcoRI restricted) — Invitrogen
- Puc19 (HpaaII restricted) — Invitrogen
- High M_w markers — Invitrogen

1.13) Cell lines. From ATCC

- Breast cancer: BT-549, CAMA-1, MDA-MB-134, MDA-MB-157, MDA-MB-231, MDA-MB-436, MDA-MB-468, SKBR-3, T47-D, ZR-75-1 and ZR-75-30
- Human embryonic kidney: 293T

1.14) Buffers and solutions.

- 1X PCR buffer: 20 mM Tris-HCl (pH 8.0); 50 mM KCl.
- 10X agarose loading buffer: 10 mM Tris-HCl (pH 8.0); 200 mM EDTA (pH 8.0); 2 % (w/v) sarkosyl; 15 % ficoll 400; 0.1 % bromophenol blue; 0.1 % cyanol.
- Colony denaturing solution: 1.5 M NaCl; 0.5 M NaOH.
- Colony neutralizing solution: 3 M NaCl; 0.5 M Tris-HCl (pH 8.0);
- De-ionised formamide: 2 g of mixed bed resin/50 ml formamide. Mix for one hour and filter.
- 100X Denhart's solution: 2 % (w/v) ficoll 400; 2 % (w/v) polyvinylpirolidone; 2 % (w/v) BSA.
- 10X dNTP labeling solution: 1 mM labeling mix minus dCTP.
- Filter denaturing solution: 0.5 m NaOH.
- Filter neutralizing solution: 0.2 M Tris-HCl; 2X SSC.
- 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 neutralizing solution: 1.5 M NaCl; 0.5 M Tris-HCl (pH 7.5)
- 10X labeling buffer: 0.5 M Tris-HCl (pH 7.5); 0.1 M $MgCl_2$; 10 mM DTT; 0.5 mg/ml BSA; 0.05 A_{260}/µl random hexamers (Pharmacia).
- 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 $NaH_2PO_4.2H_2O$; 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.
- TSS: L-Broth; 10% (w/v) PEG 3350; 5% (v/v) DMSO; 50 mM $MgCL_2$.

Appendix #2

Appendix 2:

Ref #	Name of Site	Address of site
1	Gene map 99	www.ncbi.nlm.nih.gov/genemap/page.cgi?=Home.html
2	Radiation hybrid panel Genebridge 4(GB4) server	www.sanger.ac.uk/Software/Rhsever.shtml
3	The National Center for Biotechnology Information (NCBI)	http://www.ncbi.nim.gov/
4	LALIGN (global alignments between two nucleiotide sequences).	http://www.ch.embnet.org/software/LALIGN_form.html
5	GENSCAN	http://genes.mit.edu/GENSCAN.html
6	The Human Gene Nomenclature committee	http://www.gene.ucl.ac.uk/cgi-bin/nomenclature/searchgenes.pl
7	International Human Genome Sequencing Consortium	http://www.ensembl.org/
8	Celera	http://www.celera.com/
9	GeneFINDER	www.bioscience.org/urllists/genefind.htm
10	ProDom	www.prodes.toulouse.inra.fr/prodom/2002.1/html/home.php
11	Pfam	www.sanger.ac.uk/Software/Pfam/
12	PESTfind	www.at.embnet.org/embnet/tools/ bio/PESTfind/
13	PfScan	http://hits.isb-sib.ch/cgi-bin/PFSCAN

Appendix #3

doi:10.1006/geno.2002.6828, available online at http://www.idealibrary.com on IDEAL

Article

Sequencing, Transcript Identification, and Quantitative Gene Expression Profiling in the Breast Cancer Loss of Heterozygosity Region 16q24.3 Reveal Three Potential Tumor-Suppressor Genes

Jason A. Powell,[1,2] Alison E. Gardner,[1] Anthony J. Bais,[1] Susan J. Hinze,[1,4] Elizabeth Baker,[1] Scott Whitmore,[1,4] Joanna Crawford,[1,4] Marina Kochetkova,[1] Hayley E. Spendlove,[1] Norman A. Doggett,[5] Grant R. Sutherland,[1,2] David F. Callen,[1,4] and Gabriel Kremmidiotis[1,3,4,*]

[1]*Centre for Medical Genetics, Department of Cytogenetics and Molecular Genetics, Women's and Children's Hospital, North Adelaide, South Australia 5006, Australia*
Departments of [2]Genetics and [3]Paediatrics, University of Adelaide, Adelaide 5000, Australia
[4]*Bionomics Limited, Thebarton, South Australia 5031, Australia*
[5]*DOE Joint Genome Institute, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*

[*]*To whom correspondence and reprint requests should be addressed. Fax: +(618) 83546199. E-mail: gkremmid@bionomics.com.au.*

Loss of heterozygosity (LOH) of chromosome 16q24.3 is a common genetic alteration observed in invasive ductal and lobular breast carcinomas. We constructed a physical map and generated genomic DNA sequence data spanning 2.4 Mb in this region. Detailed *in silico* and *in vitro* analyses of the genomic sequence data enabled the identification of 104 genes. It was hypothesized that tumor-suppressor genes would exhibit marked mRNA expression variability in a panel of breast cancer cell lines as a result of downregulation due to mutation or hypermethylation. We examined the mRNA expression profiles of the genes identified at 16q24.3 in normal breast, a normal breast epithelial cell line, and several breast cancer cell lines exhibiting 16q24.3 LOH. Three of the genes, *CYBA*, Hs.7970, and *CBFA2T3*, exhibited variability ten times higher than the baseline. The possible role of these genes as tumor suppressors is discussed.

Key Words: breast cancer, 16q24.3, LOH, tumor suppressor

INTRODUCTION

Frequent loss of heterozygosity (LOH) at 16q has been observed in several tumor types such as breast [1], prostate [2], lung [3], hepatocellular [4], gastric [5], head and neck squamous cell carcinoma [6], and rhabdomyosarcoma [7]. Consequently, it has been suggested that this chromosome arm harbors one or more tumor-suppressor loci. Detailed LOH studies in breast and prostate tumors have identified three regions of frequent allelic imbalance, 16q22.1, 16q23.2–q24.1, and 16q24.3 [1]. Polymorphic markers encompassing these regions in breast cancer exhibit LOH in 28–89% of preinvasive ductal tumors [8] and 40–60% of primary breast tumors [9]. Irrespective of disease stage or histological phenotype, LOH is most frequently observed at 16q24.3 [10], suggesting a key role for this region in breast tumorigenesis.

The minimum region for breast cancer LOH at 16q24.3 has been resolved to an interval between the polymorphic marker *D16S498* and the 16q telomere [1]. No tumor suppressor loci have yet been identified in this region. We have previously carried out mutation analyses on seven genes mapping to this interval: *SPG7* [11], *BBC1* [12], *CPNE7* [13], *CDK10* [14], *FANCA* [15], *GAS11*, and *C16ORF3* [16]. None of these genes were found to harbor any mutations in sporadic breast cancer DNA samples. We have extended our studies by constructing a physical map and carrying out large-scale genomic DNA sequence analysis and detailed annotation to identify other candidate tumor-suppressor genes in this region.

Mutations may result in aberrant mRNA expression due to nonsense-mediated mRNA decay [17], disruption of promoter function [18], or alteration of mRNA stability [19]. In addition, promoter hypermethylation has been recently demonstrated to have a central role in the downregulation of tumor-suppressor genes [20]. Consequently, it is reasonable to hypothesize that tumor-suppressor genes associated with

doi:10.1006/geno.2002.6828, available online at http://www.idealibrary.com on IDEAL

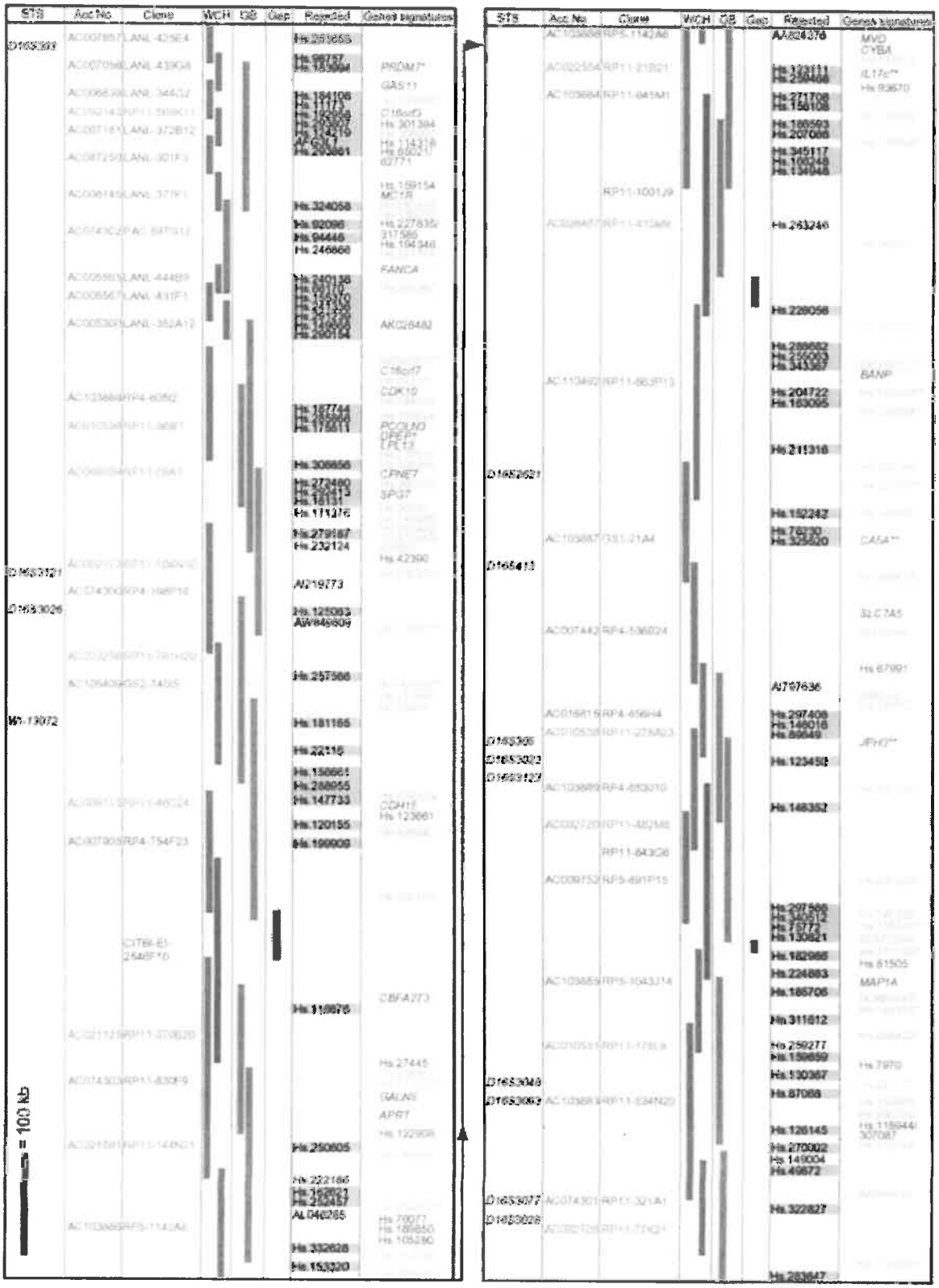


FIG. 1. Integrated physical and transcript map of the sequenced LOH region at 16q24.3. The map is anchored at the telomere and extends just over 2.4 Mb centromeric, to the marker *D16S3028*. Clones sequenced by our group are shown in red (WCH); clones integrated from GenBank (GB) are shown in green. The physical map is contiguous and three sequencing gaps are indicated in black. BAC clones spanning sequence gaps are shown (blue). Genes (blue), gene signatures (orange), and rejected gene signatures are ordered with their corresponding sequenced genomic clones adjacent. Gene signatures were rejected on the basis of either *in silico* analysis (shaded) or RT-PCR (plain). Genes and gene signatures excluded from real-time analysis due to late C_T values (*) or unsatisfactory melt curves (**) are indicated.

the 16q24.3 LOH may be aberrantly expressed due to the presence of mutations or promoter hypermethylation. We have investigated the expression profile of the transcripts identified at 16q24.3 in a panel of non-tumor and breast cancer cell lines containing 16q24.3 LOH. Several of these genes were found to exhibit aberrant mRNA expression.

RESULTS

Physical Map Construction

We constructed a high-resolution physical map of 16q24.3 demarcated by *D16S3028* to the telomere. We had previously constructed a cosmid physical map covering the region from the gene *CDH15* to the 16q telomere [21] (Fig. 1). We have extended this map to include *D16S3028*. Initially, we mapped several markers on a chromosome 16 somatic cell hybrid panel [22] and subsequently used them as probes to screen large-insert PAC and BAC libraries. Initial walking points included the genes *SLC7A5* and *APRT* and the Unigene cluster Hs.7970. Genemap 99 radiation hybrid markers corresponding to this region of chromosome 16 were integrated into our physical map (Fig. 1). We obtained an average of four times PAC/BAC clone coverage of this region. We used fluorescent *in situ* hybridization to confirm the mapping of the clones representing the minimal tiling path to 16q24.3 (data not shown). Two of the clones, RP11-863P13 and GS1-21A4, hybridized to 16q24.3 and 16p11.2, indicating the existence of a chromosome 16 specific duplication.

Identification of Genes and Gene Signatures

The 16q24.3 minimal tiling path served as the template for large-scale sequence analysis. We sequenced 26 clones at 2–3× coverage, accurately assembling and ordering these contigs. We subsequently used this sequence to pull down and integrate sequence data from 13 clones available from the first draft of the human genome project (Golden Path; Fig. 1) [23]. The combined unique sequence of 2.4 Mb stretches from the marker *D16S3028* to the telomere. These combined data provide good sequence coverage of 16q24.3 reflected in 5 contiguous clones, 13 clones in five or less contigs, and 10 clones in seven or more contigs. Only two clones between the genes *BANP* and *SLC7A5* are sparsely sequenced. The minimal tiling path has an average overlap of 13.4 kb, ranging from 490 bp to 49.7 kb. There are three sequence gaps spanned by clones CITBI-E1-2546F10, RP11-1001J9, and RP11-843G6. These gaps correspond to estimated sizes of 23 kb, 10 kb, and 7.5 kb, respectively.

We identified genes in the region by BLAST analysis of the NCBI (http://www.ncbi.nlm.nih.gov/) non-redundant, Unigene, and EST databases using the genomic sequence of the region as the query. These analyses led to the identification of 43 genes, 141 Unigene EST clusters, and 382 singleton ESTs. The Unigene EST clusters and singleton ESTs did not define clear gene structures with corresponding open reading frames. These ESTs may define genes that can only be partially constructed from the information available in public databases or they may represent genomic contamination clones that are known to exist in the public EST databases [24]. We selected the term "gene signatures" to describe these ESTs and carried out detailed *in silico* analyses to predict their status as either possible genes or database contaminations. All gene signatures corresponding to Unigene EST clusters were considered as possible genes, except those containing 3′ genomic poly(A) sequence stretches. Detailed inspection and assembly of physically adjacent Unigene clusters enabled the grouping of several gene signatures into single genes. Gene signatures represented by singleton ESTs were considered to be possible genes if they were found to have an exon/intron structure, mouse homologous sequences, protein homologs, or GENSCAN predicted exons. This *in silico* selection allowed us to identify a total of 116 genes. We rejected 82 Unigene clusters and most singleton ESTs as contaminations or redundancies (Fig. 1).

We tested these *in silico* predicted genes experimentally for their transcription status by RT-PCR analysis. We used total RNA from 18 adult and 3 fetal human tissues as template for the RT-PCR amplification of each predicted gene. This analysis demonstrated that 104 of the 116 predicted genes represent transcribed sequences and therefore define actual genes. Twelve predicted genes gave non-reproducible RT-PCR bands or failed to amplify. These either are not genes or are representative of genes that are expressed in a tightly regulated manner. For example, expression may be restricted to certain stages of development or to tissues not represented in the panel used in this study. These *in silico* predicted genes were not studied any further. The mRNA expression profiles of the genes giving clear RT-PCR evidence of transcription were analyzed in normal breast and a panel of breast cancer cell lines.

Quantitative Expression Profiling of 16q24.3 Genes in Breast Cancer Cell Lines

To identify a possible involvement of any 16q24.3 genes in breast cancer, we carried out quantitative mRNA expression analysis of all the corresponding transcripts in normal breast tissue and a panel of human breast cancer cell lines (Fig. 2). Using the expression value of each gene in normal breast tissue as a baseline, we calculated the relative fold-difference between the cell line exhibiting the highest expression and the cell line exhibiting the lowest expression. This value was termed the "relative fold variability index" (RFVI). Initially, we examined four housekeeping genes, which established an RFVI baseline range of 1 to 42 ($RFVI_{cyclophilin\ A} = 1$ and $RFVI_{ATP5A} = 42$). This baseline range was accepted to reflect mRNA expression differences that are due to normal population variations or experimental variability. Subsequently, we determined the RFVI for *SYK* and *CDKN2A*, two known tumor-suppressor genes that have previously shown aberrant expression in breast cancer cells [25,26]. The RFVI values for *CDKN2A* and *SYK* were over 10 times greater than the upper limit of the baseline range (Fig. 2). Of the 104 transcripts analyzed, 66 displayed RFVI values within the baseline range and 9 exhibited

doi:10.1006/geno.2002.6828, available online at http://www.idealibrary.com on IDEAL

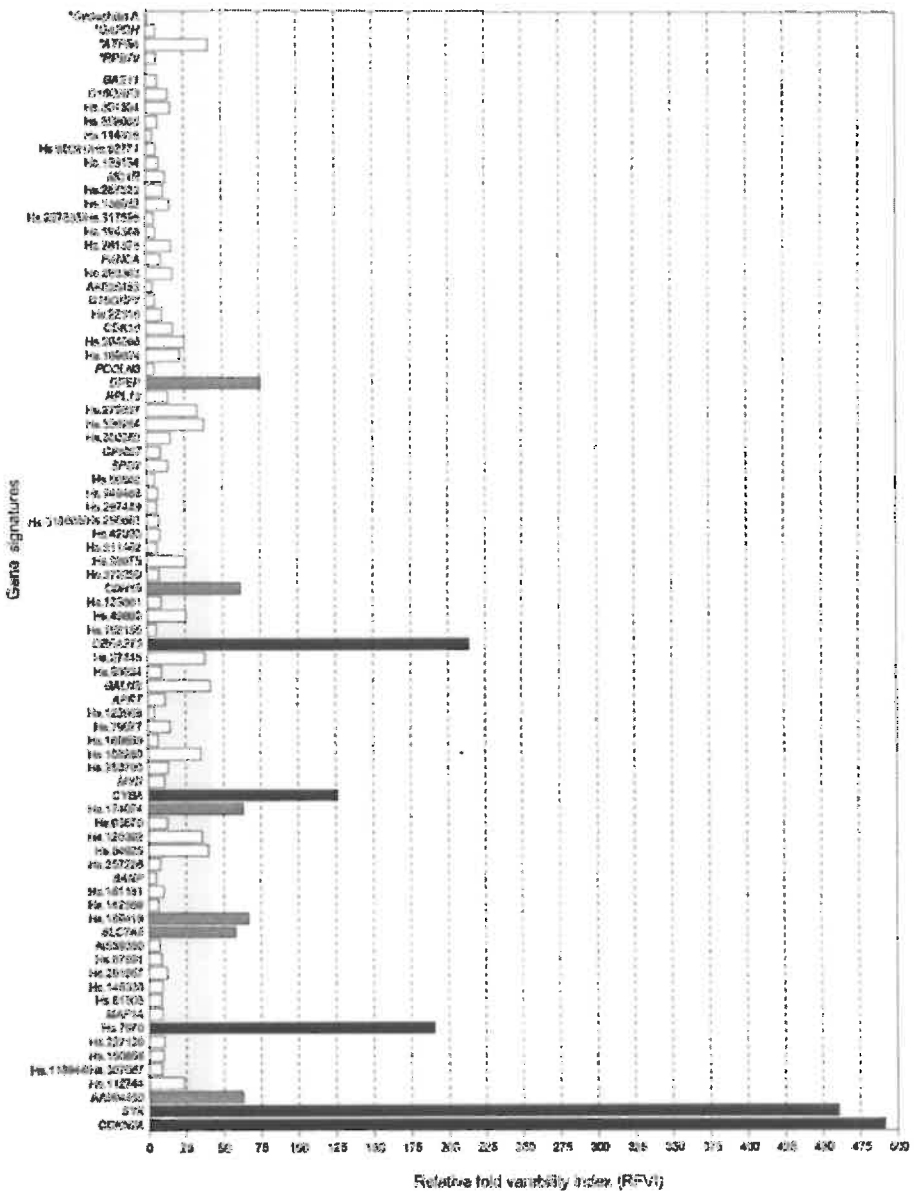


FIG. 2. Gene expression variability in breast cancer cell lines exhibiting chromosome 16 LOH. The range of mRNA expression variability for each gene analyzed is expressed as RFVI. The RFVI values of four housekeeping genes (*) were used to establish a baseline range ($RFVI_{baseline}$ = 1–42). The tumor-suppressor genes *SYK* and *CDKN2A* display very high RFVI values. The RFVI values of 75 genes mapping at 16q24.3 are shown. Six genes display moderately elevated variability (gray bars) and three genes display significant elevated variability (black bars).

higher values. The gene signatures corresponding to *DPEP1*, *CDH15*, Hs.17074, Hs.189419, *SLC7A5*, and AA994450 exhibited RFVI values that were only marginally higher than the baseline (Fig. 2). The individual expression profiles for these gene signatures are shown in Fig. 3.

CBFA2T3, *CYBA*, and Hs.7970 displayed RFVI values considerably higher than the baseline. Figure 4 shows the individual expression profiles of these genes and *SYK* across our cell line panel. *CBFA2T3* and *CYBA* both display variable expression, with some cell lines exhibiting marginal upregulation and some exhibiting marked downregulation. The Hs.7970 gene was downregulated in all the cell lines, with the downregulation being more pronounced in two cell lines.

We found that 29 transcripts displayed complex melting curves and/or very late C_T values. The expression of these genes was considered to be too low to allow meaningful interpretation of real-time PCR results.

Discussion

Draft versions of the human genome have recently become available in the form of electronically assembled, annotated scaffolds [23,27] (http://www.ensembl.org/ and http://www.celera.com/). Naturally occurring repeats, electronic chimeras, and duplications have resulted in the presence of a substantial number of mapping discrepancies and misalignments in these sequence drafts [28]. In contrast, the 16q24.3 gene map presented here is based on sequence data derived from an experimentally determined physical map. Southern hybridization, PCR-based probe mapping, and *in situ* hybridization were used for the construction of this physical map, making it a more reliable source of gene order in this region of the genome. Our high-throughput sequencing of this region has yielded an ordered sequence "backbone" that allowed us to identify and assemble in this region other clones that have been randomly and partially sequenced by other sequencing centers. This ordered sequence provides a valuable resource for the future "ordered" gap closure and completion of the sequence in this region. In addition, it serves as a short-term resource for tumor-suppressor gene discovery efforts until the sequence of this region is completed.

We used our draft sequence of 16q24.3 to identify genes mapping in this region. Our strategy for gene identification in genomic sequence data was integrated, combining computational predictions, human curation, and experimental validation. Initially, genes were identified by *in silico* means involving comparisons with expressed sequence databases, cross-species sequence conservation, algorithm-based gene prediction, and manual inspection. This analysis allowed the identification of 43 actual genes that have been previously published or can be *in silico* constructed based on information

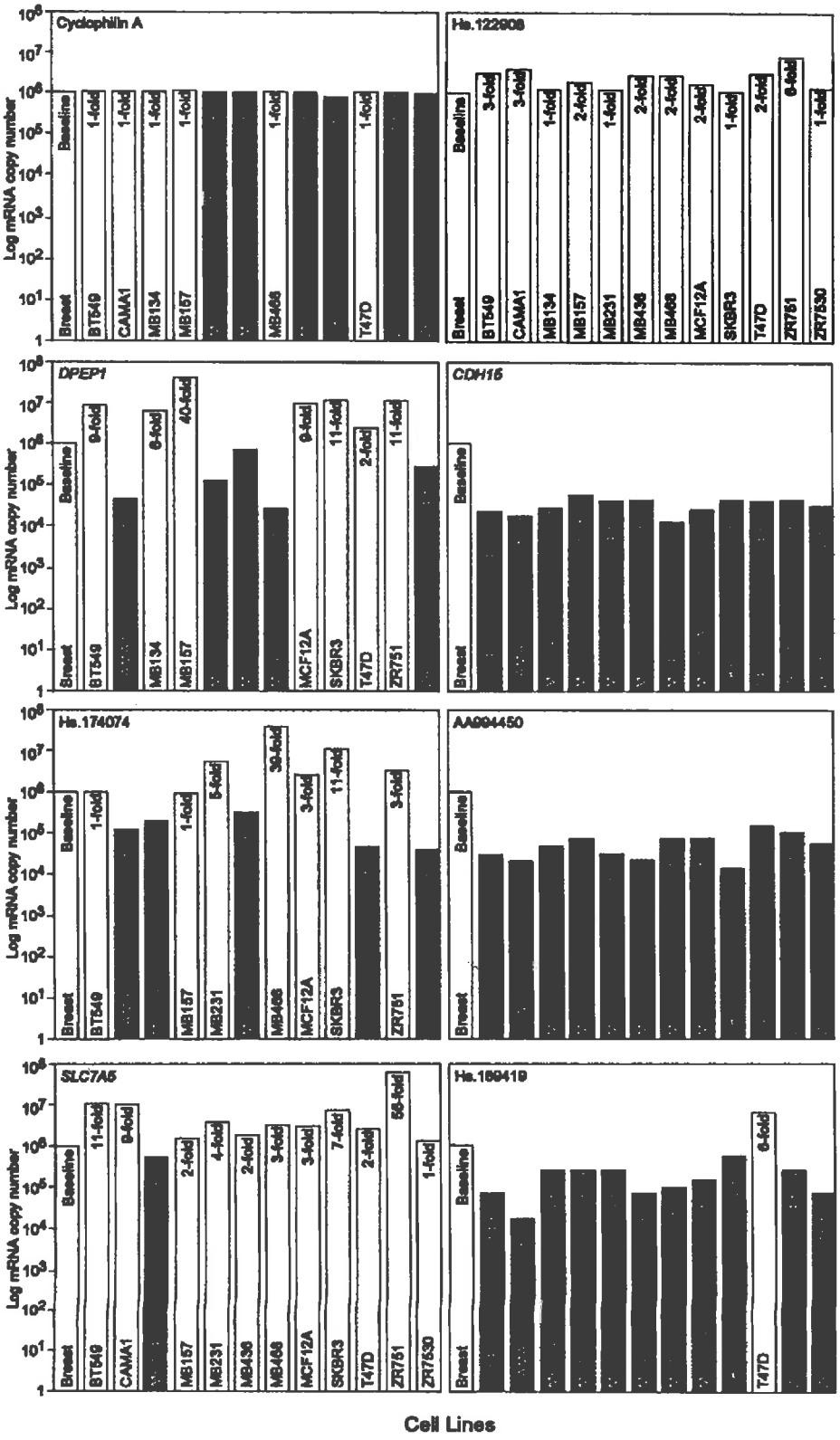


FIG. 3. mRNA expression profiling of genes mapping at 16q24.3. Data are expressed as fold change calculated relative to expression in the "baseline" normal breast tissue. Open and filled bars represent fold upregulation and downregulation with respect to the baseline. Cyclophilin A is representative of the expression profiling observed with housekeeping genes. Hs.122908 is representative of genes exhibiting expression variability within the housekeeping gene range. *DPEP1*, *CDH15*, Hs.17074, Hs.189419, *SLC7A5*, and AA994450 were found to be marginally above the housekeeping gene expression variability range.

available in public databases. In addition, 73 possible gene loci were identified which cannot be constructed from information available in public databases and are represented by a very small number of ESTs and, in some cases, even single ESTs. Due to their limited sequence representation it cannot be determined *in silico* whether these ESTs represent genes or are the product of genomic DNA contamination in cDNA libraries. Consequently, we carried out analysis in the form of RT-PCR to provide experimental evidence for the representation of these genes in the human transcriptome. Our analysis indicated that 61 of these 73 possible genes are transcribed. In total, our detailed *in silico* and RT-PCR analyses allowed the identification of 104 genes in the region. The gene density in the 16q24.3 region is not uniform. There are three gene-dense regions: one between *D16S3121* and the telomere, and one around each of the genes *MAP1A* and *MVD*. The gene density in these regions is one gene per 15 kb. These three areas are separated by areas with densities of one gene per 35 kb. On average, there is approximately one gene every 20 kb. This is consistent with previous observations that telomeric, Giemsa-light bands are gene-rich [29].

Several of the genes identified are good tumor-suppressor candidates based on their function or based on homologies to other genes or proteins of known function. These genes can be selected and prioritized for mutation analyses. However, most genes identified are of no known or predicted function. Consequently, we studied the mRNA expression profiles of all these genes as an independent selection criterion for the identification of potential tumor suppressors. Our working hypothesis was that tumor-suppressor gene expression would exhibit marked variability across a panel of different breast cancer cell lines. We found this to be true for *SYK* and

doi:10.1006/geno.2002.6828, available online at http://www.idealibrary.com on IDEAL

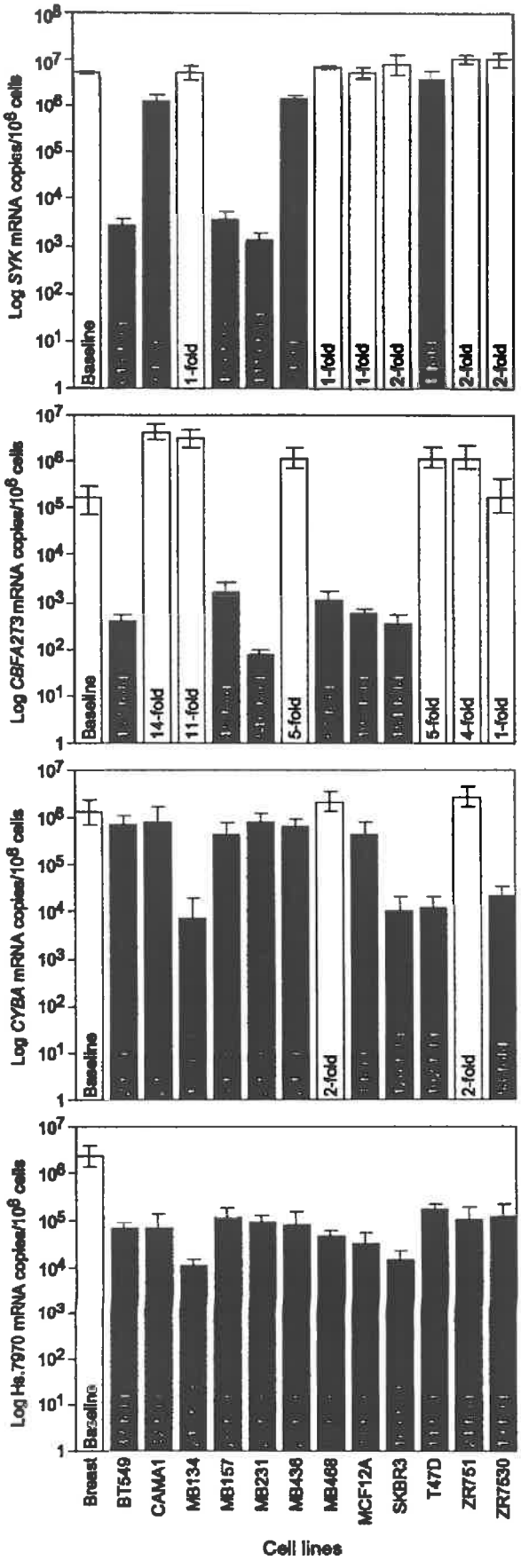


FIG. 4. mRNA expression analysis of the tumor-suppressor gene *SYK* and the 16q24.3 transcripts *CBFA2T3*, *CYBA*, and Hs.7970 by real-time RT-PCR. Data are expressed as log mRNA copy number/10^6 cells (mean ± SD, $n = 4$). Fold changes shown were calculated relative to expression in the "baseline" normal breast tissue. Black and white bars represent fold upregulation and downregulation respective to the baseline.

CDKN2A, two genes previously shown to act as tumor suppressors. Based on comparisons with the expression profiles of housekeeping genes, we identified six genes of moderately elevated variability and three genes exhibiting significant variability values. *DPEP1*, *CDH15*, Hs.17074, Hs.189419, *SLC7A5*, and AA994450 were found to be marginally more variable than the baseline defined by housekeeping genes. The possible involvement in tumor-suppressor events of these marginally variable genes should be viewed with caution and will require confirmation by independent methodology. However, *CBFA2T3*, *CYBA*, and Hs.7970 exhibited significant expression variability.

CBFA2T3 displayed expression that was five times more variable than the baseline. Expression of this gene was reduced in the cell lines BT-549, MDA-MB-157, and MDA-MB-231. These cell lines are also downregulated for *SYK* and have been previously shown to display typical features of malignant phenotype including increased motility and invasion [25]. *CBFA2T3* expression was also reduced in SK-BR-3 and MDA-MB-468. These two cell lines are not downregulated for *SYK* and display pre-malignant phenotypes with very low motility and predominantly non-invasive cluster formation [25]. Furthermore, *CBFA2T3* is also downregulated in MCF12A, which is non-tumorigenic. These findings suggest that *CBFA2T3* downregulation is an early event in breast cancer. This is consistent with the notion that chromosome 16 LOH occurs early in carcinogenesis [30]. *CBFA2T3* is a member of the *ETO* family that also includes *CBFA2T1*, *CBFA2T2*, and *Drosophila melanogaster nervy*. *CBFA2T1* and *CBFA2T3* were identified through their involvement in chromosomal translocations observed in acute myeloid leukemia (AML) [31,32]. Such translocation events result in the generation of AML1–CBFA2T1 or AML1–CBFA2T3 fusion proteins that repress genes normally activated by AML1. CBFA2T1 has been shown to interact with corepressor complexes involving N-CoR, mSin3A, and histone deacetylases [33]. Based on the high sequence conservation between *CBFA2T1* and *CBFA2T3*, it is reasonable to hypothesize that CBFA2T3 also exhibits transcription repressor function.

We also found that *CYBA* was three times more variable than the housekeeping gene baseline. Expression of this gene was reduced in the cell lines MDA-MB-134, SK-BR-3, and T47-D. The role of CYBA as an NADPH oxidase in the microbicidal function of phagocytes has been studied extensively. Mutations in *CYBA* cause chronic granulomatous disease characterized by recurrent bacterial and fungal infections [34]. More recently, CYBA has been implicated in oxidases involved in epithelial and muscle cell gene regulation and function with implications in atherosclerosis [35].

CYBA associates with CYBB to form cytochrome b-558, which is the membrane component of NADPH oxidase and functions as the final electron transporter in the oxidation of NADPH, resulting in the generation of reactive oxygen species (ROS) such as O_2^- and H_2O_2. Several studies have shown the involvement of ROS in carcinogenesis and tumor progression [36]. ROS concentration levels are critical in the regulation of several genes involved in diverse pathways delineating transcription, proliferation, and apoptosis [37–39]. ROS levels are under a tight regulatory control involving the interplay of NADPH oxidases and antioxidant ROS scavengers [40]. Any disruption to these control mechanisms is likely to result in aberrant cell behavior such as that seen in cancer.

In addition to the structural evidence provided by LOH, functional evidence supporting the presence of a tumor-suppressor gene at 16q24.3 has come from micro-cell mediated transfer experiments. The introduction of a 16q24.3 YAC clone into breast cancer cell lines induced cellular senescence [41]. The gene represented by the unigene cluster Hs.7970 maps in the region included in this YAC clone (J.C. *et al.*, unpublished data). In our studies, this gene exhibited expression variability four times higher than the baseline. However, this finding should be interpreted with caution because this gene exhibited low expression in all the cell lines studied. Consequently, the exponential nature of PCR amplification combined with the small number of target molecules may result in the exaggeration of trivial variations. The Hs.7970 gene encodes a protein that appears to contain an F-box domain. F-box containing proteins function as ubiquitin protein ligases and the ubiqitin-proteasome pathway is often the target of cancer-related deregulation [42].

The highly variable expression patterns observed for *CBFA2T3*, *CYBA*, and Hs.7970 suggest a possible involvement in breast cancer and warrant further experimentation to determine the exact role these genes may have in breast cancer processes. The notion of these genes being breast cancer tumor suppressors is compatible with what is known about their respective functions.

Materials and Methods

Physical map construction. We screened BAC and PAC libraries with markers located between *D16S3028* [1] and *CDH15* [43]. These markers included genes, cDNAs, and microsatellites originally obtained from Genemap 99 (http://www.ncbi.nlm.nih.gov/genemap/page.cgi?F=Home.html) using the radiation hybrid panel Genebridge 4 (GB4) server (http://www.sanger.ac.uk/Software/RHserver/RHserver.shtml). We prepared DNA from positive clones using standard techniques (Qiagen). We digested the DNA with *EagI* and resolved it by pulse field gel electrophoresis. We aligned clones based on their restriction enzyme pattern and confirmed the overlaps by Southern blot hybridization. Additional hybridizations of various markers, including probes derived from the clones themselves, aided in map construction and consolidation. We used fragments representing the extreme ends of individual contigs to re-screen the BAC and PAC libraries. Fluorescent *in situ* hybridization (FISH) was used to confirm the chromosomal localizations of clones selected for large-scale sequencing.

Large-scale sequence analyses. We used DNA from cosmid BAC and PAC clones to construct random puc18 sub-libraries. Large-scale sequencing of sub-library clones was carried out on ABI377 sequencers. We assembled and analyzed the genomic sequence data using PHRED, PHRAP [44], and GAP4 [45] software on a SUN workstation. The sequence we generated was used to anchor other draft sequence data deposited to GenBank by other sequencing centers.

In silico analyses. Our investigation for matching and/or homologous nucleotide and protein sequences involved the use of the BLASTn and BLASTp [46] algorithms to search databases at the National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/). We used the algorithm LALIGN for performing global alignments between two nucleotide sequences. LALIGN parameters were set at default values (http://www.ch.embnet.org/software/LALIGN_form.html). We used the ClustalW algorithm in multiple protein sequence alignments. We used the GENSCAN program (http://genes.mit.edu/GENSCAN.html) for the prediction of exons and genes in genomic DNA sequence.

Nomenclature. Genes previously described in the literature or in public databases are referred to with their approved symbols as they appear in the Human Gene Nomenclature Database (http://www.gene.ucl.ac.uk/cgi-bin/nomenclature/searchgenes.pl). Other genes of no previously described function are referred to using the Unigene database code number as found at NCBI (http://www.ncbi.nlm.nih.gov/UniGene/). Singleton ESTs are referred to by their GenBank accession numbers.

Cell lines and tissues. We used the following breast cancer cell lines: BT-549, CAMA-1, MDA-MB-134, MDA-MB-157, MDA-MB-231, MDA-MB-436, MDA-MB-468, SK-BR-3, T47-D, ZR-75-1, and ZR-75-30. We purchased these from the American Type Culture Collection (ATCC) and cultured them in RPMI or OPTI-MEM medium (Gibco) supplied with 10% FCS. Additional culture supplements were used as per the cell line specification sheets. The cell lines used were selected based on LOH data. Eleven of the cell lines used contain LOH involving chromosome 16 [47]. The cell line MCF12A was derived from normal breast epithelium and was included in our studies as a normal control.

RT-PCR analysis of the *in silico* identified gene signatures was carried out on total RNA from 21 human tissues mixed into six pools: pool 1, brain, heart, kidney, liver; pool 2, lung, trachea, colon, bone marrow; pool 3, spleen, thymus, prostate, skeletal muscle; pool 4, testis, uterus, fetal brain, fetal liver; pool 5, spinal cord, placenta, adrenal gland, salivary gland, fetal lung; and pool 6, breast.

RNA extraction and RT-PCR. We isolated total RNA from cell lines using TRIzol (Gibco BRL). Total RNA derived from 21 human tissues was purchased commercially (Clontech, Stratagene, Ambion). We removed DNA contamination from all RNA preparations using DNAfree DNase (Ambion). We confirmed the absence of contaminating DNA by using primers that only amplify from intronic DNA (primers from the *CHRNA4* locus: 5′-CTGGAGATGTTTGTGGCCTT-3′ and 5′-TGCTTCACCCATACGTC-3′). This primer set failed to amplify from all the cDNA samples used in our study. RNA concentrations were estimated using RiboGreen (Molecular Probes). We performed reverse transcription reactions using oligo $(dT)_{16}$ and Superscript RNase H^- reverse transcriptase (Gibco BRL). We used the Lasergene Primer Select software (DNASTAR) to selected primer sets for each transcript. Wherever possible, we designed primers to span exon–intron boundaries. All primer sequences used in this study are available on request. Transcript expression status was confirmed by amplification from a variety of human tissue cDNA preparations using HotStarTaq (Qiagen) according to the manufacturer's instructions. Transcripts with null or poor amplification were studied with at least two different primer sets.

Real-time PCR and data normalization. We carried out real-time PCR on a Rotor-Gene 2000 (Corbett Research). Reactions (25 μl) included 12.5 μl SYBR Green I PCR Master Mix (PE Biosystems), 0.2 μM of each primer, and 30 ng cDNA template. Amplification cycling was performed as follows: 94°C for 10 minutes, followed by 45 cycles of 93°C for 20 seconds, 60°C for 30 seconds, and 70°C for 30 seconds. Fluorescence data acquisition was at 510 nm during each 70°C extension phase. We performed melt curve analysis and agarose gel electrophoresis following each real-time PCR run to assess product specificity. For real-time PCR quantification we used internal standard curves as described [48]. mRNA copies per nanogram of total RNA used were calculated by the cycle threshold value (C_T) extrapolation to standard curves constructed with serial dilutions of known concentration templates. Copy num-

bers per cell were calculated based on a total-RNA/cell value of 4 pg. We analyzed the expression levels of the housekeeping genes cyclophilin A, *GAPDH*, RNA polymerase subunit II, and ATPase subunit II in breast cancer cell lines to determine the most accurate endogenous control for data normalization. Cyclophilin A displayed the most uniform expression profile and was used to normalize the results obtained from all the genes studied. Fold changes in cell line expression were calculated using the normal breast expression levels as a reference. The range of expression fold change between the panel of 12 cell lines was plotted as a single value for each transcript. We chose the term relative fold variability index (RFVI) to describe this value.

ACKNOWLEDGMENTS

We thank Erica Woollatt and Helen Eyre (both from Women's and Children's Hospital, North Adelaide, South Australia) for instruction and assistance with FISH, and The Australian Genome Research Facility (http://www.agrf.org.au/) for their contribution to our large-scale sequencing effort. This work was supported by research grants from the Adelaide Women's & Children's Hospital Foundation, the National Health and Medical Research Council of Australia, with funding and resources also provided by Bionomics Limited.

RECEIVED FOR PUBLICATION MARCH 18; ACCEPTED JUNE 17, 2002.

REFERENCES

1. Cleton-Jansen, A. M., *et al.* (2001). Loss of heterozygosity mapping at chromosome arm 16q in 712 breast tumours reveals factors that influence delineation of candidate regions. *Cancer Res.* **61:** 1171–1177.
2. Suzuki, H., *et al.* (1996). Three distinct commonly deleted regions of chromosome arm 16q in human primary and metastatic prostate cancers. *Genes Chromosomes Cancer* **17:** 225–233.
3. Sato, M., *et al.* (1998). Identification of a 910-kb region of common allelic loss in chromosome bands 16q24.1-q24.2 in human lung cancer. *Genes Chromosomes Cancer* **22:** 1–8.
4. Chou, Y. H., Chung, K. C., Jeng, L. B., Chen, T. C., and Liaw, Y. F. (1998). Frequent allelic loss on chromosomes 4q and 16q associated with human hepatocellular carcinoma in Taiwan. *Cancer Lett.* **123:** 1–6.
5. Mori, Y., *et al.* (1999). Chromosome band 16q24 is frequently deleted in human gastric cancer. *Br. J. Cancer* **80:** 556–562.
6. Wang, X., *et al.* (1999). Cervical metastases of head and neck squamous cell carcinoma correlate with loss of heterozygosity on chromosome 16q. *Int. J. Oncol.* **14:** 557–561.
7. Visser, M., *et al.* (1997). Allelotype of pediatric rhabdomyosarcoma. *Oncogene* **15:** 1309–1314.
8. 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 tumour suppressor gene region. *Cancer Res.* **56:** 5605–5609.
9. Devilee, P., and Cornelisse, C. J. (1994). Somatic genetic changes in human breast cancer. *Biochimica et Biophysica Acta* **1198:** 113–130.
10. 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.
11. Settasatian, C., *et al.* (1999). Genomic structure and expression analysis of the spastic paraplegia gene, SPG7. *Hum. Genet.* **105:** 139–144.
12. Moerland, E., Breuning, M. H., Cornelisse, C. J., and Cleton-Jansen, A. M. (1997). Exclusion of BBC1 and CMAR as candidate breast tumour-suppressor genes. *Br. J. Cancer* **76:** 1550–1553.
13. Savino, M., *et al.* (1999). Characterization of copine VII, a new member of the copine family, and its exclusion as a candidate in sporadic breast cancers with loss of heterozygosity at 16q24.3. *Genomics* **61:** 219–226.
14. Crawford, J., *et al.* (1999). The PISSLRE gene: structure, exon skipping, and exclusion as tumour suppressor in breast cancer. *Genomics* **56:** 90–97.
15. Cleton-Jansen, A. M., *et al.* (1999). Mutation analysis of the Fanconi anaemia A gene in breast tumours with loss of heterozygosity at 16q24.3. *Br. J. Cancer* **79:** 1049–1052.
16. Whitmore, S. A., *et al.* (1998). Characterization and screening for mutations of the growth arrest-specific 11 (GAS11) and C16orf3 genes at 16q24.3 in breast cancer. *Genomics* **52:** 325–331.
17. Rajavel, K. S., and Neufeld, E. F. (2001). Nonsense-mediated decay of human HEXA mRNA. *Mol. Cell. Biol.* **21:** 5512–5519.
18. 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.
19. Wieland, I., *et al.* (1999). Isolation of DICE1: a gene frequently affected by LOH and downregulated in lung carcinomas. *Oncogene* **18:** 4530–4537.
20. Yuan, Y., Mendez, R., Sahin, A., and Dai, J. L. (2001). Hypermethylation leads to silencing of the SYK gene in human breast cancer. *Cancer Res.* **61:** 5558–5561.
21. Whitmore, S. A., *et al.* (1998). 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* **50:** 1–8.
22. Callen, D. F., *et al.* (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.
23. International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. *Nature* **409:** 860–921.
24. Gonzalez, I. L., and Sylvester, J. E. (1997). Incognito rRNA and rDNA in databases and libraries. *Genome Res.* **7:** 65–70.
25. Coopman, P. J., *et al.* (2000). The SYK tyrosine kinase suppresses malignant growth of human breast cancer cells. *Nature* **406:** 742–747.
26. Bisogna, M., *et al.* (2001). Molecular analysis of the INK4A and INK4B gene loci in human breast cancer cell lines and primary carcinomas. *Cancer Genet. Cytogenet.* **125:** 131–138.
27. Venter, J. C., *et al.* (2001). The sequence of the human genome. *Science* **291:** 1304–1351.
28. Katsanis, N., Worley, K. C., and Lupski, J. R. (2001). An evaluation of the draft human genome sequence. *Nat. Genet.* **29:** 88–91.
29. Saccone, S., De Sario, A., Della Valle, G., and Bernardi, G. (1992). The highest gene concentrations in the human genome are in telomeric bands of metaphase chromosomes. *Proc. Natl. Acad. Sci. USA* **89:** 4913–4917.
30. Gong, G., *et al.* (2001). Genetic changes in paired atypical and usual ductal hyperplasia of the breast by comparative genomic hybridization. *Clin. Cancer Res.* **7:** 2410–2414.
31. Kozu, T., *et al.* (1993). Junctions of AML1/CBFA2T1(ETO) fusion are constant in t(8-21) acute myeloid leukemia detected by reverse transcription polymerase chain reaction. *Blood* **82:** 1270–1276.
32. Gamou, T., *et al.* (1998). The partner gene of AML1 in t(16;21) myeloid malignancies is a novel member of the CBFA2T1(ETO) family. *Blood* **91:** 4028–4037.
33. Hildebrand, D., Tiefenbach, J., Heinzel, T., Grez, M., and Maurer, A. B. (2001). Multiple regions of ETO cooperate in transcriptional repression. *J. Biol. Chem.* **276:** 9889–9895.
34. Rae, J., *et al.* (2000). Molecular analysis of 9 new families with chronic granulomatous disease caused by mutations in CYBA, the gene encoding p22(phox). *Blood* **96:** 1106–1112.
35. Sorescu, D., Szocs, K., and Griendling, K. K. (2001). NAD(P)H oxidases and their relevance to atherosclerosis. *Trends Cardiovasc. Med.* **11:** 124–131.
36. Brown, N. S., and Bicknell, R. (2001). Hypoxia and oxidative stress in breast cancer. Oxidative stress: its effects on the growth, metastatic potential and response to therapy of breast cancer. *Breast Cancer Res.* **3:** 323–327.
37. Burdon, R. H., Gill, V., and Alliangana, D. (1996). Hydrogen peroxide in relation to proliferation and apoptosis in BHK-21 hamster fibroblasts. *Free Radic. Res.* **24:** 81–93.
38. Arnold, R. S., *et al.* (2001). Hydrogen peroxide mediates the cell growth and transformation caused by the mitogenic oxidase Nox1. *Proc. Natl. Acad. Sci. USA* **98:** 5550–5555.
39. Jacobson, M. D. (1996). Reactive oxygen species and programmed cell death. *Trends Biochem. Sci.* **21:** 83–86.
40. Griendling, K. K., and Ushio-Fukai, M. (2000). Reactive oxygen species as mediators of angiotensin II signaling. *Regul. Pept.* **91:** 21–27.
41. Reddy, D. E., Keck, C. L., Popescu, N., Athwal, R. S., and Kaur, G. P. (2000). Identification of a YAC from 16q24 carrying a senescence gene for breast cancer cells. *Oncogene* **19:** 217–222.
42. Spataro, V., Norburt, C., and Harris, A. (1998). The ubiquitin-proteasome pathway in cancer. *Br. J. Cancer* **77:** 448–455.
43. Kremmidiotis, G., *et al.* (1998). Localization of human cadherin genes to chromosome regions exhibiting cancer-related loss of heterozygosity. *Genomics* **49:** 467–471.
44. Ewing, B., Hillier, L., Wendl, M. C., and Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8:** 175–185.
45. Staden, R., Beal, K. F., and Bonfield, J. K. (2000). The Staden package, 1998. *Methods Mol. Biol.* **132:** 115–130.
46. Altschul, S. F., *et al.* (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25:** 3389–3402.
47. Callen, D. F., *et al.* (2002). Defining regions of loss of heterozygosity of 16q in breast cancer cell lines. *Cancer Genet. Cytogenet.* **133:** 76–82.
48. Bustin, S. A. (2000). Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J. Mol. Endocrinol.* **25:** 169–193.

Kochetkova, M., McKenzie, O.L.D., Bais, A.J., Martin, J.M., Secker, G.A., Seshadri, R., Powell, J.A., Hinze, S.J., Gardner, A.E., Spendlove, H.E., O'Callaghan, N.J., Cleton-Jansen, A., Cornelisse, C., Whitmore, S.A., Crawford, J., Kremmidiotis, G., Sutherland, G.R. and Callen, D.F. (2002). CBFA2T3 (MTG16) Is a Putative Breast Tumor Suppressor Gene from the Breast Cancer Loss of Heterozygosity Region at 16q24.3. *Cancer Research*, 62(16), 4599-4604.

Cleton-Jansen, A., Callen, D.F., Seshadri, R., Goldup, S., McCallum, B., Crawford, J., Powell, J.A., Settasatian, C., Beerendonk, H., Moerland, E.W., Smit, V.T.B.H.M., Harris, W.H. Millis, R., Morgan, N.V., Barnes, D., Mathew, C.G. and Cornelisse, C.J. (2001). Loss of Heterozygosity Mapping at Chromosome Arm 16q in 712 Breast Tumors Reveals Factors that Influence Delineation of Candidate Regions. *Cancer Research*, 61(3), 1171-1177.

ANKRD11 and *ANKRD12* are novel 9kb genes encoding nuclear proteins with ankyrin domains: screening of the *ANKRD11* gene for involvement in breast cancer.

keywords: chromosome 16, chromosome 18, loss of heterozygosity, nuclear localisation

Jason A. Powell[a], Chatri Settasatian[a], Karen M. Lower[a], Olivia L. McKenzie[a], Saif Zarqa[b], Raman Kumar[b], David F. Callen[c,*]

[a]*Department of Laboratory Genetics, Women's and Children's Hospital, North Adelaide, SA 5006, Australia.*

[b]*Hanson Institute, Institute of Medical and Veterinary Science, Adelaide, SA 5000, Australia.*

[c]*Department of Medicine, University of Adelaide, Adelaide, SA 5000, Australia*

* *Address for correspondence:*

Assoc Prof David F Callen,

Breast Cancer Genetics group,

IMVS, P.O. Box 14, Rundle Mall, South Australia, 5000, Australia.

Phone: (618) 82223145 Fax: (618) 82223217

Email: david.callen@imvs.sa.gov.au

Abbreviations: aa, amino acid(s); bp, base pairs(s); kb, kilobase(s) or 1000 bp; mRNA, messenger RNA; UTR, untranslated region; ORF, open reading frame; RT-PCR, Reverse Transcription Polymerase Chain Reaction; RACE, Rapid Amplification of cDNA End(s); BAC, bacterial artificial chromosome; ESTs, expressed sequence tags; ANK, ankyrin motif; PBL, peripheral blood lymphocytes; SSCA, single strand polymorphism analysis

Abstract

The isolation and characterization of *ANKRD11*, and its chromosome 18 homolog *ANKRD12*, forms the basis of a novel protein family. A direct comparison between *ANKRD11* and *ANKRD12* indicates that they both span large genomic intervals, share a large exon (6,578 and 4,721 base pairs, respectively), have conserved exon/intron structure and encoded predicted proteins of similar size (2,664 and 2,063 amino acids, respectively). *ANKRD11* and *ANKRD12* map to 16q24.3 and 18p11.3, respectively, both regions of the chromosome reported as targets of sporadic breast cancer loss of heterozygosity. Deduced amino acid sequences of the two genes contained three complete ankyrin domains and a number of nuclear localisation signals. Experiments showed that ANKRD11 was indeed a nuclear protein. The *ANKRD11* gene was screened for breast tumour-restricted mutations and in constitutional DNA of patients with breast cancer with the aim to find if this is a familial breast cancer susceptibility gene. Although missense mutations were found, none originated in sporadic breast tumours or had compelling evidence as a cause of familial breast cancer. The possible consequences of these missense mutations will be confirmed when we complete our current investigations of ANKRD11 function.

1. Introduction

Tumour related loss-of-heterozygosity (LOH) is defined by the finding of apparent homozygosity for a polymorphic locus in tumour DNA that is heterozygous in germ-line DNA from the same patient. In each tumour type, chromosome regions can be defined showing consistently elevated LOH frequencies compared with background values. Such regions of high LOH most likely define the locations of tumour suppressor genes since LOH is a mechanism that can result in loss of its normal allele. Extensive breast cancer studies show that the long arm of chromosome 16 frequently undergoes LOH and therefore is the likely location of a breast cancer tumour suppressor gene. The band 16q24.3 could be defined as one of the smallest commonly involved LOH regions based on studies of 712 sporadic breast tumours using a panel of polymorphic markers (Cleton-Jansen et al. 2001). Subsequently, a search was undertaken to identify the putative breast cancer tumour suppressor in band 16q24.3 (Whitmore et al. 1998; Powell et al. 2002). These studies involved construction of physical maps based on cloned genomic DNA, DNA sequencing and identification of genes and transcripts. During this process we identified *ANKRD11* that encodes an unusually large protein.

Several lines of evidence suggested a possible involvement of *ANKRD11* in breast cancer. The predicted protein has three ankyrin (ANK) motifs that are frequently involved in protein-protein interactions. Since the ANK motifs of ANKRD11 show high homology to similar domains of BARD1, a protein known to interact with BRCA1 via their RING domains (Wu *et al*, 1996), it was conceivable that ANKRD11 may interact indirectly with BRCA1. A possible role of *ANKRD11* in cancer was reinforced by an unconfirmed report that an inherited gene for predisposition to breast cancer may be located at 16q24.3 (Giles et al. 1992), and that a matching partial cDNA sequence (GeneBank entry NM_013275) was reported to be down-

regulated in nasopharyngeal carcinoma. Interestingly, nasopharyngeal carcinoma also exhibits 16q LOH (Shao et al. 2001). In this report we present data defining the structure and expression of *ANKRD11* and *ANKRD12*, its homolog at 18p11.3, and investigate the possible involvement of *ANKRD11* in breast cancer.

2. Materials and Methods

2.1 Transcript characterization

In silico analysis utilised a variety of publicly available databases and online tools available from NCBI (National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov), EMBL (European Molecular Biology Laboratory, (http://www.embl-heidelberg.de) and the Sanger Institute (http://www.sanger.ac.uk). Transcripts were then confirmed by RT-PCR amplification of normal breast RNA (BD Biosciences). In general, 3 μg of total RNA or 100 ng of polyA$^+$ mRNA was reverse transcribed with Superscript RNase H$^-$ reverse transcriptase (Invitrogen) and either 50 pmol of random hexamers (Perkin Elmer) or 100 pmol of oligo dT primers. The cDNA products were used for subsequent PCR amplification using primers designed to bind with the different exons of the genes. Transcripts were further extended by 3'-RACE experiments (Invitrogen).

2.2 Northern blot analysis

A human multiple tissue Northern blot membrane (BD Biosciences) was probed according to the manufacture's instructions. RT-PCR products were generated from *ANKRD11* cDNA with the primers 5'GTGGCCCTTCTCATGCAGAT and 5'TGTCACCGTGGGGACAG and from *ANKRD12* with the primers 5'ACTTCGTCACGGTGGAATC and

5'TGACACATGCTTTTCCTTGG. Membranes were hybridized overnight at 65°C, and the following day washed three times in 2xSSC and 1% SDS for 10 min at room temperature. A high stringency wash of 0.1xSSC and 1% SDS at 65°C for 30 min was used when direct monitoring of the filter showed high background signal.

2.3 Cell culture and transient transfections

HEK293T cells were obtained from the American Type Culture Collection (ATCC) and maintained at 37°C in 5% CO_2 in DMEM (Invitrogen) supplemented with 10% fetal calf serum, 2mM L-glutamine and 10mg/liter penicillin and lentamycin. The vector pLN-ANKRD11.myc was generated by cloning a C-terminal myc-tag and the full length ORF of *ANKRD11* into the *Hind*III and *Cla*I sites of the mammalian expression vector pLNCX2 (BD Biosciences). The 7995 bp *ANKRD11* ORF is flanked by the forward primer 5' AGGAGCAGGACGATGCCCCG and the reverse primer 5' ACAAAGTCGTCGTTGACGTC. pLN-ANKRD11.myc was transfected into HEK293T cells with Lipofectamine 2000 (Invitrogen) and the myc-tag detected by sequential incubation with myc monoclonal antibody (Santa Cruz Biotechnology Inc.) and CY3-conjugated goat anti-mouse IgG (Jackson ImmunoResearch Labs). Cell nuclei were counter stained with DAPI and the images merged using CytoVision (Applied Imaging).

2.4 Mutation screening

ANKRD11 was screened for mutations by single strand conformational analysis (SSCA). Screened were 46 DNA samples from breast tumours known to have LOH of 16q (Cleton-Jansen et al. 2001). DNA was only isolated from biopsies with at least 50% tumour. Amplicons showing a SSCA variant were re-amplified with unlabelled primers and sequenced. Normal peripheral blood leukocyte (PBL) DNA from the same patient was also

tested when variants were found in the breast tumour DNA. In addition, 100 PBL DNA samples from familial breast cancer patients were screened. Each of the samples was from a different family consisting of multiple members with breast cancer and where tests for the presence of *BRAC1* or *BRAC2* mutations were negative. These samples were obtained from kConFab (Kathleen Cuningham Foundation Consortium for Research into Familial Breast Cancer, Australia). SSCA variants identified in the patient PBL DNA samples were then screened in a panel of PBL DNAs isolated from normal individuals. For some patients with a germline mutation it was investigated if the same mutation was also present in DNA isolated from their breast tumours. The DNA was isolated from paraffin sections of breast tumours using the DNeasy Tissue Kit (Qiagen).

For SSCA analysis, DNAs were first amplified with Hex fluorescent labelled oligo-primers designed using the intronic sequences flanking each exon, except for exon 9 that was divided into thirty-one 200-300 bp overlapping amplicons. Each amplicon was amplified by PCR in 10 µl reactions containing 2 mM dNTPs, 1.5 mM $MgCl_2$, 50 pmol of each primer, 1 unit of *Taq* polymerase and 30 ng of patient DNA. Amplification involved 10 cycles of 94°C for 1 minute; 55°C for 2 minutes; 72°C for 2 minutes, followed by a further 25 cycles of 94°C for 1 minute; 60°C for 2 minutes; 72°C for 2 minutes, with a final extension of 72°C for 7 minutes. Amplified products were mixed with 10 µl of formamide loading buffer, incubated at 100°C for 5 minutes, placed on ice, and then resolved on 4% acrylamide gels using the GelScan 2000 (Corbett Research).

3. Results and Discussion

3.1 Characterisation of ANKRD11 nucleotide sequence

ANKRD11 was originally identified by exon trapping and sequencing from the 140 kb BAC 561E17 (Whitmore et al. 1998). This BAC was located to a region of 16q24.3 approximately 700 kb from the telomere. A detailed *in silico* analysis was performed to assemble the various overlapping human Unigene clusters and ESTs. As human ESTs for some regions of the gene were not available, homologous mouse ESTs were used to extend the sequence. The assembled 9307 bp full-length sequence of *ANKRD11* was confirmed by RT-PCR and subsequent sequencing of the amplified products. Many ESTs in the database possessed 3' ends that clustered at several long adenosine-rich nucleotide tracts present within the *ANKRD11* sequence resulting in partial *ANKRD11* transcripts being reported in databases. For example, *LZ16*, a nasopharyngeal carcinoma susceptibility gene of 1603 bp was reverse transcribed from the poly(A)-rich tract "AAAAAAATAAAGTGAAAAA" starting at nucleotide 1813 of *ANKRD11*. In addition, the size of the *ANKRD11* transcript exceeds the limits of conventional reverse transcription, further contributing to the database abundance of partial *ANKRD11* transcripts.

ANKRD11 contains an ORF of 7995 bp with a start codon in exon 3 at base position 462 that conforms to the Kozak consensus sequence and is preceded by a sequence of three in-frame stop codons. There is a stop codon in exon 13 and this same stop codon was identified in the corresponding mouse cDNA clone. The homology between these two orthologs significantly decreased 3' of this stop codon, supporting the likelihood that this sequence was the 3' untranslated region (3'UTR). The 3' UTR of 852 bp includes a polyadenylation signal, AATAAA, 13 bp before the polyadenylation site. Alignment of the full-length *ANKRD11* sequence with the 16q24.3 genomic sequence available at NCBI established that *ANKRD11* consists of 13 exons ranging in size from 85 bp to a large exon of 6.58 kb, and spans a genomic interval of greater than 263 kb (Table 1). A large 113 kb intron exists between exons

2 and 3 and all splice junctions conform to the published consensus sequence (Shapiro and Senapathy, 1987).

Northern blot analysis confirmed the transcript size and tissue expression of *ANKRD11* (Fig 1a). A 1050 bp DNA probe that encompassed exons 6 to 9 hybridised to a predominant transcript of approximately 9kb in all tissues examined. The level of signal suggested highest expression in the heart, placenta, kidney and pancreas with lowest expression in lung. Smaller transcripts of 3.2, 3.0 and 1.0 kb with signals lower than the 9kb band were also detected. Additional cDNA probes generated from various parts of the *ANKRD11* transcript also detected the 9 kb band on Northern blots (data not shown). The observed ubiquitous expression profile was consistent with analysis of the appropriate databases with ESTs derived from a variety of tissue cDNA libraries.

In silico analysis of the human EST database at NCBI and RT-PCR experiments were undertaken to define the presence of alternate splicing. There were a large number of apparent ESTs in introns of the *ANKRD11* gene. Many of these were related to simple repeat regions within the introns. Complex splicing variants of *ANKRD11* could be demonstrated by RT-PCR amplification (data not presented). Whether these variants represent the less than 9 kb bands observed in the Northern blots is unknown. There is one Unigene cluster, containing EST AA995728, which is contained within intron 4 and may represent a transcript transcribed in a direction opposite to the *ANKRD11* open reading frame. There was alternative usage of two tandem CAG triplets at the 5' splice site of exon 3 with about half of the ESTs in the database having this triplet missing. Due to their location in the 5'UTR, this polymorphism is unlikely to be of any functional significance.

3.2 Homologous sequences

The *ANKRD11* sequence of 2026 bp between exons 9 and 13 was identical to the assembled genomic sequence derived from Xq27.1. The X chromosome sequence is considered to be a transposed pseudogene as the region of *ANKRD11* homology on the X chromosome is flanked by L1 elements. This pseudogene is not transcribed due to the absence of a promoter sequence.

In silico analysis identified a homologous gene on chromosome 18 at band p11.3 that we have notated as *ANKRD12*. KIAA0874 and GAC-1, entered in the NCBI database as two independent transcripts, are partial transcripts of *ANKRD12*. RT-PCR was used to confirm the full-length *ANKRD12* gene (data not shown). *ANKRD12,* a 9,035 bp gene, consisting of 13 exons, spans approximately 147kb of genomic DNA (Table 1). The 6,189 bp ORF has a start codon in exon 2 at position 215-217 and a stop codon in exon 13 at position 6401-6403. Northern blot analysis with an *ANKRD12* specific cDNA probe identified a 9kb band corresponding to the predicted full-length transcript that was expressed predominantly in the skeletal muscle and placenta (Fig 1b). A 7kb band was also present and is likely to represent the product of an alternative splicing event.

ANKRD11 and *ANKRD12* are of similar size coding for predicted proteins of 2664 and 2062 amino acids respectively, both possess a large exon (6578 and 4721 bp respectively), exhibit high exon/intron structure conservation, and have DNA sequence identity ranging from 22-67%. The sequences for *ANKRD11* and *ANKRD12* have been deposited in GenBank with accession numbers AY373756 and AY373757 respectively.

3.2 In silico analysis of predicted protein

The *ANKRD11* gene encodes a predicted 298kDa protein of 2,664aa. *In silico* analysis identified an ANK repeat domain, a motif that is present in over 400 different proteins and functions as a site for protein-protein interactions (Sedgwick and Smerdon, 1999). The ANK

repeat unit contains 33 aa with two proposed consensus sequences (Lux et al. 1990; Sedgwick and Smerdon 1999) given in Fig 2b. Three complete tandem ANK repeats are present in the N-terminal regions of both the *ANKRD11* and *ANKRD12* proteins. The ANK repeats were highly conserved between the two proteins with 77% identity and 89% similarity, and this conservation extended into the sequences flanking the repeats (Fig 2b). Homology searches showed the ANK repeats of *ANKRD11* and *ANKRD12* were 55-58% identical, and 70-75% similar, to the three tandem ANK repeats of the BARD1 protein. If the amino acid similarity between these ANK repeats reflect functional homology, then this suggests a possible role of *ANKRD11* and *ANKRD12* in breast cancer as the amino-terminal RING domain of *BARD1* interacts with *BRCA1* (Wu et al. 1996).

The *ANKRD11* and *ANKRD12* encoded proteins were predicted by PROSITE to possess eight and three bipartite nuclear localization signals, respectively, suggesting nuclear localization of both proteins. Using the PESTfind program, PEST sites (potential proteolytic cleavage sites) were identified; six in ANKRD11, four of which were located in a 152 aa region at the C-terminus, and a single PEST sequence in ANKRD12. Proteins with PEST domains can be rapidly degraded via the ubiquitination pathway (Rechsteiner and Rogers, 1996). No other domains could be detected in either of the proteins. The predicted domains on the protein coded for by *ANKRD11* are shown in Fig. 2(a).

3.3 Protein localisation

To confirm the predicted sub-cellular localization of *ANKRD11*, the coding region was amplified from cDNA generated from human breast total RNA and cloned into a mammalian expression vector with a 3' in-frame myc peptide tag. HEK 293T cells were transfected with a plasmid expressing an ANKRD11-myc protein and 24 hours later the recombinant protein was detected with anti-c-myc monoclonal antibodies. ANKRD11.myc displayed nuclear

speckle localization, with the number of speckles varying from 5 to 15 per nucleus (Fig 3). Double staining experiments with other known nuclear speckled proteins (POD and SnRNAs) did not show any co-localisation with the ANKRD11.myc speckles. The observed localisation is consistent with that predicted from the presence of multiple localisation signals in the protein.

3.4 Screening for mutations in sporadic breast tumours.

The *ANKRD11* coding sequence, together with the surrounding splice sites and short regions flanking the introns, were screened for mutations in 46 sporadic breast tumour DNA samples with known chromosome 16q LOH that included band 16q24.3 (Table 2). No mutations were found that were restricted to the tumour DNA. However, twelve different mutations were found to be apparently homozygous in tumours but heterozygous in the constitutional DNA from the same patient. This observation is a consequence of tumour LOH for this region of chromosome 16. Seven of the twelve mutations were silent polymorphisms as they did not cause an amino acid change and were present in unaffected individuals. Five mutations resulted in an amino acid change and four of these were also present in unaffected individuals. Of particular interest was the mutation of cytosine to thymidine at position 1353, positioned 70 bp 3' of the ANK domains, and absent from 330 chromosomes from unaffected individuals. This mutation resulted in an amino acid change from threonine to methionine (Table 2) and the threonine residue is conserved in both the human *ANKRD12* homolog and the mouse ortholog of *ANKRD11*. The possible consequence of this mutation will be confirmed when we complete our current investigations of ANKRD11 function.

3.5 Screening for mutations in familial breast cancer

Evidence suggested that *ANKRD11* might be involved in familial breast cancer. As the ANK domains of the ANKRD11 protein showed significant homology to those of BARD1, a BRCA1 interacting protein, ANKRD11 might have a role in the *BRCA1* pathway. Furthermore, an unconfirmed report of linkage analysis in familial breast cancer identified 16q24.3 as a disease candidate region (Giles et al., 1992). Approximately 50% of the *ANKRD11* gene was screened with the regions selected based on the SSCA variants found in the tumour screening. These included the part of the gene containing the ANK domains and a major proportion of exon 9. SSCA of the DNA samples detected a number of variants that were polymorphisms as they were also present in normal individuals.

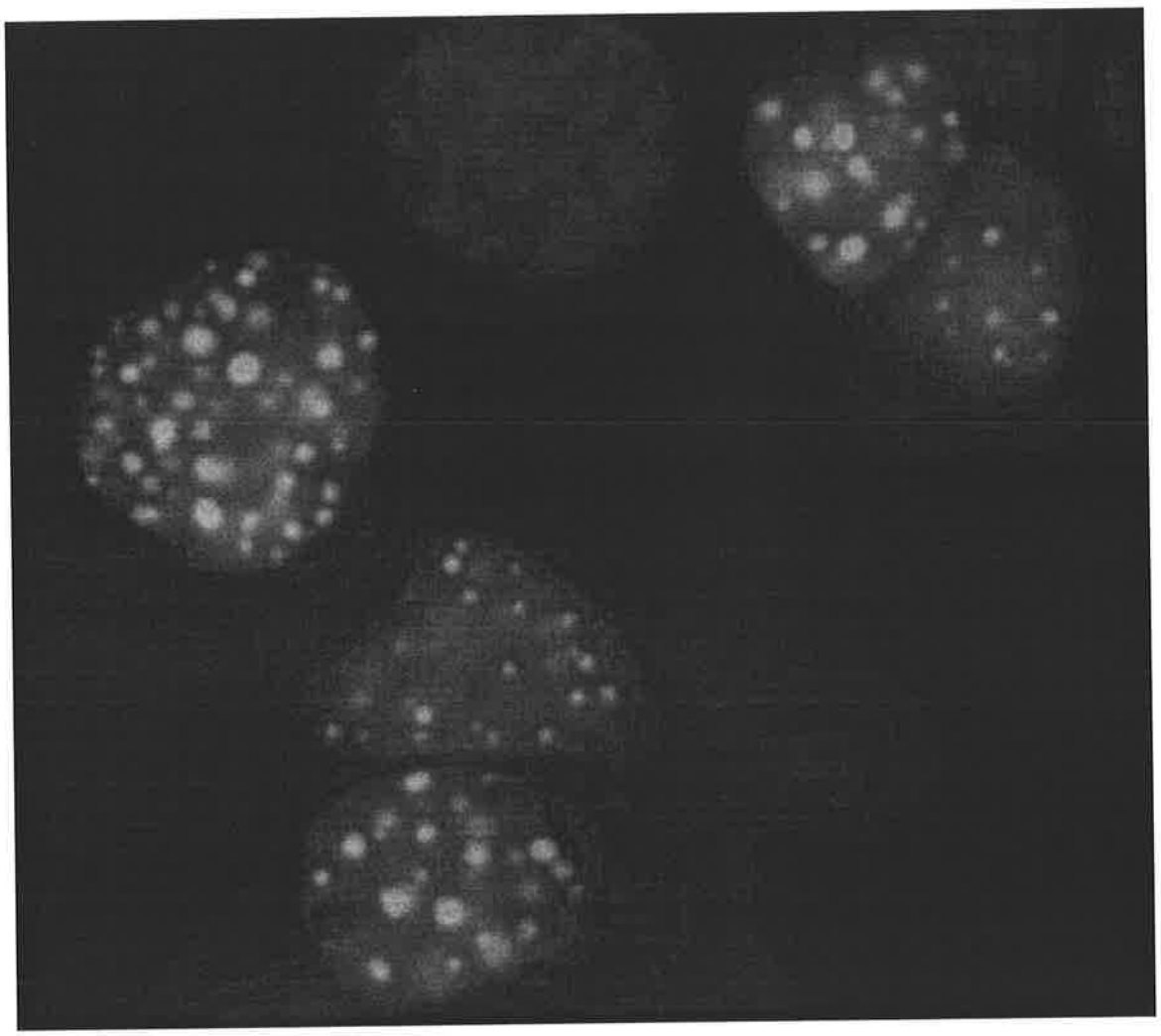
A heterozygous mutation was detected in one patient at nucleotide 5972 that resulted in a CCG triplet coding for proline being mutated to TCG that codes for serine, a non-conserved amino acid change. The same mutation was found in two additional females with breast cancer from this pedigree. This mutation was not detected in 400 chromosomes from unrelated normal individuals. DNA was extracted from paraffin sections of breast tumour derived from one of these individuals. Subsequent PCR amplification and sequencing established that the tumour was CCG, suggesting that the mutated TCG allele was lost by tumour restricted LOH. Therefore, it is unlikely that this mutation at nucleotide 5972 is related to the risk of breast cancer as, contrary to our expectation; the normal allele was retained in the tumour.

3.5 Conclusion

The isolation and characterization of *ANKRD11*, and its chromosome 18 homolog *ANKRD12*, forms the basis of a new novel protein family. A direct comparison between *ANKRD11* and *ANKRD12* indicates both span large genomic intervals, share a large exon

(6,578 and 4,721 bp respectively), have conserved exon/intron structure and encode predicted proteins of similar size (2,664 and 2,063 amino acids respectively). Because of their atypical gene structure and size, only partial sequences of these two genes are available in public databases. This is likely to be a consequence of both the large mRNA that exceeds the limits of conventional reverse transcription and the presence of poly(A) tracts within *ANKRD11* and *ANKRD12* mRNAs that permit internal oligo-dT primer binding in reverse transcription. In addition, the genes have not been recognised by *in silico* gene prediction programs as the sizes of both the transcriptional units and genomic regions exceed the standard parameters used by such programs. *ANKRD11* and *ANKRD12* map to 16q24.3 and 18p11.3, respectively. The 16q24.3 band is a target of LOH in sporadic breast cancers (Cleton-Jansen et al. 2001). Interestingly, chromosome 18p11.3 has also been shown to exhibit significant LOH of 68% in sporadic breast tumours (Kittiniyom et al. 2001).

The amino acid sequences deduced from the *ANKRD11* and *ANKRD12* open reading frames have three complete ANK domains and a number of nuclear localisation signals. Nuclear localisation of ANKRD11 was confirmed. Since there was evidence that *ANKRD11* was involved in breast cancer, SSCA was used to screen the *ANKRD11* gene for breast tumour restricted mutations. The possibility that *ANKRD11* is a familial breast cancer susceptibility gene was investigated by screening for mutations in constitutional DNA of patients with breast cancer. Although missense mutations were found, none originated in sporadic breast tumours or had compelling evidence as a cause of familial breast cancer. The possible consequences of these missense mutations will be confirmed when we complete our current investigations of ANKRD11 function.



References

Cleton-Jansen, A. M., D. F. Callen, R. Seshadri, S. Goldup, B. McCallum, J. Crawford, J. A. Powell, C. Settasatian, H. van Beerendonk, E. W. Moerland, V. T. Smit, W. H. Harris, R. Millis, N. V. Morgan, D. Barnes, C. G. Mathew and C. J. Cornelisse (2001). Loss of heterozygosity mapping at chromosome arm 16q in 712 breast tumors reveals factors that influence delineation of candidate regions. Cancer Res 61: 1171-7.

Giles, R., D. F. Callen, G. R. Sutherland and M.-C. King (1992). Preliminary test of linkage of chromosome 16q markers to breast cancer in families with older onset disease. abstract in "Report of the Second International Workshop on Human Chromosome 16 Mapping". Cytogenet Cell Genet 60: 174.

Kittiniyom, K., K. M. Gorse, F. Dalbegue, J. H. Lichy, J. K. Taubenberger and I. F. Newsham (2001). Allelic loss on chromosome band 18p11.3 occurs early and reveals heterogeneity in breast cancer progression. Breast Cancer Res 3: 192-8.

Lux, S. E., K. M. John and V. Bennett (1990). Analysis of cDNA for human erythrocyte ankyrin indicates a repeated structure with homology to tissue-differentiation and cell-cycle control proteins. Nature 344: 36-42.

Powell, J. A., A. E. Gardner, A. J. Bais, S. J. Hinze, E. Baker, S. Whitmore, J. Crawford, M. Kochetkova, H. E. Spendlove, N. A. Doggett, G. R. Sutherland, D. F. Callen and G. Kremmidiotis (2002). Sequencing, transcript identification, and quantitative gene expression profiling in the breast cancer loss of heterozygosity region 16q24.3 reveal three potential tumor-suppressor genes. Genomics 80: 303-10.

Rechsteiner, M. and S. W. Rogers (1996). PEST sequences and regulation by proteolysis. Trends Biochem Sci 21: 267-71.

Sedgwick, S. G. and S. J. Smerdon (1999). The ankyrin repeat: a diversity of interactions on a common structural framework. Trends Biochem Sci 24: 311-6.

Shao, J. Y., X. M. Huang, X. J. Yu, L. X. Huang, Q. L. Wu, J. C. Xia, H. Y. Wang, Q. S. Feng, Z. F. Ren, I. Ernberg, L. F. Hu and Y. X. Zeng (2001). Loss of heterozygosity and its correlation with clinical outcome and Epstein-Barr virus infection in nasopharyngeal carcinoma. Anticancer Res 21: 3021-9.

Shapiro, M. B. and P. Senapathy (1987). RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acids Res 15: 7155-74.

Whitmore, S. A., J. Crawford, S. Apostolou, H. Eyre, E. Baker, K. M. Lower, C. Settasatian, S. Goldup, R. Seshadri, R. A. Gibson, C. G. Mathew, A. M. Cleton-Jansen, A. Savoia, J. C. Pronk, A. D. Auerbach, N. A. Doggett, G. R. Sutherland and D. F. Callen (1998). 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 50: 1-8.

Wu, L. C., Z. W. Wang, J. T. Tsan, M. A. Spillman, A. Phung, X. L. Xu, M. C. Yang, L. Y. Hwang, A. M. Bowcock and R. Baer (1996). Identification of a RING protein that can interact in vivo with the BRCA1 gene product. Nat Genet 14: 430-40.

Table 1. Comparative gene structures of *ANKRD11* and *ANKRD12*.

ANKRD11			*ANKRD12*		
Number	Exon (bp)	Intron (bp)	Number	Exon (bp)	Intron (bp)
1	317	71932	1	163	45416
2	85	113150	2	138	13031
3	147	11588	3	148	8777
4	142	14019	4	69	4113
5	171	184	5	147	2780
6	204	1952	6	201	4973
7	143	2341	7	143	4951
8	148	389	8	148	32211
9	6578	3878	9	4721	4858
10	99	135	10	99	11635
11	144	3909	11	144	3881
12	93	2153	12	96	1296
13	1037	-	13	2818	-
Total	9,307	234,669	Total	9,035	137,922

Table 2. Screening *ANKRD11* for mutations in sporadic breast tumours.

Location of change (exon/bp position)	Frequency in tumours	Constitutional nucleotide triplet	Tumour nucleotide triplet	Amino acid change	Frequency in normal chromosomes
8/1353	1/46	A(C/T)G	ATG	Thr->Meth	0/330
9/3375	2/46	G(T/C)G	GCG	Val->Ala	2/100
9/5802	1/46	(G/A)CC	ACC	Ala->Thr	1/330
9/6546	4/46	(C/G)CC	GCC	Pro->Ala	7/100
9/7290	3/46	(C/T)CG	TCG	Pro->Ser	4/200

Legends for Figures.

Figure 1. Expression of *ANKRD11* and *ANKRD12* in various tissues.

Human multiple tissue northern blot membranes were hybridised to *ANKRD11*- (a) and *ANKRD12*- (b) specific probes. Y axis indicates size of band (kb).

Figure 2. Nuclear localisation of *ANKRD11* protein.

HEK293T were transfected with a plasmid expressing the ANKRD11 protein. The merged image shows the nuclei stained blue and the speckled distribution in the nucleus of the red stained ANKRD11 protein.

Fig. 3. Structure of ANKRD11 protein (a) and comparison of its amino acid sequence with ANKRD12 (b).

Consensus 1 ankyrin domain, Lux et al (1990). Consensus 2 ankyrin domain, Sedgwick and Smerdon (1999). Identical amino acids are highlighted in grey. Amino acids matching with one of the consensus sequences are in bold. P, PEST sequences; N, bipartite nuclear localisation signals; A, ankyrin domains.

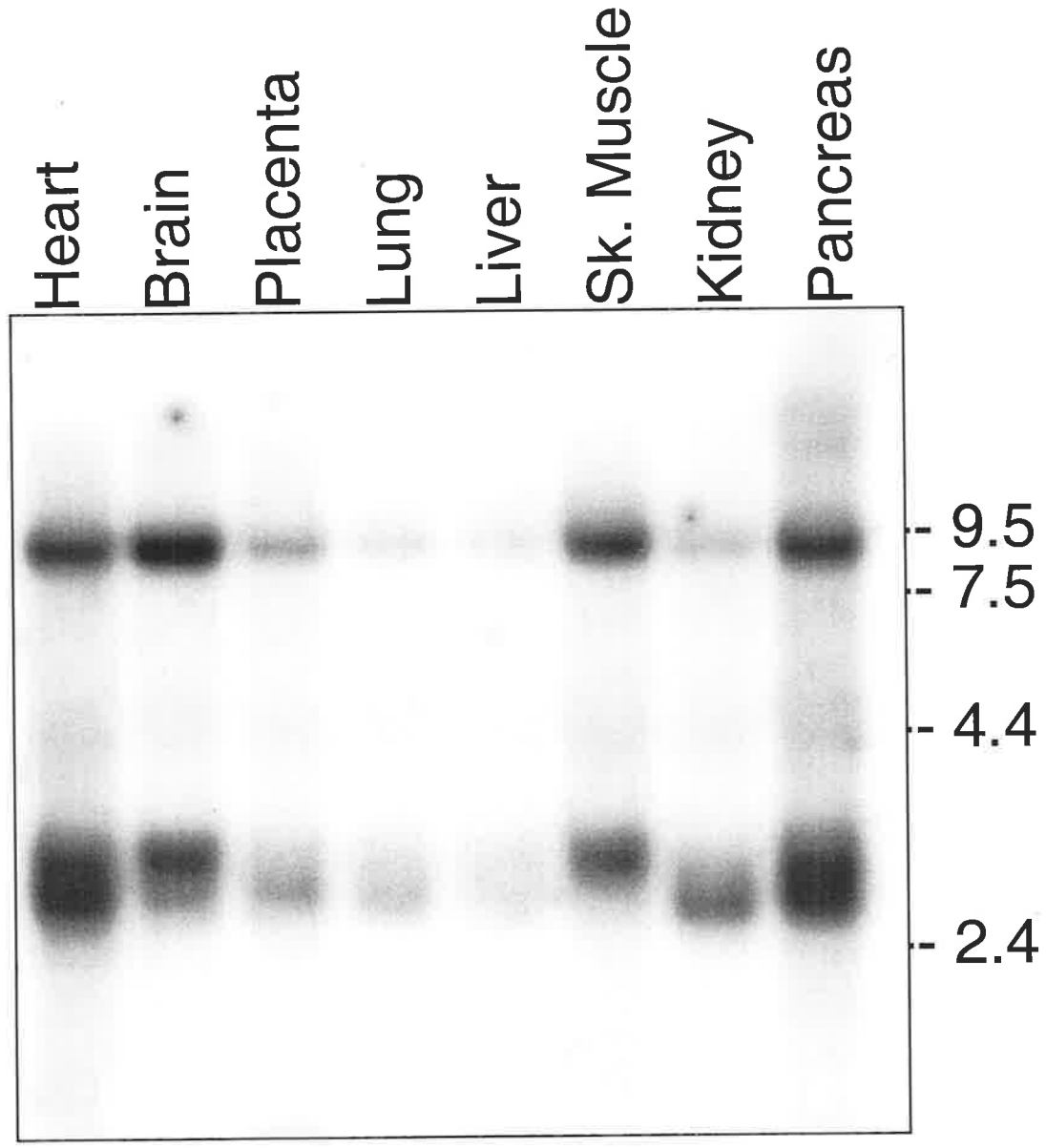

Heart
Brain
Placenta
Lung
Liver
Sk. Muscle
Kidney
Pancreas
9.5
7.5
4.4
2.4

Figure 1a

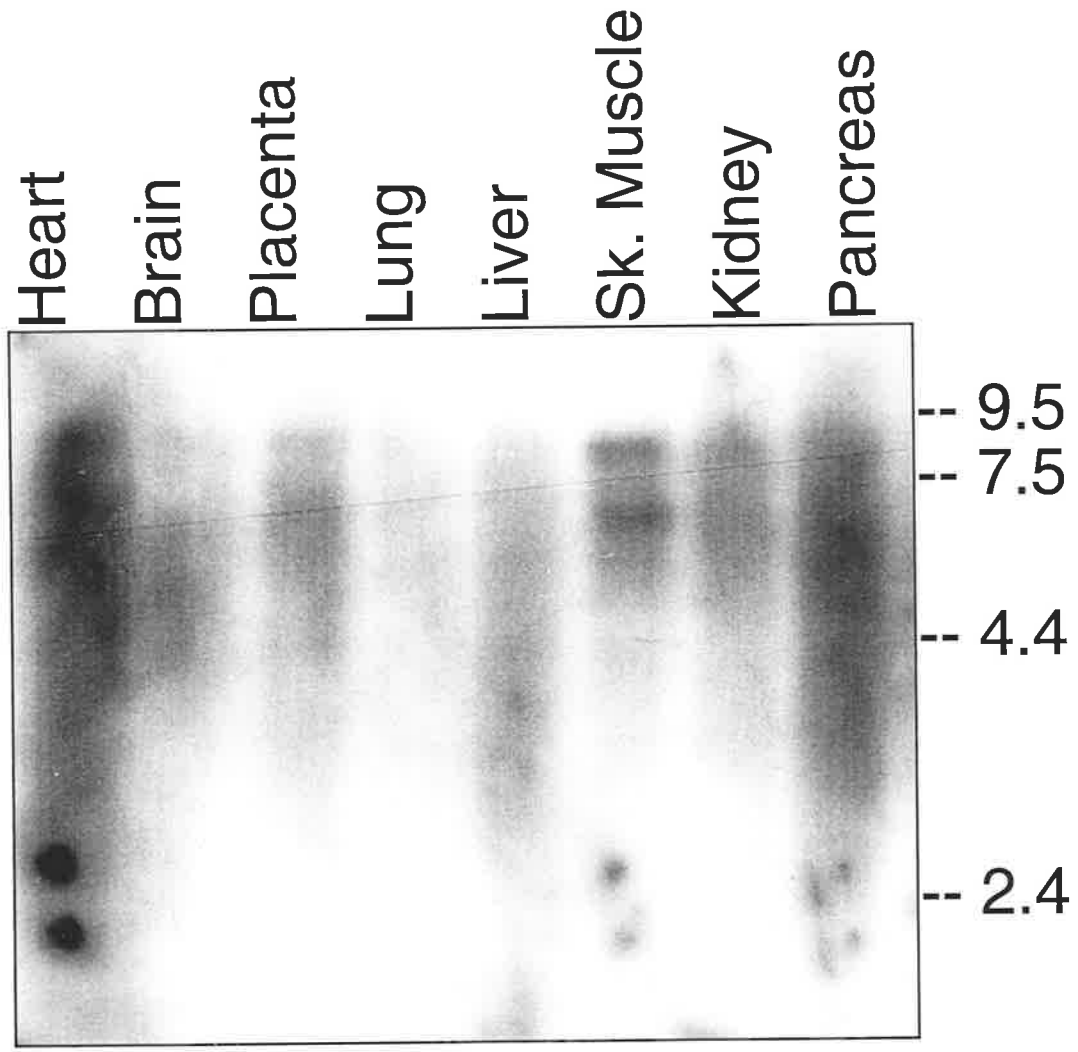
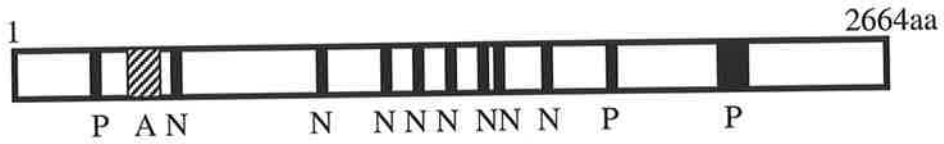

Heart
Brain
Placenta
Lung
Liver
Sk. Muscle
Kidney
Pancreas
-- 9.5
-- 7.5
-- 4.4
-- 2.4

Figure 1b

Figure 3.

(a)



(b)

```
ANKRD11    135aa  [T][K][S][VC]                ASKTKDKVNK RNE
ANKRD12    149aa  [S][N][GT][PA]               TSRQKDKVNK RNE

ANKRD11           RGETRLHRAA IRGDARRIKE LISEGADVNV KDF
ANKRD12           RGETPLHMAA IRGDVKQVKE LISLGANVNV KDF

ANKRD11           AGWTALHEAC NRGYYDVAKQ LLAAGAEVNT KGL
ANKRD12           AGWTPLHEAC NVGYYDVAKI LIAAGADVNT QGL

ANKRD11           DDDTPLHDAA NNGHYKVVKL LLRYGGNPQQ SNR
ANKRD12           DDDTPLHDSA SSGHRDIVKL LLRHGGNPFQ ANK

ANKRD11           KGETPLKVAN SPTMVNLLLG KGTYTSSEES STE  300aa
ANKRD12           HGERPVDVAE TEELELLLKR EVPLSDDDES YTD  314aa

Consensus 1       -G-TPLH-AA --GH---V-- LL--GA--N- ---
Consensus 2       -G-TPLHLA- --G---VV-L LL--GADVNA -D-
```