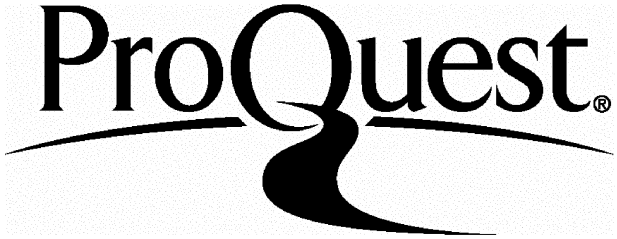
IDENTIFICATION AND CHARACTERISATION OF CANDIDATE TUMOUR SUPPRESSOR GENES FROM CHROMOSOME 13q14.3, AN AREA OF FREQUENT DELETION IN PATIENTS WITH B-CELL CHRONIC LYMPHOCYTIC LEUKAEMIA.

A thesis submitted for the degree of Doctor of Philosophy by

Clare Rowntree

Department of Academic Haematology
University College Medical School
Royal Free Campus
Pond Street
London NW3 2QG

November 1999



Abstract

B-cell chronic lymphocytic leukaemia (B-cell CLL) is a malignancy of circulating B lymphocytes characterised by a clonal expansion of CD5+ B cells. The aetiology of this disease is unknown. The commonest structural cytogenetic abnormality seen in B-cell CLL is deletion of chromosome 13q13.4 and it is likely that a tumour suppressor gene lies within this deleted region.

The aim of the work described in this thesis was to define the region of minimal deletion at chromosome 13q14.3 in our patients with B-cell CLL and to then isolate and characterise candidate tumour suppressor gene cDNAs from this genomic region.

Using known markers from the area, a physical map was constructed of the deleted region. 52% of our patients were shown to have deletion of 13q14.3 when tested by Southern blotting techniques for loss of markers from this region.. The region of minimal deletion in these patients was shown to be a maximum of 450kb. A putative exon, TA 6.35, was isolated by exon trapping techniques from this minimally deleted area. By screening a peripheral leucocyte cDNA library with this putative exon a candidate tumour suppressor gene cDNA was isolated. Using further cDNA library screening techniques and RACE PCR to characterise this cDNA, a second candidate tumour suppressor gene cDNA was also isolated from the region of deletion.

The first cDNA, clone 1, consisted of 9 exons including the original TA 6.35 exon. The exons of clone 1 span a genomic distance of over 450kb with exons 2, 3 and 4 lying within our minimal region of deletion. The second cDNA, clone 2:2 consisted of 3 exons. The third exon of this clone also lies within our minimal region of deletion. The first two exons of clone 1 were shared by clone 2:2. Both clones were demonstrated to be expressed in normal and CLL lymphocytes by RT PCR analysis. cDNA clone 2:2 had a postulated open reading frame encoding for 78 amino acids. cDNA clone 1 was shown to exist in many alternatively spliced forms, non of which have a long open reading frame. It was postulated that clone 1 may not encode for a peptide but may act as an RNA regulating expression of cDNA clone 2:2. When patients with heterozygous deletion of 13q14.3 were analysed for mutation of either clone 1 or 2:2, the majority did not have mutations demonstrated within the retained allele. However, the promoter regions of these transcripts are as yet unidentified and we suggest that, therefore, they remain candidate tumour suppressor genes from 13q14.3.

Acknowledgements

Firstly, I would like to thank my supervisor, Dr Letizia Foroni for her guidance and support throughout my PhD. In addition to her role as supervisor, she also helped with some of the practical work involved in this thesis, in particular the work described in sections 3.2, 4.3, 4.5 and 6.7.

Secondly, I would like to thank Dr Veronique Duke for her continued cheerfulness and for her practical support with the genomic organisation and mutational analysis of cDNA clone 1 described in chapter 6. I would also like the following for their specific help; Tina Butler for the FISH analysis of our YAC clones, John Yaxley and Giulio Palmisano for their assistance with the sequencing of cDNA clone 1 and the 6kb genomic fragment described in chapter 4, Matthew Gordon for all the lymphocyte T cell depletion and Panos Panayiotidis for the initial identification of the exon TA 6.35 and cDNA clone 1.

A huge thank you to all in the lab who tried against all odds to keep me sane and who never tired of listening; Amanda Jackson, Clair Gricks, Mary Papaioannou and Steve Hart. I am grateful to Stan Wickremasinghe for his advice and critical comments on the writing of this thesis. Thank you also to Professor Hoffbrand for his continued interest and support.

Finally I would like to thank my family and Tony for their continued love and care throughout these past few years, and for their faith in me to complete what seemed to be an impossible task.

Table of contents

List of Figures

List of Tables

Chapter 1. Introduction

1.1. The genetics of cancer

Cancer is a genetic disease, arising from an accumulation of genetic abnormalities that promote selection and survival of cells with increasingly aggressive behaviour (Fearon, 1997). Most cancers are clonal in that they arise as a result of genetic mutation within one cell. This mutation may either occur as an acquired somatic mutation specific to the malignant clone or as an hereditary genetic lesion present in all cells of the body (1% of cancers) (Fearon, 1997). Studies of these rare hereditary syndromes have provided much of the evidence that shows cancer to be a genetic disease.

Cancer occurs as a result of an imbalance of cell homeostasis. All normal tissues must be in balance, whether they are actively growing, dividing, differentiating or resting. This homeostasis is maintained by the rate of cell growth being balanced by the rate of cell death and cell differentiation. The same cellular mechanisms regulate this balance in all cells, from developing organs during embryogenesis to non-dividing tissues such as neurones. In addition to remaining in balance, all tissues must faithfully replicate their DNA without introducing error during cell division (Sherr, 1996). Any cell that contains genetic error must repair the damage prior to replication or must enter a self destruct mode i.e. programmed cell death (apoptosis). Loss of regulation of this balance may lead to uncontrolled cell proliferation and is the hallmark of cancer. Two groups of genes that are important in the regulation of cell homeostasis are the proto-oncogenes and the tumour suppressor genes. Their roles in the normal regulation of the cell cycle and how loss of this regulation can lead to cancer will be discussed below.

1.1.1 The cell cycle

The cell cycle is responsible for the production of two progeny from one parent cell and can be divided into four phases (see figure 1.1). Chromosomes are faithfully replicated during S phase. Identical chromosomal copies are distributed equally to daughter cells during mitosis (M phase). Two gap (G) phases occur between the S and the M phase. G_1 precedes S phase and G_2 precedes M phase. A cell can exit the cell cycle after M phase and enter a resting phase, G_0. Cell growth can occur continuously throughout all stages of the cell cycle. Tissues with

rapid turnover e.g. bone marrow and gut epithelium, will be continuously passing through the stages of the cell cycle whereas tissues that rarely divide e.g. neurones, may be mostly in G_0. The machinery for DNA replication and chromosomal segregation is protected within the cell from external influences, as damage to its components could be potentially fatal to the cell (Sherr, 1996). However, the elements controlling entry and exit to G_1 are responsive to extracellular signals. The decision to divide occurs at the G_0/G_1 transition, after which time a cell becomes largely refractory to extracellular signals again (Sherr, 1996).

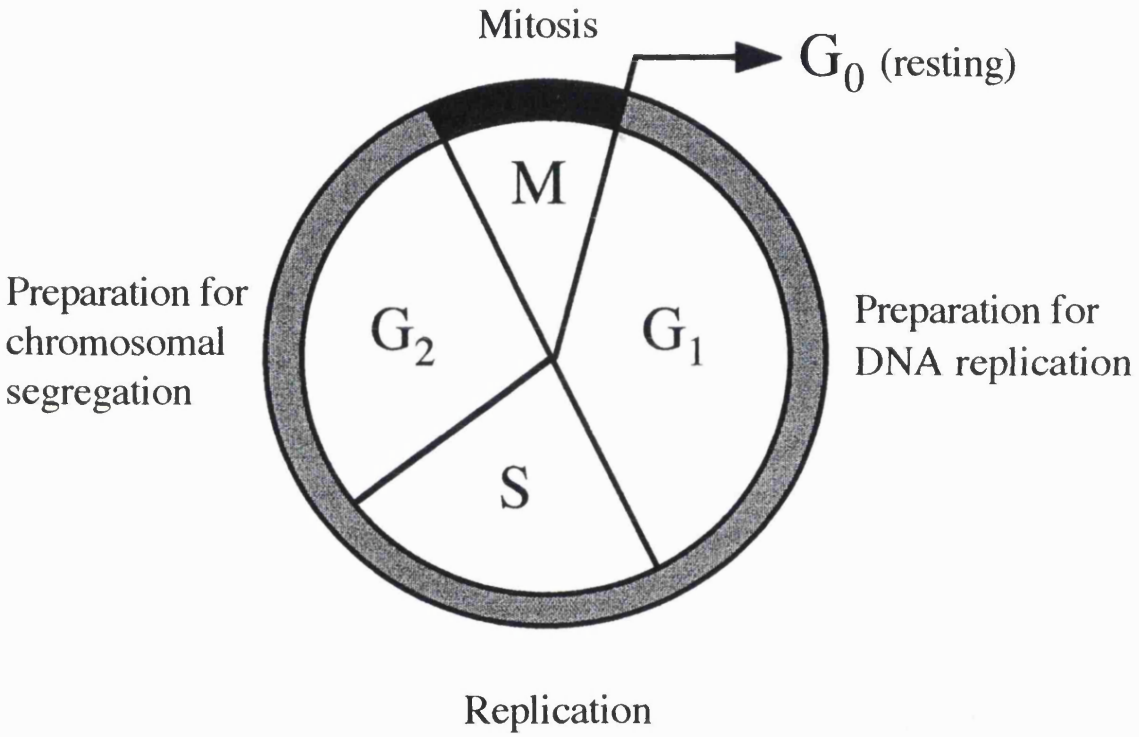


Figure 1.1: Phases of the cell cycle.

Two important checkpoints exist within the cell cycle. These are not transitions between phases through the cell cycle, but are points where 'surveillance operations' by the cell occur. They are not essential to the workings of the cell but prevent the cell continuing to cycle in the face of stress or damage (Sherr, 1996). Mutations in genes whose products influence these checkpoints are often found in cancers (Vogelstein and Kinzler, 1998). The first checkpoint is the point in G_1 when the cell becomes committed to DNA synthesis, and the second is at the G_2/M boundary. The checkpoint in late G_1 is called the restriction point and is where the cell decides between continued proliferation and escape from the

cycle into G_0. Passage through the restriction point into S phase is controlled by a group of proteins known as cyclin dependent kinases (CDK). These CDKs are positively and negatively regulated by cyclins. Unphosphorylated retinoblastoma protein (RB 1) prevents cells from entering S phase and promotes G_1 exit by inhibiting the E2F-DP1 transcription factor which triggers expression of several genes critical for cell cycle transit. If the cell is to continue into S phase, these CDKs will inactivate RB 1 by phosphorylation, allowing cell cycle progression and blocking G_1 exit (Sherr, 1996). Many cancer cells abandon control at the restriction point and remain continuously in cell cycle. Control of the cell cycle and its checkpoints are critical in malignant progression of all tumours. It is thought that mutations in genes that control these checkpoints could lead to clonal evolution of cancer cells by allowing the accumulation of genetic errors that would normally have resulted in cell cycle arrest or cell death.

1.1.2. Apoptosis

Apoptosis (programmed cell death) is the means by which tissues can eliminate unwanted cells without initiating an inflammatory response. Apoptosis occurs in all tissues and plays an important role in the homeostatic balance between cell proliferation and cell death. It is particularly important in the removal of cells that have acquired deleterious genetic changes that may lead to the development of cancer. The gene products controlling apoptosis are themselves normally tightly controlled with both positive and negative pathways being equally important. These pathways are highly responsive to extracellular signals (Vogelstein and Kinzler, 1998). Abnormalities occurring in genes controlling these processes may lead to a failure of apoptosis, thus promoting immortalisation of a cell and its progeny - an important step towards the malignant phenotype. An example of such a gene is the proto-oncogene *bcl*-2, identified as the gene on chromosome 18q21 involved in the t(14;18) chromosomal translocation found in the majority of follicular lymphomas (Cleary *et al.*, 1986). Juxtaposition of *bcl*-2 to the heavy chain locus results in upregulation of its expression. *In vitro* studies show that upregulation of *bcl-2* does not increase the proliferative potential of cells, but *bcl*-2 transformed growth factor dependent cells survive for prolonged periods without the usual growth factor requirement (Vaux *et al.*, 1988). In

addition *bcl*-2 has been demonstrated to act as an inhibitor of apoptosis, co-operating with *c-myc* to immortalise pre-B cells *in vitro* (Vaux *et al*, 1988). This was the first evidence that inhibition of apoptosis can contribute to malignant transformation.

1.1.3. Genetic mutation and cancer

Three classes of genes that are fundamental to the status quo of a cell have been identified and aberrations within these genes can cause progression to cancer.

The first of these are the oncogenes. These genes usually play crucial roles in cell signalling, cell cycle or apoptosis. Inappropriate expression of these genes results in increased cell growth or decreased cell death. Mutations in oncogenes may lead to them becoming constitutively active with consequent deregulated cell control. Alternatively, translocation of these genes to loci containing strong transcriptional enhancer elements e.g. immunoglobulin or T cell antigen receptor genes, results in their augmented expression.

The second class of genes to be implicated in cancer development is the tumour suppressor genes. These encode for negative regulators of the cell cycle or promoters of apoptosis. Loss of these genes due to genetic alteration can also result in uncontrolled cell proliferation.

The third important class of genes in cancer genetics are those whose products are involved in DNA repair following damage, for example due to ionising radiation or UV damage.

Within any given tumour, mutations may have occurred in more than one class of gene. For example, many colorectal carcinomas have been demonstrated to have up to 10 genetic changes involving 5 different genes (Levine, 1993), including tumour suppressor genes and oncogenes. Each of these individual groups of genes involved in cancer genetics will be discussed in more detail.

1.1.4. Oncogenes

Oncogenes were originally identified as retroviral genes that were capable of transforming mammalian cells. Their normal functional counterparts were then discovered in mammalian genomes and were termed proto-oncogenes. These genes have diverse roles in cell signalling pathways. They are not oncogenic in their

normal state but can become so, if deregulated or mutated (Haber and Fearon, 1998). Proto-oncogenes show marked evolutionary conservation and are usually promoters of cell growth or blockers of apoptosis (Levine, 1993). There is great variation between the products of these genes, all of which are elements of cellular signalling networks, ranging from extracellular ligands and growth factors, through cytoplasmic protein kinases and GTP-binding proteins, to nuclear transcription factors (Hunter, 1991).

A single mutational event in an oncogene is often sufficient to deregulate its expression or function with consequent abnormal cell growth. This gain of function acts in a dominant fashion to the wild type allele, as shown by transfection experiments using mutated gene constructs (see review (Ponder, 1988)). Such mutations are rarely inherited as they are likely to be fatal in embryogenesis (Levine, 1993). Rearranged RET is an inherited cause of thyoid carcinoma.

Two classes of oncogene have been described based on the effect that their altered expression has on cells in tissue culture (Hunter, 1991). The first are genes that act to immortalise the cells. The second group produce an altered phenotype within the cell, such as reduction in growth factor requirement, changes in cell shape and anchorage-independent cell growth. There is evidence that within a malignant clone, several such oncogenes with differing effects on cell control co-operate to induce cancer (Hunter, 1991).

1.1.5. Mechanisms of activation of oncogenes

Oncogenes tend to be activated by one of three mechanisms. Firstly, a point mutation within the gene may enable its protein product to function in an uncontrolled manner e.g. the *ras* genes. These are 21kD guanine nucleotide binding proteins encoded by three genes known as *k-ras*, *h-ras* and *n-ras*. They play critical roles in the transduction of cellular proliferation signals. Point mutations in each of these *ras* genes have been shown to lead to cancer e.g. *n-ras* is mutated in some cases of melanoma (Perucho *et al.*, 1981) and specific mutations of *k-ras* or *n-ras* are found in 58% of colorectal adenomas larger than 1cm and 47% of colorectal carcinomas (Vogelstein *et al.*, 1988).

Secondly, chromosomal translocations can activate oncogenes (Barr, 1998). Translocation as a mechanism of gene activation is frequently seen in the leukaemias and lymphomas (Rabbitts, 1994). Whilst mutation of the same oncogene can be seen in many tumour types, specific translocations are often

characteristic of a particular leukaemia or lymphoma (Rabbits, 1994; Barr, 1998). In the acute leukaemias translocations usually alter expression of transcription factor genes (Look, 1997). For example, the oncogene AML 1 is fused to the transcriptional control gene ETV6 via the t(12;21)(p13;q22) translocation seen in 25% of children with B lineage acute lymphocytic leukaemia. Many similar examples exist (reviewed by Rabbitts (Rabbitts, 1994)). Chronic myeloid leukaemia is characterised by the translocation t(9;22)(q34;q11), fusing the oncogenes *bcr* and *abl.* Activation of *bcl-2* via t(14;18)(q32,q21) similarly characterises follicular lymphoma (discussed above). This upregulation of *bcl-2* is probably not sufficient in itself for lymphoma to occur, but immortalises the cell rendering it susceptible to further transforming mutations which are, as yet unknown (Rabbitts, 1991). A similar consequence is postulated for the translocation t(14;19)(q32;q13.1), occasionally seen in B-cell CLL, which juxtaposes the oncogene *bcl-3* to the heavy chain locus, driving excess production of the *bcl-3* protein (Ohno *et al.*, 1990). However, this translocation is only seen in a minority of CLL tumours and does not characterise the disease.

Thirdly, gene amplification can result in activation or deregulation of oncogenes (Vogelstein and Kinzler, 1998). Oncogenes encoding for growth factor receptors and signal transduction elements are frequently the target of gene amplification in cancer cells. Gene amplification may occur as a result of duplication of the entire chromosome, but is more often due to the cell amplifying the oncogene as part of an 'amplicon' within the normal chromosomal number. The amplified gene may be located some distance from the original copy and multiple copies may be present. Rearrangements of the gene do not exist in this situation, but expression of the product is high due to the transcription of mRNA from the multiply copied gene. The mechanism of amplicon production is not well understood. In the promyelocytic leukaemia cell line, HL60, 8-30 copies of the *c-myc* oncogene have been identified (Collins and Groudine, 1982). Cancers with oncogene activation due to gene amplification have a particularly poor prognosis.

1.1.6. Tumour suppressor genes

Loss of genetic material within cancer cells characterises the tumour suppressor nature of some genes (Ponder, 1988). Tumour suppressor genes encode proteins that negatively control cell proliferation or promote apoptosis. Both alleles of a tumour suppressor gene need to be inactivated within a single cell

for malignant potential to develop. As discussed later, many of the mechanisms of malignancy in CLL are thought to arise due to the inactivation of tumour suppressor genes.

Three types of evidence support the idea that alterations in genes whose products negatively regulate cell expansion (tumour suppressor genes) can cause malignancy; somatic cell hybrids, studies of familial cancers and loss of heterozygosity studies.

Somatic cell hybrids

Somatic cell hybrids, generated by fusing tumourigenic cells with non tumourigenic cell lines, have shown reversal of the tumour phenotype in a variety of cancers. This suggests that the addition of a normal complement of chromosomes reverses the genetic defect within these tumours. The relevant chromosomes were identified in some cancer cells using micro cell experiments, where the tumour cell was fused with somatic cells containing only one human chromosome. In this way the tumour suppressor gene involved in retinoblastoma was located to chromosome 13 and the tumour suppressor gene involved in Wilm's tumour was located to chromosome 11 (reviewed (Levine, 1993; Marshall, 1991; Weinberg, 1991). These experiments led to the concept that these tumour cells had lost the function of a critical gene giving rise to malignant potential which could be restored by a single wild type allele.

Familial cancers

Over 20 years ago, it was noted that the childhood tumour of the retina, retinoblastoma, occurred in infants in 40% of cases (Knudson, 1971). In the majority of these infants the tumour was bilateral and some of these children had a family history of the disease. In contrast, the other 60% of cases occurred in older children with no family history. In these cases the tumours were usually unilateral. Knudson postulated that some children inherited a mutant allele of a critical gene in the germline i.e. present in all cells. They then acquired a second mutation giving rise to malignant change and ultimately cancer. Because all cells carried the original mutation, it was feasible that more than one cell would develop a second mutation, explaining the bilateral cases seen. The older children with disease, he postulated, acquired both abnormalities in the same somatic cell giving

rise to unilateral disease of later onset. It became clear that this hypothesis was correct and the second mutation occurred in the remaining allele of the same gene. Knudson's theory has now become known as the 'two hit' hypothesis. The retinoblastoma gene (RB 1) has since been isolated from chromosome 13q14.3 (Friend *et al.*, 1986). 40% of children with the disease have small deletions or point mutations in the RB 1 gene in their germline. The second allele was shown to be inactivated in the tumours, usually with a large deletion of 13q14.3 (Ponder, 1988). Studies have shown that introduction of normal RB 1 cDNA into cultivated tumour cells reverses the tumourigenic properties of the cells (Knudson, 1993). These loss of function mutations are acting in a recessive fashion to the wild type allele. RB-1 has become the paradigm for tumour suppressor genes and Knudson's two hit theory.

Interestingly, 100% of children with germline mutations of RB 1 will develop retinoblastoma in infancy. If they survive, they may develop osteosarcoma later in life but no other tumours are seen at an increased frequency. This suggests that other tissues have mechanisms for compensating for the loss of the RB 1 gene product (Knudson, 1993). However, many adult carcinomas have detectable, biallelic defects of RB 1, including the majority of small cell lung carcinomas (Weinberg, 1991), suggesting that abnormalities of this gene acquired later in life in certain tissues may be detrimental.

Other tumour suppressor genes have been found through similar studies of hereditary cancers e.g. the Wilms' tumour gene (Call *et al.*, 1990). The tissue restriction associated with the hereditary RB 1 defect is also seen in other hereditary cancer syndromes (Weinberg, 1991).

p53 is another well characterised tumour suppressor gene located on chromosome 17p13 that encodes a 53-kD nuclear phosphoprotein. Its pivotal role in cell cycle regulation is discussed below. Mutation of p53 in association with cancer was first described in the rare autosomal dominantly inherited disorder, Li-Fraumeni syndrome (LFS). Individuals with LFS develop early onset sarcomas and cancers. Many of these patients have germline mutations of p53 (Malkin *et al.*, 1990). p53 was subsequently shown to be mutated in approximately 50% of cancers (Vogelstein and Kinzler, 1998) and was originally thought to be the crucial point in the pathogenesis of the majority of cancers. However, many of these mutations have been shown to be acquired by the cancer cell as the disease progresses rather than contributing to the original malignant phenotype e.g. blast

acceleration of chronic myeloid leukaemia. The role of p53 in the progression of B-cell CLL is discussed below.

Loss of heterozygosity

Genetic polymorphisms exist within both coding and non coding genomic sequences that allow differentiation between two alleles of the same chromosome within the same cell. Techniques such as restriction fragment length polymorphisms and microsatellite markers (discussed later) can be used to identify individuals that carry polymorphic sequences on opposing alleles. Using these tools, investigators noticed that some cancer tissues in these 'informative cases' have lost this heterozygosity when compared to normal tissues from the same individual (Ponder, 1988). Loss of heterozygosity (LOH) is seen in many tumours and has been particularly well documented in colorectal cancer (Weinberg, 1991). LOH experiments added to the body of evidence suggesting that genetic loss in certain tumours could be consistent with inactivation of genes that regulate cell function i.e. tumour suppressor genes.

LOH may occur because a deletion of genomic material has removed one allele. Alternatively, evolving tumour cells with one mutant copy of a tumour suppressor gene may remove the remaining wild type allele in a variety of ways during mitosis (Vogelstein and Kinzler, 1998). The cell may retain a duplicated copy of the abnormal chromosome and eliminate the wild type allele through non disjunction of chromosomes. Alternatively, non disjunction may result in the cell retaining only one copy of the mutant allele. Recombination with the wild type allele, replacing the normal copy of the gene with the mutant copy during mitosis has also been described. These abnormalities at mitosis occur at frequencies as high as 10^{-3} to 10^{-4} per cell generation. Conversely, the frequency of an independent mutation in the retained allele of the gene is 10^{-6} per cell generation (Weinberg, 1991). Consequently, many tumours that lack functional copies of a tumour suppressor gene have used LOH to eliminate the wild type allele and display two identically mutated alleles.

Recently LOH has been shown in 'normal' colonic mucosa of patients with colorectal cancer (Cui *et al.*, 1998). The investigators suggest that LOH may be one of the pre-malignant events in the development of cancer and that, in the case of colonic cancer, it may be a sign of impending malignancy.

LOH studies are useful for suggesting the presence of a tumour suppressor gene on a given chromosome. However, LOH may extend over large genomic distances and, as a consequence, these studies alone are not sufficient to pin point the location of the gene involved.

1.1.7. Mechanisms of inactivation of tumour suppressor genes

There are several mechanisms by which a cell may inactivate a tumour suppressor gene.

Point mutations may disrupt the open reading frame of a gene i.e. a missense mutation, or they may disrupt the promoter region thereby altering expression of the gene. Mutations of CpG islands (CpG repeats associated with the promoter region in 60% of genes) account for 30% of point mutations within genes and are commonly seen in cancers (Robertson and Jones, 1997). For example, 50% of the mutations inactivating the p53 gene in colonic cancer occur in the associated CpG site (Greenblatt *et al.*, 1994). Point mutations may also disrupt an intron-exon splice site of a gene, preventing the RNA processing machinery from assembling the correct mRNA molecule. 15% of point mutations that result in human disease cause an RNA splicing defect (Cooper and Mattox, 1997), e.g. intermittent porphyria (Llewellyn *et al.*, 1996) and Glanzmann thrombasthenia (Jin *et al.*, 1996). As already discussed, loss of genetic material is the hallmark of a tumour suppressor gene and both large and small deletions affecting such genes are frequently reported in human cancers (Ponder, 1988).

Rarely, an insertion of DNA has been documented to disrupt human disease genes, including tumour suppressor genes. L1 retrotransposon elements are repetitive elements that make up 15% of the human genome, a minority of which are capable of retrotransposition and can 'jump' around the genome (Sassaman *et al.*, 1997). Disruption of tumour suppressor genes such as the adenomatous polyposis cancer gene (APC) in colonic cancer and the breast cancer 2 gene (BRCA 2) have been documented to cause tumours due to L1 transposition in a small number of cases (Kazazian and Moran, 1998).

Another mechanism used by cancer cells to alter tumour suppressor gene expression is DNA methylation. Methylation of the genome (presence of 5-methylcytosine) is required for mammalian development but can make the genome unstable (Robertson and Jones, 1997). As discussed, 60% of human genes including all 'house keeping' genes, are associated with CpG rich islands, usually at

their promoter region (Bird, 1986). These islands of CpG repeats are characteristically unmethylated and there is evidence that transcription of genes is inhibited if these CpG islands become methylated. Much of the remaining genome, however, is normally methylated. This state is often reversed in cancers, including CLL, with hypomethylation of the genome occurring (Feinberg and Vogelstein, 1983; Wahlfors *et al.*, 1992). This widespread hypomethylation may be associated with hypermethylation at CpG islands (Robertson and Jones, 1997), a mechanism used by a malignant cell as a means of 'switching off' expression of a tumour suppressor gene. This method of inactivation of tumour suppressor genes has been demonstrated with several genes, including the RB 1 gene and the breast cancer 1 gene (BRCA 1) (Robertson and Jones, 1997).

Methylation is also important in cancer cells in relation to a group of genes that are imprinted. Imprinting is an epigenetic phenomenon where one copy of a gene, either on the paternal chromosome or the maternal chromosome, is silenced during development resulting in the expression of only one allele in each cell (Vogelstein and Kinzler, 1998). This silencing probably occurs by methylation. An increasing number of genes have been described that show either maternal or paternal imprinting, and alterations in imprinting patterns have been described in cancers. For example, the insulin-like growth factor II (IGF2) is a maternally imprinted gene on chromosome 11p15. Relaxation of this imprinting, by hypomethylation, causing expression of the gene from both alleles, has been shown to occur in many tumours including 70% of Wilms' tumours (Ogawa *et al.*, 193; Vogelstein and Kinzler, 1998). It is likely that the excess production of this growth factor resulting from loss of imprinting contributes to the malignant phenotype of the tumour. H19 is another imprinted gene located close to IGF2 on chromosome 11p15. H19 is paternally imprinted and has been shown to have tumour suppressor activities (Hao *et al.*, 1993). Interestingly, in Wilms' tumour, the loss of imprinting by methylation that allows expression of both IGF2 alleles also silences both H19 alleles contributing further to tumour progression (Feil and Kelsey, 1997) Loss of imprinting has been demonstrated in many cancers e.g. loss of imprinting of IGF2 in chronic myeloid leukaemia (Randhawa *et al.*, 1998), and it is probably one of the commonest genetic abnormalities occurring in cancer (Vogelstein and Kinzler, 1998).

1.1.8. Tumour suppressor genes and the cell cycle

The prototype of tumour suppressor genes, found to be inactivated in at least 50% of tumours, is the p53 gene located on chromosome 17p13. p53 is an archetypal cell cycle checkpoint regulator (Sherr, 1996). Normal cells will not replicate a damaged chromosome, but will stop in G_1 phase of the cell cycle to repair the damage (figure 1.1). This critical point in the cell cycle is controlled by the product of the tumour suppressor gene, p53. Exactly how DNA damage is detected by the cell is not known. However, a protein kinase acting upstream of p53 known as ATM has been shown to be activated in response to toxic agents such as ionising radiation. It is thought that upregulated ATM in turn activates p53 by phosphorylation (Canman *et al.*, 1998). p53 then acts as a transcriptional activator, upregulating the expression of genes such as the CDK inhibitor $p21^{CIPI}$, that cause cell cycle arrest (Sherr, 1996). If the DNA damage is irreparable, p53 will then induce apoptosis of the cell, possibly by upregulation of pro-apoptotic proteins such as bax (reviewed (Wickremasinghe and Hoffbrand, 1999)) (figure 1.2A).

Cancer cells have also developed less direct mechanisms that contribute to the inactivation of the p53 pathway. For example, oncogenes such as human papilloma virus 6 antagonise the p53 product, preventing cell cycle arrest and DNA repair (Sherr, 1996) (see figure 1.2A). MDM2 protein is an oncogene product that neutralises p53 and increased production of this protein is seen in certain tumours (Evan and Littlewood, 1998). Preventing p53 dependent apoptosis appears to be key to tumourigenesis (Sherr, 1996) and it has been proposed that most, if not all cancers, will have lesions that ultimately affect this pathway.

As already discussed, the product of the RB 1 gene acts at the restriction point of the cell cycle. The RB protein exists in a phosphorylated and a unphosphorylated form. The unphosphorylated form blocks passage from G_1 to S phase in the cell cycle, by complexing with transcription factors such as E2F (Sherr, 1996). This allows the cell to enter G_0 resting phase. Phosphorylation is mediated via the CDKs by cyclins such as cyclin D1. Phosphorylation of the RB protein results in its dissociation from E2F and consequently removes the cell cycle block at this critical point. This pathway appears to be disrupted by different mechanisms in many human cancers (Sherr, 1996). Loss of the RB gene via chromosomal deletion is the obvious way that a cell can disrupt this pathway,

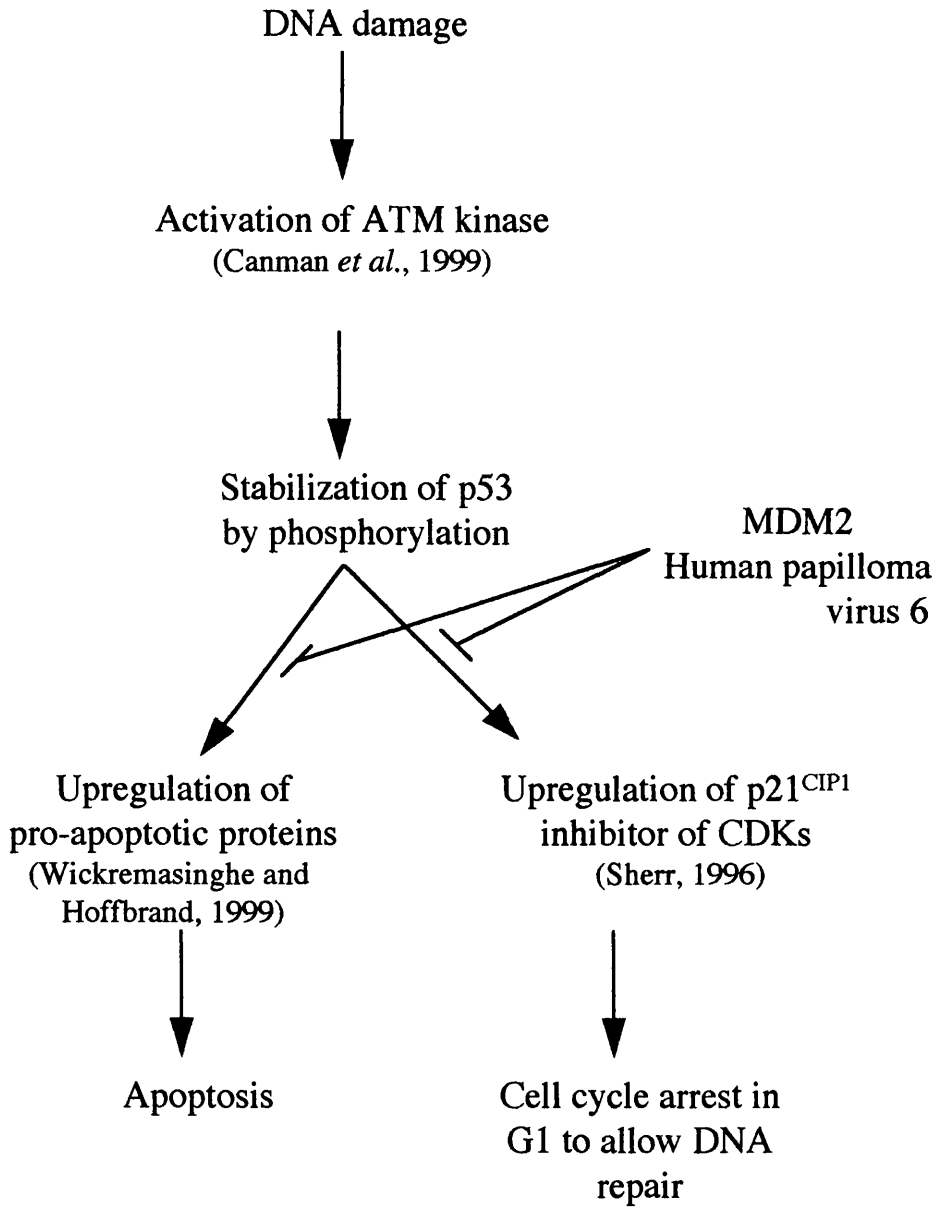
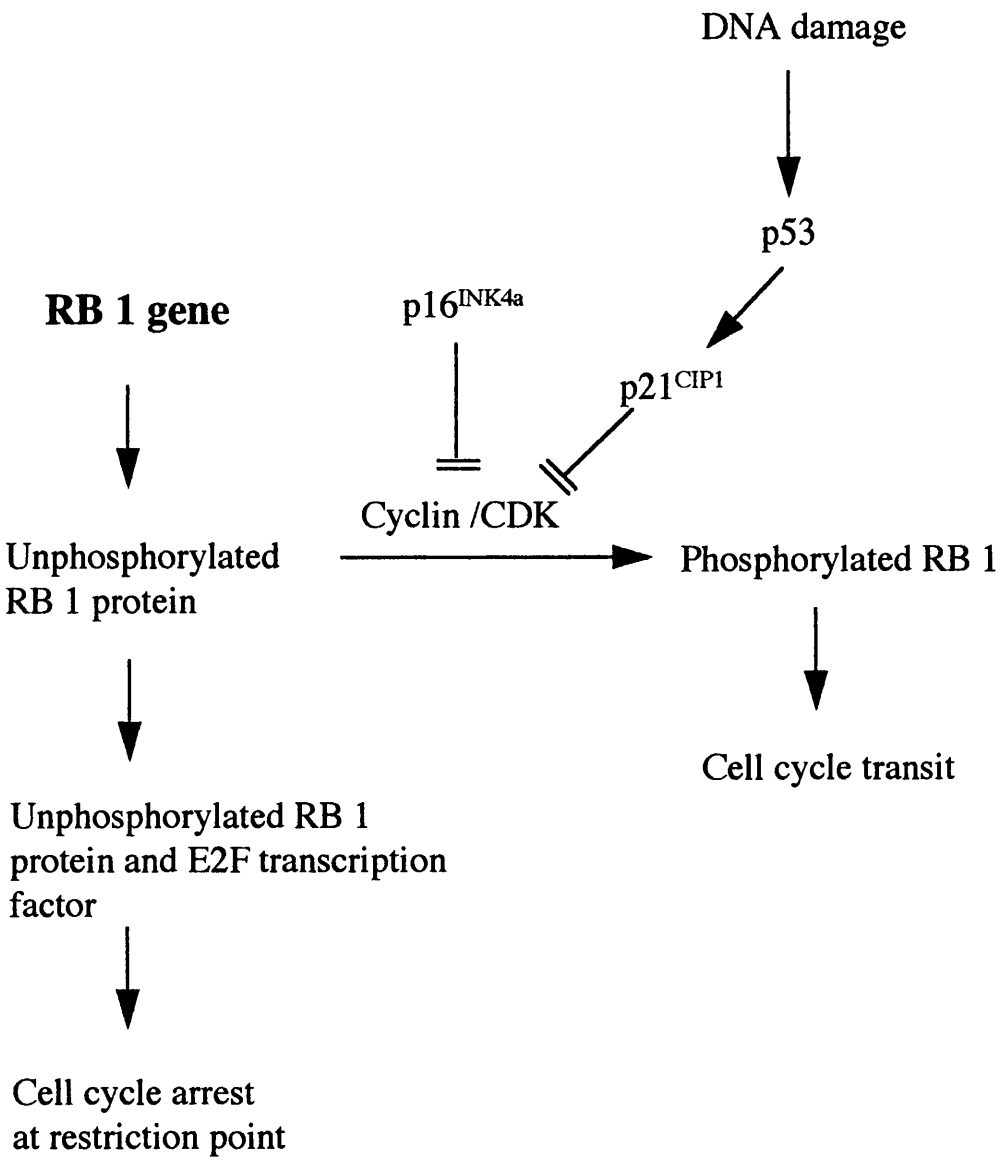


Figure 1.2A: Schematic of the p53 pathway and some of its inhibitors



(Sherr, 1996)

Figure 1.2B: Schematic of the RB 1 pathway and some of the positive and negative pathways acting upon it, including the connection with the p53 pathway.

and this occurs in many tumours. However, because other gene products converge on this pathway other mechanisms have also been implicated. For example, the gene $p16^{INK4a}$ on 9p21 normally inactivates cyclin D1, preventing phosphorylation of RB 1 (see figure 1.2B). $p16^{INK4a}$ acts as a tumour suppressor gene and it is inactivated in familial melanoma and other diverse sporadic cancers. Overexpression of cyclin D1, causing phosphorylation of RB 1, is seen in tumours such as breast cancer and parathyroid tumours (Vogelstein and Kinzler, 1998). Thus the RB1 functional pathway is disrupted through various components in many human cancers. Interestingly when the pathway is inactivated by alterations to genes other than RB 1, no mutations of RB 1 are found concomitantly. Mutations affecting the same biochemical pathway are thought to be mutually exclusive as they offer no advantage to the cell in progressing to overt malignancy (Vogelstein and Kinzler, 1998).

1.1.9. DNA repair genes

Three rare autosomal recessive syndromes are associated with a nucleotide excision repair defect: xeroderma pigmentosum, Cockayne syndrome and a photosensitive form of trichothiodystrophy. All patients with these syndromes show extreme sensitivity to sunlight and patients with xeroderma pigmentosum (XP) have more than a thousand fold increase in sunlight induced cancers. The other two syndromes are not associated with an increased risk of cancer. The nucleotide excision repair pathway repairs many structurally unrelated DNA lesions, including those arising due to toxins and UV light. Patients with XP provided some of the first evidence that cancer could arise as a result of a cell not being able to repair DNA damage.

Hereditary, non polyposis colorectal cancer is an autosomal dominant condition characterised by an increased risk of cancer of the colon, endometrium, ovary, stomach and some other epithelial organs. Patients have an 80% risk of developing colorectal cancer. These patients have an unusual form of somatic mutation consisting of insertions or deletions of the simple repetitive elements that make up microsatellite sequences (microsatellite instability) (Ionov *et al.*, 1993; Thibodeau *et al.*, 1993). Investigators showed that increasing microsatellite instability significantly correlated with tumours of the proximal colon and proposed that these cancers develop through this unique form of genomic instability.

A DNA mismatch repair gene, hMSH2, located on chromosome 2p was found to be mutated in a proportion of these patients with hereditary non polyposis (Fishel *et al.*, 1993; Leach *et al.*, 1993). Other mutated DNA mismatch repair genes have since been found in the remaining patients with this syndrome. As with other hereditary tumour syndromes, one abnormal copy of these genes is inherited in the germline. Due to the nature of the gene product, once the second allele is inactivated, genomic mutations accrue rapidly.

Microsatellite instability has now been reported in a number of hereditary and sporadic cancers. These mutations may not be related to tumour pathogenesis but may merely be a marker of the increasing genomic instability due to defective DNA repair (Perucho, 1996). However, there is some evidence that in certain instances these mutations may occur within other DNA repair genes or oncogenes/tumour suppressor genes and so speed up the accumulation of mutations and consequent tumour progression (Perucho, 1996).

1.2. B-cell chronic lymphocytic leukaemia

1.2.1. Epidemiology and description

B-cell chronic lymphocytic leukaemia (CLL) accounts for 25% of leukaemias seen in adults in the United Kingdom and is the commonest leukaemia in the western world. The annual incidence in the USA is slightly lower than that of plasma cell myeloma but, due to its indolent nature, there are more people alive in western societies with B-cell CLL than any other leukaemia at any given time (Rai *et al.*, 1996).

B-cell CLL is rare before the age of 40 years and is increasingly common with age. The male:female ratio is 2:1. Environmental factors e.g. ionising radiation, drugs and chemicals, do not appear to influence the incidence of CLL. In fact, CLL is the only major adult leukaemia not to show an increased incidence following the atomic bomb explosion at Hiroshima.

Genetic factors do appear to be important in this disease. CLL is rare in Asia with the annual incidence in Korea being 1.5% of that seen in the USA. Japanese and Chinese immigrants to America continue to have a low incidence of B-cell CLL (Rai *et al.*, 1996) showing that this difference is due to hereditary and not environmental factors. In addition, there are reports of familial clustering of CLL with the disease showing anticipation i.e. children of affected patients

present at a mean age of approximately 22 years earlier than their parents (Yuille *et al.*, 1998).

B-cell CLL is characterised by a proliferation and accumulation of small B lymphocytes (see figure 1.3). Using surface immunoglobulin studies and immunoglobulin gene rearrangement analysis this population of B cells has been shown to be monoclonal (Hoffbrand and Pettit, 1994). The cells are mostly in the G_0/G_1 phase of the cell cycle (figure 1.1) and are non proliferating. The majority of the clinical features seen in B-cell CLL are related to the accumulation of these lymphocytes. Patients may have 50-100 times the normal lymphoid mass in the blood, bone marrow, spleen, lymph nodes and liver. Presentation is usually with widespread, symmetrical lymphadenopathy with the nodes being rubbery and non tender. Splenomegaly is also a common finding, especially with advancing disease. As a result of bone marrow infiltration, anaemia and thrombocytopenia may be present. Patients may complain of symptoms related to this haematological failure e.g. dyspnoea, tiredness and easy bruising. Immunological failure is a hallmark of CLL with both humoral and cellular immune processes gradually failing. Neutropenia occurs as a result of bone marrow infiltration and, as a consequence, frequent bacterial and fungal infections are seen in these patients. There is also an increased incidence of herpes zoster in B-cell CLL. The immunobiology of CLL will be discussed in more detail below.

1.2.2. The accumulation of lymphocytes is due to a failure of apoptosis

In 1967, Dameshek suggested that the increased numbers of malignant B cells seen in B-cell CLL was due to the long life span of these cells (Dameshek, 1967). It is now known that a failure of apoptosis in these cells causes the incessant build up of this tumour. The cause for this failure of apoptosis in CLL cells is not known at the present time. However, abnormalities of some of the proteins that control apoptosis have been documented. Over 80% of B CLL cells express high levels of the anti apoptotic protein, bcl-2 (Kitada *et al.*, 1998). The mechanism for this remains unclear, as only rarely can it be attributed to an abnormality involving the *bcl-2* locus (see below). In addition to the high levels of bcl-2, CLL cells also express high levels of anti apoptotic proteins bcl-X_L and mcl-1, whilst expressing low levels of the pro-apoptotic protein bax. High levels of mcl-1 have been correlated with failure to reach remission after single agent

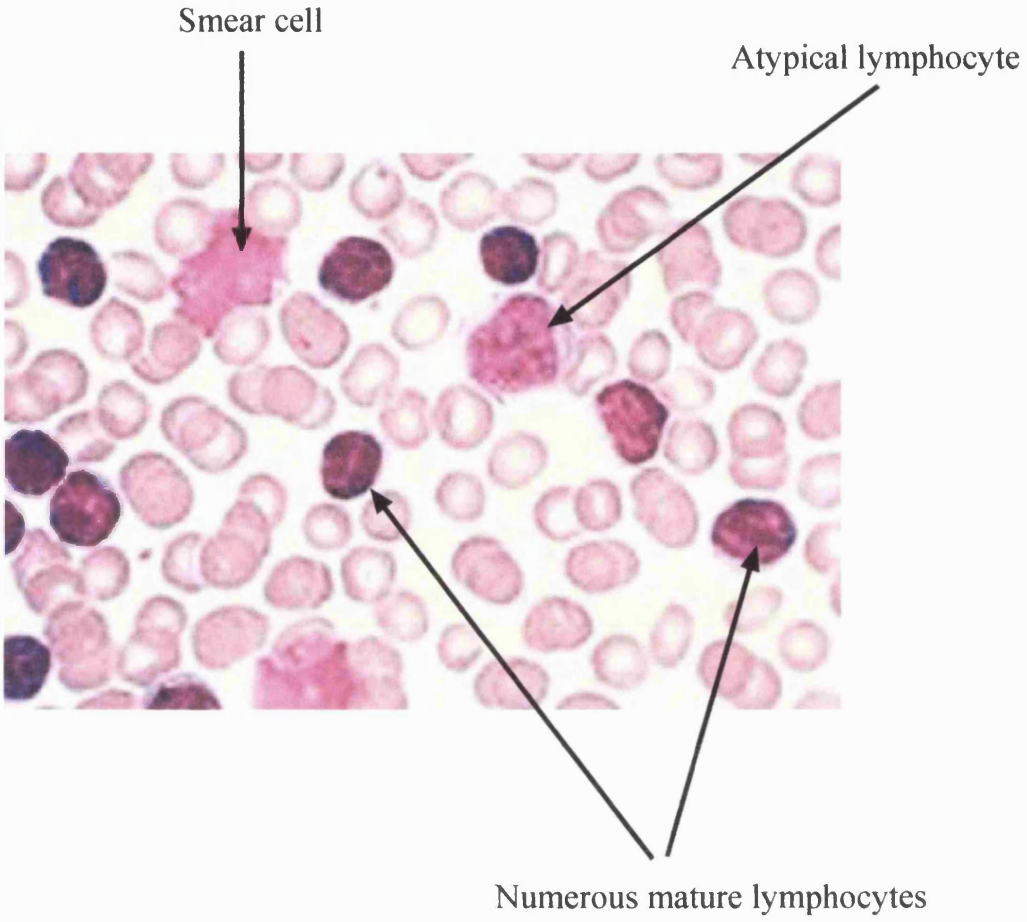


Figure 1.3: A peripheral blood smear from a patient with B-cell chronic lymphocytic leukaemia. There is an increase in small, mature lymphocytes. Note the two smear cells caused by breakage of the cells when the film was made. There is also an atypical lymphocyte seen in the middle of the figure.

chemotherapy (Kitada *et al.*, 1998). The ratio between bcl-2 and bax appears to be an important factor in B-cell CLL development.

1.2.3. Diagnosis of B-cell CLL

Diagnosis is based on the laboratory finding of a lymphocytosis. The National Cancer Institute-Sponsored Working Group (NCI-WG) published the following guidelines on diagnosis of B-cell CLL (Cheson *et al.*, 1996)

1). The absolute lymphocyte count must be > 5 x 10^9/l (may be as high as 300 x 10^9/l). Morphologically, mature lymphocytes must be seen in the peripheral blood.

2). The phenotype of the blood lymphocytes should reveal the following characteristics;

i). predominance of B cells that are CD19+, CD20+ and CD23+ as well as being CD5+ (in the absence of other T cell markers)

ii). the B cell population should be monoclonal as shown by light chain restriction

iii). surface immunoglobulin (sIg) expression should be low.

3). Bone marrow examination is not required if the above criteria have been met, but is advised when the lymphocyte count is low. The aspirate must show >30% of all nucleated cells to be lymphoid with the bone marrow showing normal or hyper-cellularity. Prognostically, the pattern of lymphoid infiltration is helpful. Diffuse involvement correlates with progressive or advanced disease, whilst nodular or interstitial involvement predicts a better prognosis.

1.2.4. Clinical staging and prognostic features of B-cell CLL

The Rai staging system was first presented in 1975 (Rai *et al.*, 1975). It proposed staging patients into five groups; stage 0- bone marrow and blood lymphocytosis; stage I- lymphocytosis with enlarged nodes; stage II- lymphocytosis with enlarged liver or spleen or both; stage III- lymphocytosis with anaemia and stage IV- lymphocytosis with thrombocytopenia. Analysis at this time of 125 patients showed the following median survival figures for each stage (in months); stage 0 - >150, stage I - 101, stage II - 71, stage III - 19 and stage IV - 19. Increasing age was also shown to be a poor predictor for survival. Binet *et al.* proposed a new prognostic classification in 1981, where they defined

three prognostic groups as follows (Binet *et al.*, 1981). Group A - no anaemia, no thrombocytopenia and less than three involved physical sites; group B - no anaemia, no thrombocytopenia with three or more sites involved; group C - anaemia (Hb <10g) and / or thrombocytopenia (platelets <100,000mm^3). The classification system now proposed by the NCI-WG is a modification of the two classifications (Cheson *et al.*, 1996). Three stages are now used (instead of the original 5); low risk group - only lymphocytosis; intermediate risk group - lymphocytosis and enlarged nodes, with or without an enlarged liver / spleen; and the high risk group - lymphocytosis with anaemia (Hb <11g) and / or thrombocytopenia (platelets <100,000mm^3). The median life expectancy for the low risk group is 14+ years, for the intermediate group is 8 years and for the high risk group is about 4 years.

In addition, high levels of β2 microglobulin predict a poor outcome. The other useful predictor is the lymphocyte doubling time (Rai *et al.*, 1996). This is calculated from the blood count; a doubling time of over 12 months predicts a good prognosis.

1.2.5. Normal B-cell maturation

Before discussing the abnormalities of the immune system that are associated with B-cell CLL, normal B cell function will be briefly described.

B lymphocytes form part of the immune system in humans, interacting with phagocytes and T lymphocytes to fight infection. They are derived from bone marrow stem cells and are mostly distributed throughout the germinal centres of lymph nodes, the spleen and the respiratory tract. They express different surface markers throughout their maturation (figure 1.4).

One of the main functions of B cells is to produce immunoglobulins (antibodies) in response to an antigen. Immunoglobulins are first expressed on the surface of the B cells, but can then be secreted by a mature B cell (plasma cell). Immunoglobulins can be divided into 5 classes; IgM, IgD, IgA, IgG and IgE. As seen in figure 1.4, IgM is usually produced first (with or without IgD) but the B cell can switch (isotype switching) to produce an immunoglobulin of the IgG or IgA class. This later response is usually more sustained. All immunoglobulin molecules consist of two identical heavy chains and two identical light chains. The heavy chain varies depending on the constant region used (see below) and defines the immunoglobulin class of the antibody. The light chains will be either κ or λ for

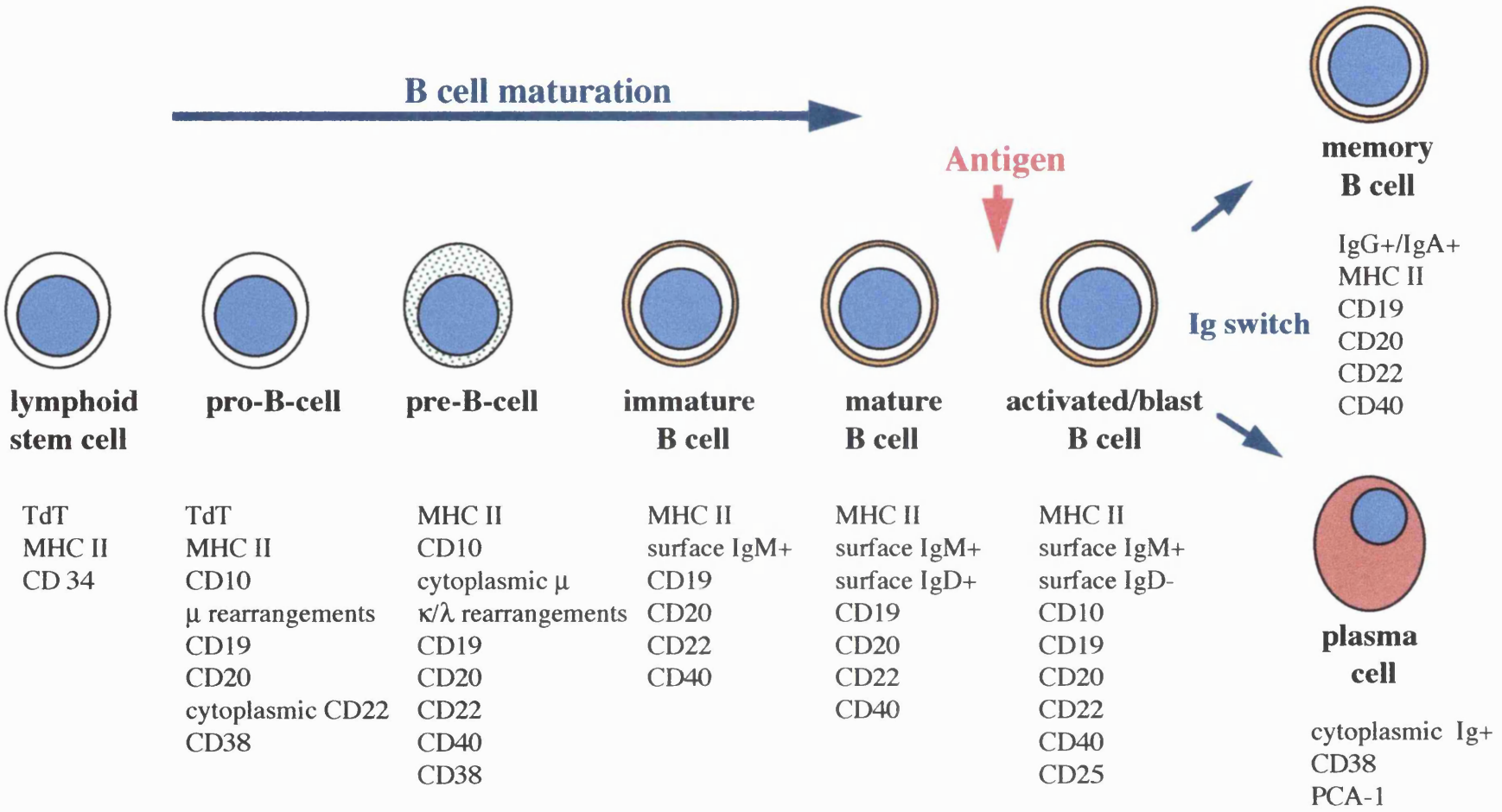


Figure 1.4: B cell maturation. Following stimulation by antigen, B cells may either become a memory B cell or a immunoglobulin secreting plasma cell. Appearance of relevant cell surface markers is shown, as is the sequence of immunoglobulin rearrangement and expression.

all subtypes. The purpose of an immunoglobulin molecule is to bind antigen and so aid its destruction. Many rearrangements of the immunoglobulin genes occur within the B cell as it matures to increase the diversity of the immunoglobulin repertoire produced. These rearrangements will be described briefly here as study of the immunoglobulins in B-cell CLL have given vital clues as to the biology of the disease.

1.2.6. Normal Ig rearrangement and antibody generation

The gene complex that encodes the heavy chains is located on chromosome 14q32, whilst the gene complexes encoding the κ light chain and the λ light chain are on chromosomes 2p12 and 22q11 respectively. During B cell development, elements from within these gene complexes undergo a series of rearrangements to form a specific combination of exons that will ultimately encode the heavy and light chains of the antibody molecule for that particular B cell. The heavy chain genes usually rearrange first. Three families of gene segments make up the heavy chain gene locus; V_H, J_H and D. There are approximately 30 D (diversity segment) genes, 6 functional J_H (joining region) genes and 95 heavy chain variable region genes (V_H), of which only half are functional. The V_H genes are further divided into subgroups, of which there are 7, which are grouped on sequence homology with other members of the family. The number of segments making up each V_H family varies; families 1, 3 and 4 are very large whilst 5, 6 and 7 are very small with only one or two elements each. One segment from each group is ultimately selected at the DNA level to make up the VDJ rearrangement that will encode the variable part of the heavy chain (figure 1.5). The class of immunoglobulin produced depends on the constant region that is used in conjunction with the VDJ rearrangement and, as already discussed, this is switched during B cell maturation. After successful rearrangement of the heavy chain locus, one of approximately 70 Vκ genes rearrange to one of five Jκ genes to encode the light chain (VJ rearrangement). If this fails then a λ light chain rearrangement is assembled using a similar process.

Additional diversity at the junctions between the rearranged V, D and J segments results from imprecise splicing and also through the addition of non-encoded base pairs (N regions) at the D to J and V to DJ junctions through the action of terminal deoxynucleotidyl transferase.

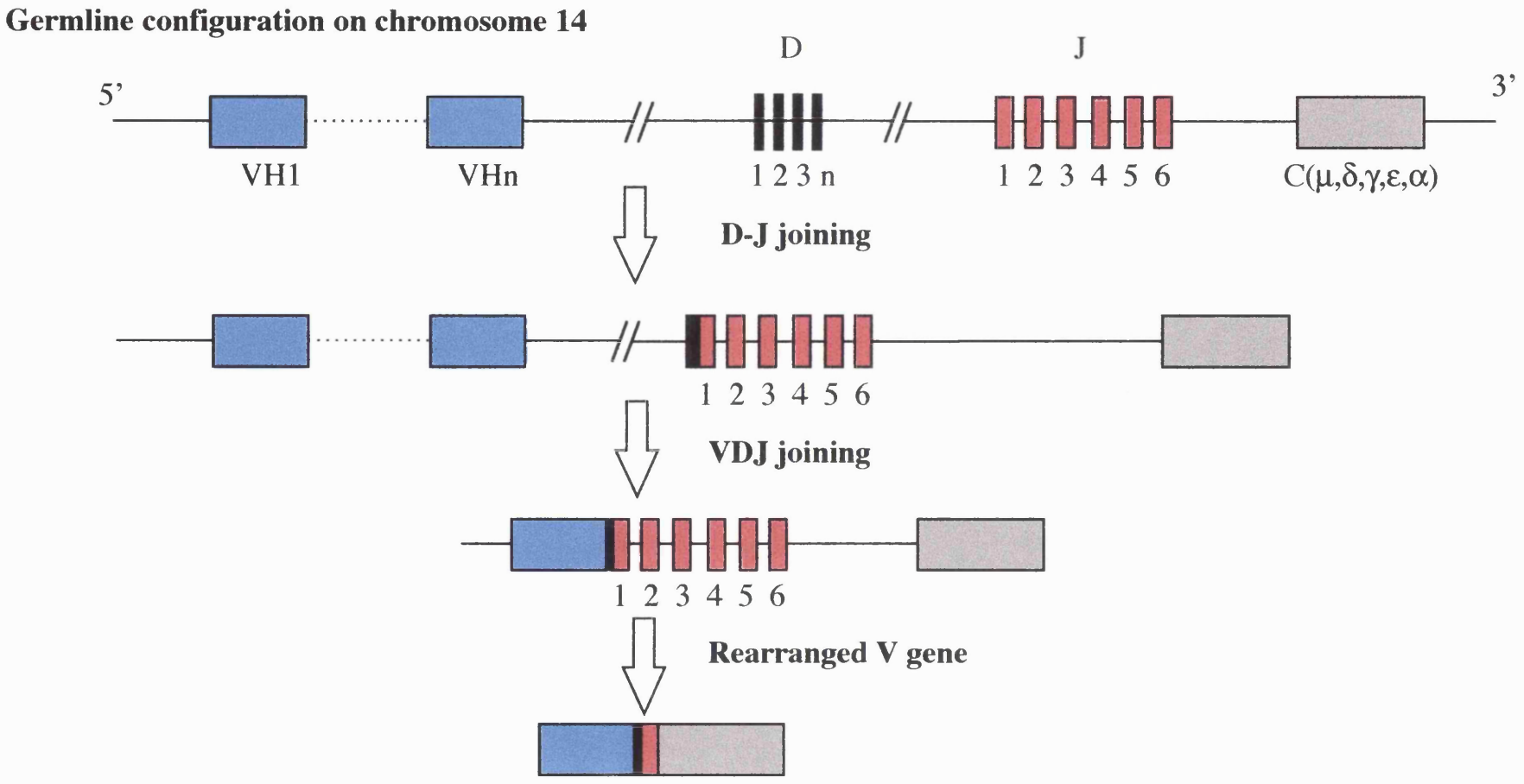


Figure 1.5: IgH rearrangement of the heavy chain locus in B cell development. D and J gene segments are joined to one of 95 functional VH gene segments.The addition of a C segment forms an active transcriptional gene.

Three regions of greater antigen affinity are recognised as hypervariable regions in both the heavy and light chains and are known as complementarity determining regions (CDR). After successful Ig rearrangement, somatic hypermutation occurs in response to antigen stimulation, particularly along the CDR regions, to increase antibody diversity and to allow the cell to produce an antibody with greater affinity to a given antigen. Somatic mutations target the CDR regions most frequently, but mutations are scattered along other parts of the VDJ rearrangement as well and these are known as framework mutations. The third CDR region corresponds to the region of joining of the V and J segment for the light chain, and at the V_H to D-J_H segment junction for the heavy chain. Together the CDRs of both the heavy and light chains fashion the specificity of the Ig molecule and give the B cell repertoire great potential for diversity.

Both Ig isotope switching and somatic hypermutation occur when the B cell migrates through the germinal centre of the lymphoid follicle whilst engaged in a secondary immune response to an antigen.

1.2.7. The B-cell immunoglobulin repertoire

When the immunoglobulin expression of normal B cells has been analysed, certain families of V, D or J elements are used more frequently than others. This is in part due to the over representation of certain families within the genome, but investigators have shown that the choice of family is not an entirely random process (Kipps and Carson, 1993). The V_H 3 and V_H 1 are the genes most frequently used as would be predicted by their increased representation within the genome. However, within these two families there are preferences for certain members over others. For example, the following VH 3 members each account for 8-20% of Ig VH 3 rearrangements; 23, 30/30b, 30.3, 33, 15 and 11, whilst the remaining 17 functional members of the family contribute to <3% of the rearrangements (Johnson *et al.*, 1997).

1.2.8. Immunoglobulin rearrangement and usage in B-cell CLL

CLL cells most commonly express low amounts of surface IgM immunoglobulins (with or without IgD). When the immunoglobulin gene rearrangements are sequenced in CLL cells, a large number of cases show little variation from the germ line i.e. they have not undergone somatic hypermutation.

The typical CLL cell, therefore, looks like a primary B cell that has not migrated through the lymphoid follicle or encountered antigen (figure 1.4). Studies have shown, however, that the repertoire of V_H genes used in B CLL cells is not representative of those seen in normal B cells and suggest some antigen selection occurring on the malignant clone. V_H 3 and V_H 1 families are still the most frequently seen but when actual numbers are analysed, it would appear that V_H 3 is under represented when compared to the normal B cell repertoire (Kipps and Carson, 1993). Within the V_H subgroups a further bias appears to exist, over and above that seen in the normal B cells. For example, although the Ig V_H 1-69 locus is over represented in normal B cells, CLL cells over use this locus when compared with normal (Hamblin *et al.*, 1999) with significantly more usage of the 51p1 gene, a member of this group, than is seen in normal cells (Johnson *et al.*, 1997). The usage of the D segments and J_H segments seen in CLL cells is also significantly skewed when compared to that of normal cells (Johnson *et al.*, 1997).

In addition, distinct differences are seen at the heavy chain CDR3 region between CLL cells and normal B cells that use the same V_H genes. The CDR3 region is significantly longer in the neoplastic cells and resembles closely the germline sequence in its complexity and composition (Johnson *et al.*, 1997).

One explanation for this frequent rearrangement and expression of certain Ig genes is that the cells encoding for antibodies with a particular binding pattern are selected during leukaemogenesis, possibly due to antigen stimulation, resulting in the biased repertoire seen in CLL (Kipps and Carson, 1993).

1.2.9. Somatic hypermutation in CLL

The Ig V genes share extensive homology with the germline sequence and show little evidence of somatic hypermutation in most cases of CLL (Kipps and Carson, 1993).

There are exceptions to this observation. Recent studies have suggested that a minority of patients with CLL have tumours that do show evidence of somatic hypermutation which may be more common if they express one of the VH 5 genes. There is over usage of this VH family in CLL and investigators have shown that when the VH 5 family is used by a CLL cell the sequence differs significantly from the germline in the majority of patients (Humphries *et al.*, 1988), although this has been disputed by others (Rassenti and Kipps, 1993). The sequence differences noted were particularly clustered around the CDR3 region. It

would appear that these CLL tumours express antibodies that have been selected for by their ability to bind some unknown antigen.

Whilst most CLL clones express unmutated surface IgM, 8% of CLLs express surface IgG or IgA, showing that the cell has undergone isotype switching. Several studies have examined the variable regions of the Ig genes expressed by CLL cells with IgA or IgG surface immunoglobulin and found that most had undergone somatic hypermutation (reviewed (Kipps, 1998)). The combination of isotype switching and the acquisition of somatic hypermutation would suggest that these clones had encountered an antigen. One explanation would be that these cells might result from malignant transformation of mature B cells at a different stage of differentiation to the majority of IgM expressing CLL tumours. Another group have shown that, even within the group of tumours expressing IgG there is variation in the amount of somatic hypermutation seen (Ebling *et al.*, 1993). They showed that the V regions expressed in these cells that had undergone isotype switching could be in germline configuration or could show evidence of somatic mutation.

Using RNA fingerprinting and sequence analysis, another group reported that in cases where the Ig expressed by CLL cells was IgG, an expanded group of B cells producing the IgM version of the same antibody i.e. pre-isotype switch, could be detected in the same patients (Dono *et al.*, 1996). They suggest that these cells are the precursors of the malignant clone. This would indicate a multistep process occurring in the pathogenesis of CLL with clonal expansion of IgM expressing B cells being the consequence of the first step/s. The subsequent steps causing malignant transformation may occur at a pre germinal centre stage of development or may not occur until after isotype switching, with or without hypermutation in response to antigen.

In a recent study, the Ig rearrangements in a large number of patients including 20 expressing IgG were sequenced (Valetto *et al.*, 1998). As previously documented, Ig V gene bias was seen in all CLLs. However, this was even more marked in the IgG expressing cells. Of the 20 sequenced, 5 showed almost identical receptors with 90% sequence similarity in the heavy chain and >97% similarity in the light chain. The composition and length of the heavy chain CDR3 region was also noted to be extremely similar. The investigators suggested that this group represent a malignancy that has been selected for from clones of B cells by reactivity with restricted sets of antigens.

Oscier *et al.* have taken this further and have made some interesting observations linking the Ig mutation frequency of CLL cells with other genetic markers (Oscier *et al.*, 1997). and with severity of disease (Hamblin *et al.*, 1998). They have sequenced the Ig heavy chain in 84 patients, 52 of which had stable disease and 32 had progressive disease. These patients were analysed for the cytogenetic abnormalities trisomy 12 and 13q14 deletions using interphase FISH (discussed below). As previously reported, they found biased V_H usage among their patients. They also showed 39% to have a germline configuration of the V_H gene with 61% showing somatic hypermutation*. Of the cases with the germline configuration, they had clinically more aggressive disease and a much higher incidence of trisomy 12 and V_H1-69 usage. The patients with somatic mutation tended to have stable disease and a statistically significant increase in 13q deletions or a normal karyotype. These patients showed an increased use of the V_H3-23 and V_H4-34 genes. They concluded that the transformation event in CLL may occur at two distinct stages of differentiation and that the clone then goes on to acquire subsequent chromosomal changes. A further study from the same group has shown a direct link between the degree of somatic mutation in the Ig V_H genes and patient survival (Hamblin *et al.*, 1999). Median survival for stage A patients with unmutated V_H genes was 95 months compared to 293 months for those with detectable mutations in the tumour V_H genes. Hamblin *et al.* suggest that CLL consists of 2 diseases; one arising from a memory B cell that has a relatively benign course, and the other arising from a more naive B cell that has a more aggressive course. Damle *et al.* correlated the lack of somatic mutation with CD38+ expression by CLL cells. They showed that patients with unmutated Ig V genes and <30% CD38+ expression had prolonged survival whereas patients with mutated Ig V genes and >30% CD38+ expression had a poor outcome and were often unresponsive to therapy (Damle *et al.*, 1999).

It is clear that CLL constitutes a heterogeneous group of tumours as shown by the large differences in immunoglobulin gene rearrangements and B cell maturation between patients. This suggests a complex accumulation of events within the B cell leading ultimately to failure of apoptosis and consequent malignancy.

* Oscier *et al.* do not define 'hypermutation' within their study but counted differences in V_H sequence betwen the germline and the patient's clonal rearrangement. Cases with deletion of 13q showed more mutational events than those with trisomy 12 ($P<0.0003$).

1.2.10. Autoantibodies in CLL

Autoimmunity was documented as being a feature of B-cell CLL in 1967, with 20% of patients noted to be affected at some stage of their disease (Dameshek, 1967). The most commonly seen autoimmune manifestations are haemolytic anaemia and immune thrombocytopenic purpura (ITP).

Two types of autoantibody are seen in patients with B-cell CLL. The first is a polyreactive IgM class with rheumatoid factor activity which is expressed by the malignant cells. Polyreactive IgM autoantibodies may be encoded by Ig V genes that are present in the germline DNA and the expression of these autoantibodies may be a direct consequence of the lack of somatic mutation in the majority of CLL Ig rearrangements (Kipps and Carson, 1993). The pathogenic autoantibodies seen, however, are polyclonal and often of the IgG type i.e. different to the Ig class of the malignant cell in most cases. Evidence suggests that these antibodies are the product of bystander B cells and not the malignant clone (Kipps and Carson, 1993). These cells are not clonally related to the malignant cells in any way. The antibodies are usually monoreactive and of high affinity.

In vitro studies have shown that B-cell CLL cells can be transformed into very effective antigen presenting cells (APCs) by T cell stimulation. A recent review suggested that in the splenic environment, CLL cells may be stimulated to present antigen (Caligaris-Cappio, 1996). In the spleen they will come into contact with soluble forms of red cell and platelet membranes and they may present these self antigen membrane products to residual normal B cells. In this way, the normal cells will be driven to produce high affinity polyclonal autoantibodies with restricted specificity and pathogenic properties, causing conditions such as haemolytic anaemia so often seen in CLL.

1.2.11. The immunodeficiency seen in CLL

The generalised immunodeficiency seen in these patients has also long been recognised (Dameshek, 1967). Immunity to bacteria, viruses and vaccination is often greatly reduced in B-cell CLL patients. Herpes zoster is also common in these patients and can be fatal. The immunodeficiencies seen are both humoral and cellular, although hypogammaglobulinaemia is often the most marked feature. This is due to the decreasing numbers of normal immunoglobulin producing B cells as the expanding malignant B cell clone 'takes over'. The development of

hypogammaglobulinaemia is loosely connected with clinical stage and virtually all patients with advanced disease have decreased levels of immunoglobulins of all isotypes (Rai *et al.*, 1996).

The malignant B cells themselves have little stimulatory activity in mixed lymphocyte culture i.e. they do not stimulate normal allogenic T cells. The lack of T cell stimulation has been suggested to be due to an active evasion of T cell recognition by the malignant clone (Cantwell *et al.*, 1997). The leukaemic cells produce high levels of the cytokine TGF-β, a potent immunosuppressive factor and an inhibitor of both T and B cells (Rai *et al.*, 1996). Much work has been done showing that the malignant B cell is able to inactivate the cytotoxic T cell response against it in other ways as well. Important accessory molecules required for B cell / T cell interactions i.e. B7-1 and B7-2, are rarely expressed by CLL cells (Ranheim and Kipps, 1993). One critical interaction between B cells and T cells in initiating an immune response is the interaction of CD40 (B cell) with its ligand, CD40L (T cell). A rare X-linked condition resulting in CD40L deficiency in humans has been noted to produce a similar picture of immunodeficiency as that seen in CLL (Cantwell *et al.*, 1997). These affected individuals have recurrent bacterial infections due to an immunodeficiency characterised by an inability to generate antigen-specific antibody responses.

Cantwell *et al.* have demonstrated that CD4+ T cells in patients with B-cell CLL do not express CD40 ligand following appropriate stimulation, as would be expected (Cantwell *et al.*, 1997). These T cells, therefore, cannot then stimulate CD40 bearing B cells e.g. CLL cells, to express other co-stimulatory molecules that would aid their detection by cytotoxic T cells (Ranheim and Kipps, 1993). However, the investigators also showed that the T cells do produce appropriate amounts of mRNA for CD40 ligand expression in response to appropriate stimulation (Cantwell *et al.*, 1997). It appears to be the addition of the malignant B cells to the T cells that rapidly down regulates CD40L expression. This effect was also seen when malignant B cells were added to stimulated normal donor T cells. Addition of TGF-β neutralising antibodies failed to reverse this down regulation of CD40L showing that the high levels of TGF-β are not the sole mediators of this effect. Addition of CD40 monoclonal antibody to the culture did block this down regulation of CD40L in vitro. Cantwell *et al.* suggest that CLL cells are evading recognition by T cells by down regulating CD40L expression on the T cells, possibly by inhibition from the excessive expression of CD40 itself (Cantwell *et al.*, 1997). This mechanism may contribute to the self tolerance seen

in B-cell CLL (Caligaris-Cappio, 1996). If this evasion of anti tumour activity could be reversed using monoclonal anti CD40 then this may open possible therapeutic options for the future.

1.2.12. The CD5+ B cell

The cell of origin in B-cell CLL remains unknown. The CLL cells express B cell markers in conjunction with a T cell marker, CD5. The existence of this T cell antigen and its expression on CLL cells was described nearly 20 years ago (Boumsel *et al.*, 1980; Kamoun *et al.*, 1981). The CLL cells do not express any other marker consistent with a T cell phenotype and are said to be CD5 expressing B cells.

Normal CD5+ B cells have been shown to exist in peripheral blood and these cells do share some of the characteristics of malignant CLL cells. Normal CD5+ cells constitute a small subset of B cells that are seen in the lymphoid organs and blood of normal adults and most foetal lymphoid tissues. In the lymph nodes, these cells are mostly found in the mantle zones surrounding the germinal centres of the secondary B cell follicles (Rai *et al.*, 1996). They generally express surface IgM and IgD. Similarly to most CLL cells, the Ig rearrangements of these B cells show little variation from the germline. Normal CD5+ cells also show the same skewed repertoire of Ig V gene usage as is seen in CLL (Caligaris-Cappio, 1996).

There are many differences, however, between the normal CD5+ cells and their malignant counterparts (Caligaris-Cappio, 1996). CLL cells characteristically express low levels of surface immunoglobulin (sIg) and are resistant to infection by Epstein Barr virus. Normal CD5+ cells do not show these features. As already mentioned, B CLL cells are usually in the resting G_0/G_1 phase of the cell cycle whereas the normal CD5+ cells are found in all stages of the cell cycle and are not resistant to apoptosis. CLL cells express high levels of the anti apoptotic protein bcl-2, but normal CD5+ cells only express trace amounts.

Caligaris-Cappio suggests that the normal B cell population that reproduces the entire spectrum of B-CLL cells has yet to be found (Caligaris-Cappio, 1996). However, it may be that many of the features expressed by the malignant cell are a consequence of transforming events that have occurred within the cell and are, therefore, not seen in any normal B cell population.

1.3. Cytogenetic abnormalities in B-cell CLL

The molecular pathogenesis of B-cell CLL remains largely unknown. No specific chromosomal translocation has been associated with CLL and no other specific genetic abnormality has been consistently detected in this malignancy. In B-cell CLL, clonal chromosome abnormalities are found in approximately 50% of cases if studied after appropriate mitogenic stimulation (Juliusson and Merup, 1998). More than half of the abnormal karyotypes are made up of single abnormalities. Complex karyotypes i.e. multiple chromosomal abnormalities, are found in 10-15% of all patients with typical CLL. Trisomy 12 and deletions of chromosome 13q14 are the commonest single abnormalities seen. 11q deletions and 6q deletions also occur as single abnormalities, whilst 75% of those with 14q aberrations have complex karyotypes.

1.3.1. Trisomy 12

As early as 1979, trisomy 12 was defined as the most common cytogenetic abnormality seen in B-cell CLL (Juliusson and Merup, 1998). Early studies used cytogenetic banding techniques to diagnose trisomy 12 and estimated the incidence at approximately 15% (Juliusson *et al.*, 1990). With the advent of fluorescent *in situ* hybridisation techniques (FISH), probes for the centromere of chromosome 12 were designed for use on fresh interphase cells and the incidence has now been shown to be slightly higher at 18-20% (Matutes *et al.*, 1996; Garcia-Marco *et al.*, 1997). A recent review of all publications since 1991 averaged the incidence of trisomy 12 found by FISH analysis to be 22% (Juliusson and Merup, 1998) (see table 1.1).

Using restriction fragment length polymorphisms (discussed below), the extra chromosome 12 seen in trisomy 12 has been shown to derive from the duplication of one autosome with retention of the other and not as a result of a triplication of one autosome with loss of the other (Dierlamn *et al.*, 1997).

The association of trisomy 12 and atypical lymphocyte morphology has been clearly documented. Matutes *et al.* analysed 544 patients for the presence of trisomy 12 using FISH, as part of the MRC CLL 3 trial (Matutes *et al.*, 1996). They showed that overall, 83% of cases had typical CLL morphology and 10% had more than 10% prolymphocytes (immature lymphocytes). Out of those patients with trisomy 12 (20%), 31% had more than 10% prolymphocytes and

Chromosomal abnormality	Conventional Cytogenetics	FISH	Molecular Techniques
trisomy 12	15% - Juliusson 1990	18% - Matutes 1996	
		20% - Garcia-Marco 1997	
		22% - Juliusson review 1998	
deletion 13q	20% - Juliusson 1991	51% - Stilgenbauer 1998	50% - Panayiotidis 1997
		31% - Avet-Loiseau 1996	41% - Corcoran 1997
deletion 11q	9% - Juliusson 1990	20% - Dohner 1997	
	16% - Juliusson 1991	16.6% - Schaffner 1998	
	7% - Juliusson review 1998		
deletion 6q	12% - Juliusson 1991		7% - Merup 1994
abnormalities of 14q	3% - Juliusson 1990		

Table 1.1 : Showing the cytogenetic abnormalities and their frequencies reported in B-cell CLL and a comparison of different techniques

24% had atypical morphology. She concluded that trisomy 12 defines a subgroup of CLL with more frequent atypical morphology.

In a large, multicentre study of 433 patients with CLL, it was demonstrated that patients with sole trisomy 12 abnormalities had the poorest survival of all the single abnormality subgroups analysed (Juliusson *et al.*, 1990). This was significantly worse than the survival of patients with single 13q abnormalities (P=0.01). This association of trisomy 12 with more progressive disease and a shorter therapy-free survival was then confirmed with the second international compilation of data by the International Working Party on Chromosomes in Chronic Lymphocytic Leukemia (IWCCLL) (Juliusson *et al.*, 1991).

An interesting feature of patients with trisomy 12 is that the abnormality is not found in all CLL cells. Garcia-Marco *et al.* demonstrated 21-37% of cells showing trisomy of chromosome 12 in an analysis of 6 patients (Garcia-Marco *et al.*, 1994). A later study by the same group showed the numbers of cells with the abnormality ranged between patients, from 7%-84% (Garcia-Marco *et al.*, 1997). The median percentage of affected cells from the IWCCLL data was 45% (Juliusson *et al.*, 1991). This finding of trisomy 12 in a sub population of cells has been suggested to represent clonal evolution of the disease. A recent review stated that no patient had been documented to acquire trisomy 12 over the course of a longitudinal study (Juliusson and Merup, 1998). However, during the course of progressive disease the percentage of cells with trisomy 12 may increase. A FISH study by Garcia-Marco *et al.* contests this, showing one patient to acquire trisomy 12 during the course of their disease with the other 30 patients analysed showing an increasing percentage of cells with trisomy 12 during the course of their illness (Garcia-Marco *et al.*, 1997).

Another study showed trisomy 12 to be detectable in a proportion of CD34+ stem cells in B-cell CLL patients with known trisomy 12 (Gahn *et al.*, 1997). This group suggest that CLL may not be a malignancy of mature B cells but may arise due to a stem cell disorder, at least in the subset of patients with trisomy 12. The biological significance of trisomy 12 within CLL remains unknown.

1.3.2. 13q deletions

Deletions of the long arm of chromosome 13 involving band 14 are the commonest structural abnormalities seen in B-cell CLL. Some recent studies suggest that the incidence of 13q deletion, when investigated by FISH and Southern blot analysis, may be greater than the incidence of trisomy 12 (table 1.1). Using conventional cytogenetic techniques, the incidence of 13q deletions reported by the IWCCLL was 20% (Juliusson *et al.*, 1991). Using FISH analysis other groups have reported incidences of 51% (Stilgenbauer *et al.*, 1998) and 31% (Avet-Loiseau *et al.*, 1996). A similar frequency is seen with Southern blot analysis. Panayiotidis *et al.* found 13q14 deletions in 50% of patients (Panayiotidis *et al.*, 1997) whilst Corcoran *et al.* reported an incidence of 41% (Corcoran *et al.*, 1998). Approximately 6% of the patients reported by both Corcoran and Panayiotidis show homozygous deletion at this locus. It is thought that a tumour suppressor gene, whose loss or inactivation is crucial in the development of CLL, is located at 13q14.3

Large scale studies have shown that a single abnormality involving 13q does not proffer a worse prognosis than standard risk B-cell CLL i.e. patients with a normal karyotype (Juliusson *et al.*, 1990). This contrasts with all other cytogenetic abnormalities described in B-cell CLL to date. If another cytogenetic abnormality co-exists with the 13q changes then the prognosis is worse (Juliusson *et al.*, 1990).

Considering that trisomy 12 and 13q deletions are the most frequent cytogenetic changes seen in CLL, it is interesting to note that they are very rarely seen together within the same tumour. Early studies suggested that the two abnormalities might be mutually exclusive but they have since been reported co-existing within the same tumour in a small number of cases. Using both conventional cytogenetics and FISH analysis, Mould *et al.* described 8 patients out of a series of over 400 cases with both trisomy 12 and abnormalities of 13q (Mould *et al.*, 1996). In a similar study of 600 cases, Navarro *et al.* reported 9% to show abnormalities of chromosome 12 and 13 (Navarro *et al.*, 1998). The IWCCLL study reported 233 patients out of 512 with more than one chromosomal abnormality, of which 10 patients had both structural changes of 13q and trisomy 12. This low incidence was shown to be significantly less than expected (Juliusson *et al.*, 1991) and Juliusson therefore suggests that the two abnormalities may provide a similar, but unrelated pathogenic influence towards

differentiating B cells (Juliusson and Merup, 1998). If this assumption is correct then acquisition of both abnormalities gives no advantage to a cell progressing towards malignancy.

When DNA from patients was analysed for 13q deletions at the D13S25 locus (see below) using Southern blot and densitometry techniques, comparison with normal controls showed the malignant cells to have 100%, 50% or 0% of the normal values recorded (Jabbar *et al.*, 1995). Investigators drew the conclusion that loss of 13q alleles was a clonal event with the values seen representing no loss (100%), heterozygous loss (50%) or homozygous deletion (0%). Subsequent studies using FISH have cast doubt on this assumption. FISH techniques have an advantage over Southern blotting in this type of analysis in that individual cells within the malignant clone can be analysed. As discussed, Navarro *et al.* found 13q deletions and trisomy 12 within the same patients. In some cases they found that a small number of cells contained three copies of chromosome 12 and another, usually larger, group of cells showed deletion of 13q, suggesting that both abnormalities existed within a proportion of cells (Navarro *et al.*, 1998). This was verified by another group who used a YAC (933e9) covering the region of interest at 13q to test their patients for deletion of this region using FISH. They found 31% of patients had deletion of the region, with the percentage of cells exhibiting chromosomal loss ranging from 31 to 90% (Avet-Loiseau *et al.*, 1996). Merup *et al.* reported patients where the major clone showed heterozygous deletion of 13q14, but where two smaller clones showing either homozygous deletion of 13q14 or no deletion of 13q14 could also be detected (Merup *et al.*, 1998). It should be noted that this group used mitogenic stimulation to culture the B cells and, therefore, normal cells may have been over represented in their samples. It seems likely, therefore, that whilst 13q deletion may exist in the majority of cells in most tumours with 13q14 deletion (either as a heterozygous or as a homozygous loss), it may not be present in the original cell from which the clonal B lymphocyte population is derived.

Mapping of the deleted region at 13q14.3 will be described in detail later. It is worth noting that a controversy arose in 1996 suggesting that the important region of deletion at chromosome 13 was not 13q14 but 13q12, the site of the tumour suppressor gene BRCA 2 (Garcia-Marco *et al.*, 1996). Investigators reported the incidence of deletion at this region to be 80%. This would make this the most frequently occurring cytogenetic abnormality in B-cell CLL. This data was derived from interphase FISH analysis. However, FISH analysis from another

group on over 300 patients (Stilgenbauer *et al.*, 1998) and Southern blot data from our group (Panayiotidis *et al.*, 1997) could not confirm this finding in the patients tested. Deletion of 13q14.3 remains the commonest structural abnormality in most groups of patients reported.

Interestingly, the same region at 13q14.3 has been shown to be deleted in B-cell non-Hodgkin's lymphoma (Liu *et al.*, 1995), acute lymphocytic leukaemia (Liu *et al.*, 1995), splenic lymphoma with villous lymphocytes (Garcia-Marco *et al.*, 1998) and, more recently, myeloid malignancies (La Starza *et al.*, 1998) and multiple myeloma (Chang *et al.*, 1999). This region has consequently become the focus of much attention.

1.3.3. 11q deletions

Translocations involving the long arm of chromosome 11 at band 13 and the heavy chain locus on chromosome 14q32 have been described in atypical B-cell CLL. The data from the first IWCCLL showed 9% of patients (37 out of 391 analysed) to have abnormalities of the long arm of chromosome 11 (Juliusson *et al.*, 1990) (see table 1.1). In 12 cases these abnormalities involved band 13, of which 10 showed a translocation with chromosome 14. The other abnormalities were not documented. The second IWCCLL report documented 16% of patients with 11q abnormalities as seen by conventional cytogenetics (Juliusson *et al.*, 1991). A recent review of publications on cytogenetic abnormalities in CLL averaged all reported 11q deletions at 7% of patients tested with conventional cytogenetics (Juliusson and Merup, 1998).

Deletion of 11q is more frequently seen than rearrangement and, as with 13q, it is likely that a tumour suppressor gene involved in the pathogenesis of B-cell CLL is located within the minimally deleted area.

Two publications from the same group have now identified the region of minimal deletion at 11q and have shown the incidence of deletion to be higher than originally thought. Stilgenbauer *et al.* narrowed the region of deletion to a 2-3 Mb area at 11q22.3-11q23.2 by FISH analysis of 42 lymphoid tumours using a variety of YAC probes from 11q (Stilgenbauer *et al.*, 1996). This deleted area contains several genes already isolated and mapped to 11q, including the tumour suppressor gene ATM (ataxia telangiectasia gene). The same group then analysed 214 B-cell CLL patients using interphase FISH with a single YAC probe from the region of deletion (Dohner *et al.*, 1997). 20% (43 cases) of patients were shown to

have monoallelic deletion of the YAC probe, with the number of cells carrying the deletion ranging from 9.3-98%. The majority, 29 cases, showed >80% of cells to contain the deletion. Only 12 patients had translocations of 11q13, all of which were partnered to chromosome 14. They also tested the same samples for deletion of 13q and trisomy 12. They showed 13q deletions to be the most frequently detected cytogenetic abnormality at 45%, but trisomy 12 was less frequently documented than deletion of 11q at 15%. In the same paper, the investigators reported a difference in the clinical phenotype of these patients with 11q deletion. The patients were younger than those without 11q deletion, their haemoglobins were lower and they had significantly more lymphadenopathy with increased involvement of mediastinal and abdominal nodes. Patients more frequently reported constitutional symptoms and showed more rapid disease progression. Patients with 11q deletions had a 2 fold increase in the likelihood of death compared to patients without 11q deletions. They concluded that deletion of 11q22-q23 defined a sub group of younger patients with a more aggressive form of B-cell CLL.

As already stated, the ATM gene is within this minimal region of deletion at 11q. ATM is the tumour suppressor gene involved in the pathogenesis of ataxia telangiectasia, an autosomal recessive disorder causing, amongst other pathologies, immune deficiency and a predisposition to cancers, particularly lymphoproliferative diseases. The ATM protein is important in the stabilisation of p53 in response to genotoxic stress (figure 1.2A) and, therefore, ATM deficiency leads to loss of an important pro-apoptotic pathway. The carriage rate for mutation of ATM in the Caucasian population is 0.2-1%. A recent study looked for loss of heterozygosity (LOH) at the ATM locus in patients with B-cell CLL (Starostik *et al.*, 1998). Of 36 informative cases, 5 showed LOH at this locus. Four of these patients were then analysed by western blot for the presence of ATM protein and were shown not to have any detectable protein. A further 111 patients were then analysed for the expression of the ATM protein and 34% were shown to have levels below 50%. The investigators also showed that patients with ATM deficiency had more aggressive disease with decreased survival. A second group then looked at these patients at a molecular level (Bullrich *et al.*, 1999). As well as the 5 cases shown to have LOH at ATM, they randomly chose a fourth patient with aggressive disease for analysis. They found 4 of the cases to have missense mutations of the ATM gene with two of the patients having the mutation in the germ line. Interestingly, none of the patients

had concomitant 13q deletion. They concluded that ATM was likely to be the gene involved in the pathogenesis of CLL in a subset of patients. They also suggest that ATM carriers are at an increased risk of developing CLL.

Another group verified these findings (Stankovic *et al.*, 1999). They looked at 32 sporadic cases using restriction digests to identify abnormalities of the ATM gene. They also tested 20 of the patients for expression of the ATM protein, using western blot analysis. They found 6 cases with mutations of the ATM gene where protein expression was reduced or absent. Two of the cases carried the mutation in the germline and, surprisingly, none of these patients were subsequently demonstrated to have a 11q deletion at this locus by FISH analysis. They also concluded that the ATM gene plays a role in the pathogenesis of B-cell CLL in an unknown percentage of cases.

However, it is unlikely that the ATM gene is the only gene at 11q involved in B-cell CLL. A recent presentation by Schaffner *et al.* showed that 56 out of 338 patients with B-cell CLL had deletion at 11q as shown by FISH analysis (Schaffner *et al.*, 1998). Of these, 8 were shown to have mutation of the ATM gene on the retained allele. Another 48 did not have any detectable mutations in the ATM gene, despite having 11q deletions. They also found 2 other patients with mutations of both alleles without 11q deletion detectable by FISH. None of the patients analysed had mutations of the ATM gene in the germline. It would appear that 11q is the site of more than one tumour suppressor gene involved in the pathogenesis of B-cell CLL.

1.3.4. 6q deletions

Structural abnormalities of chromosome 6, mainly deletions, were found in 12% of CLL cases by the IWCCLL (Juliusson *et al.*, 1991) (table 1.1). The breakpoints seen were distributed over the whole chromosome with a preference for the regions 6q15-6q23, as shown by conventional cytogenetics. Very little other data has been published regarding the region of deletion at 6q in B-cell CLL. However, 6q deletions are also seen in non-Hodgkin's lymphoma (NHL) and acute lymphocytic leukaemia (ALL). Analysis of chromosome 6 deletion in NHL has revealed two regions of minimal deletion at 6q21-23 and 6q25-q27 (Gaidano *et al.*, 1992). Within the region at 6q21-q23, two preferential regions of deletion were seen, one at 6q21 and another at 6q23 (Offit *et al.*, 1993). Deletion at 6q21-q23 has been shown to particularly frequent in a subset of B cell small lymphocytic

lymphoma (Offit *et al.*, 1994). The region of deletion at 6q25-q27 has now been mapped and is being analysed for candidate tumour suppressor genes (Hauptschein *et al.*, 1998). However, whether the gene involved in the pathogenesis of B-cell CLL is the same as that involved in NHL remains to be clarified.

An interesting correlation between deleted 6q and T-cell receptor β gene (TCRβ) rearrangement in B-cell CLL was noted by one group (Merup *et al.*, 1994). Rearrangement of TCRβ is occasionally reported in B-cell CLL, despite normally being rearranged in T cells. Merup *et al.* found TCRβ rearrangement in 6 out of 100 patients and showed that 3 of these patients had 6q deletions. Only another 4 of the remaining 94 patients were found to have 6q deletions. The significance of this finding is at present unknown.

Deletions of chromosome 6, as for other single abnormalities other than deleted 13q, give a worse than standard risk prognosis. However, the prognosis remains better than that seen for trisomy 12 or complex karyotypes (Juliusson *et al.*, 1990).

1.3.5. Abnormalities of chromosome 14

Structural aberrations involving the long arm of chromosome 14 were found in 41 patients out of an initial 433, by the IWCCLL (Juliusson *et al.*, 1990) (see table 1.1). Most abnormalities were a result of a translocation of various donor chromosomes to 14q32 (the immunoglobulin heavy chain locus). Multiple translocation partners were implicated but the most frequent, as already discussed, was chromosome 11q13.

As already described, translocations usually mark the site of an activated oncogene. One such oncogene has been implicated in B-cell CLL from the translocation t(14;19)(q32;q13) which results in the rearrangement and upregulation of the *bcl-3* oncogene. This translocation has only been described in 16 cases of CLL in the literature so far (Dierlamn *et al.*, 1997), is often seen in association with trisomy 12 and tends to be associated with a young age and an aggressive course.

Because 14q abnormalities are rarely seen as a single abnormality it is difficult to assess their effect on prognosis. However, from the small numbers seen, patients with single 14q abnormalities have the same risk as those with other single, non 13q abnormalities (Juliusson *et al.*, 1990; Juliusson *et al.*, 1991).

1.3.6. Oncogenes and tumour suppressor genes in B-cell CLL

Various oncogenes known to be involved in other B cell malignancies have been looked at in B-cell CLL. *Bcl-2*, *c-myc* and *ras* genes have been shown not to be primarily altered in B-cell CLL (Gaidano *et al.*, 1994; O'Brien *et al.*, 1995), despite bcl-2 protein expression being high in over 80% of cases. One hundred patients with CD5+ B-cell CLL were analysed for rearrangements of *bcl-1* (a proto-oncogene implicated in other B cell malignancies) by Southern blot analysis and no abnormalities were found (Gaidano *et al.*, 1994). However, a recent report of patients with atypical CD5+ B-cell CLL showed 15/57 to have the translocation t(11;14)(q23;q32) involving the *bcl-1* locus, using conventional cytogenetics and FISH analysis (Cuneo *et al.*, 1997). This translocation is known to be associated with mantle cell lymphoma and the investigators suggest that these patients with atypical CLL may represent a recently described group with 'mantle cell leukaemia'.

The tumour suppressor gene p53 has been shown to be involved in many different malignancies and has also been analysed for mutations in CLL patients. Gaidano *et al.* reported abnormalities of the p53 tumour suppressor gene to be detected in 10% of all cases of B-cell CLL, with 60% of these showing bilateral inactivation at this locus (Gaidano *et al.*, 1994). A second group reported 17% of B-cell CLL patients to have monoallelic disruption of the p53 locus (Dohner *et al.*, 1995). Of these, over 50% were shown to have p53 mutations on the remaining allele using single stranded conformational polymorphism analysis (SSCP). The investigators reported a worse survival and a resistance to therapy in patients with p53 mutations when compared to other CLL patients. p53 mutation has also been reported in approximately 40% of patients with Richter's transformation (immunoblastic transformation of CLL occurring in 3-5% of cases).

1.4. Finding disease genes within the human genome

1.4.1. The scale of the problem

The human genome contains approximately 3000 mega base pairs (Mb) of DNA divided into 22 autosomes and 2 sex chromosomes, which range in size from 50 to 260 Mb. Human somatic cells are typically diploid (22 pairs of autosomes

and two sex chromosomes) whereas germ cells are haploid (one copy of each autosome and one sex chromosome). At the end of each chromosome are telomeres, sequences that stabilise the chromosome and prevent shortening with loss of material on subsequent divisions of the cell. Each chromosome also contains a centromere which is the site of attachment for the spindle apparatus during mitosis. Both the telomeric and the centromeric DNA are made up of highly repetitive sequences. The rest of the DNA that makes up chromosomes is a mixture of repetitive DNA interspersed with unique sequences (Korenberg and Rykowski, 1988). The two classes of repetitive sequences are the short interspersed nucleotide repeat (SINE) and the long interspersed nucleotide repeat (LINE). The commonest SINE is the *Alu* repeat, approximately 300bp of sequence repeated every 3 to 10kb. The LINE repeats are mostly L1 sequences which can be up to 6.4 kb in length and make up approximately 15% of the human genome (Kazazian and Moran, 1998). There are an estimated 50,000-100,000 genes in the human genome, the coding regions of these genes and their regulatory elements probably account for less than 10% of total DNA (Bishop, 1974; Vogelstein and Kinzler, 1998). Given that only 1-2% of the genome will be expressed as mRNA in any specific tissue at any time it is not surprisingly that many people refer to 'gene hunting' as trying to find a needle in a haystack (Hochgeschwender and Brennan, 1991).

Despite these obstacles, many genes involved in the major inherited diseases of childhood, such as Duchenne muscular dystrophy and cystic fibrosis, have now been identified and characterised (Koenig *et al.*, 1987; Rommens *et al.*, 1989). The advantage of looking for genes involved in these hereditary diseases is that the abnormal gene is consistent between affected patients and the abnormality is present in all cells giving a recognisable phenotype. The difficulty in finding genes associated with cancer is that the abnormalities within a given cell may be multiple, they may vary between individuals with phenotypically the same disease and the abnormalities will only be present in the proportion of cells that have acquired the mutation. The exception to this is the hereditary cancer syndromes. However, B-cell CLL falls into the former group.

Methods for cloning disease genes can be divided into three categories (Collins, 1995); the functional cloning approach, the positional cloning approach and the positional candidate approach.

1.4.2. The functional cloning approach

A minority of genes have been identified and characterised by a knowledge of their biochemical function. One example is the gene involved in the hereditary disease, Lesch Nyan syndrome. The deficient enzyme in this disease, hypoxanthine phosphoribosyltransferase (HPRT), is mostly expressed in the brain in normal individuals. A mouse neuroblastoma cell line was identified that produced increased levels of abnormal HPRT. The over expressed HPRT cDNA sequences from this cell line were identified by comparative hybridisations to a non overproducing cell line. The human homologue was then identified from a human cDNA library using the mouse cDNA as a probe (Brennand *et al.*, 1982).

Haemophilia has been known for centuries to be an sex linked disorder and the protein, factor VIII, was characterised before the gene. Oligonucleotides were designed from the predicted sequence of the gene based on the protein sequence. The oligos were then used to screen a genomic library from an individual with 4X chromosomes and the gene was identified (Gitschier *et al.*, 1984).

(functional cloning is not equivalent to expression cloning).

1.4.3. The positional cloning approach

Unfortunately, the biochemical defects in most diseases, including B-cell CLL, are unknown. In these cases the location of the gene needs to be defined and techniques applied to identify candidate coding sequences from the appropriate genomic region. This approach to gene localisation and characterisation is known as positional cloning. The gene for chronic granulomatous disease was the first gene to be identified using positional cloning techniques alone (Royer-Pokora *et al.*, 1986). Since then, numerous genes such as Duchenne muscular dystrophy (Koenig *et al.*, 1987) and cystic fibrosis (Riordan *et al.*, 1989) have been cloned using similar positional cloning techniques.

1.4.4. Mapping

The first stage to identifying a disease gene by its position is to identify the affected location within the genome and to construct a map of the area. Genome maps are linear representations of DNA that show organisation based on a particular system of landmarks. There are three types of genomic map: cytogenetic maps, physical maps and genetic maps.

Cytogenetic mapping

A cytogenetic map is a representation of an individual chromosome when examined microscopically. The earliest method used to locate a gene to a particular chromosome made use of somatic cell hybrids. When a hybrid cell is made between a human cell and a rodent cell, there is progressive loss of the human chromosomes as the cells are grown in culture. Ultimately, only one or two human chromosomes will be left. Panels of hybrid cells have been made covering the full human chromosome spectrum. Probes of interest can be mapped to a particular chromosome by hybridisation to Southern blots of DNA from these hybrid cells (Watson *et al.*, 1992). The HLA gene cluster was first assigned to chromosome 6 using a human-Chinese hamster hybrid cell panel (Jongsma *et al.*, 1973).

Simpler techniques designed to examine chromosomes directly have now been developed. When a metaphase spread of chromosomes from peripheral lymphocytes is stained with the dye giemsa, each chromosome exhibits a characteristic pattern of light and dark bands that distinguish it from other chromosomes. The dark bands (G bands) are A-T rich and are late replicating (Korenberg and Rykowski, 1988). These can be seen by giemsa staining after proteolytic digestion. G bands are rich in LINE repeats and are relatively gene deplete. The light bands (R) are G-C rich and are early replicating. These can be visualised with giemsa staining after heat denaturation in saline (Craig and Bickmore, 1994). In contrast to G bands, R bands are gene rich areas of the chromosome and are rich in *Alu* repeats (Korenberg and Rykowski, 1988). A cytogenetic map of chromosome 13 based on banding techniques can be seen in figure 1.6. The region of deletion in B cell malignancy (DBM) has been located to band 13q14.3 (R band) as discussed later.

A karyotype is a representation of all the chromosomes for an individual. By studying chromosomal banding, this karyotype may reveal abnormalities such as translocations, duplications, deletions, inversions etc. The location of the Duchenne muscular dystrophy gene at Xp21 was determined by the discovery of such a translocation using this method (Ray *et al.*, 1985), as was the location of the Kallmann gene at Xp22.3 (Franco *et al.*, 1991; Legouis *et al.*, 1991). To increase the sensitivity of this mapping by banding, additional fluorescent *in situ* hybridisation (FISH) techniques can be used. Probes for each whole chromosome are available commercially and can each be assigned a different fluorescing colour. By 'painting' a karyotype with these chromosome paints, subtle abnormalities not

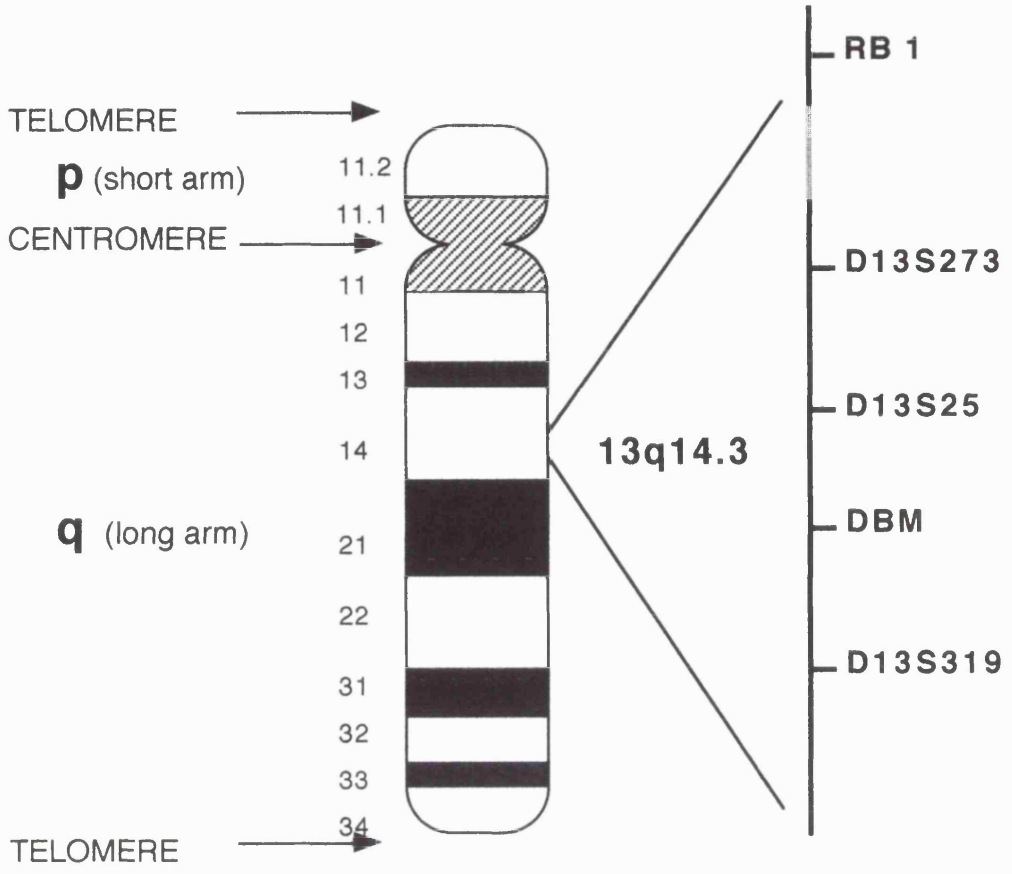


Figure 1.6: Cytogenetic map of Chromosome 13. The region deleted in B cell malignancy (DBM) is schematically represented at band 13q14.3 between two microsatellite markers, D13S25 and D13S319.

RB 1 - retinoblastoma gene

seen with banding, such as balanced translocations, can be identified by a change in colour of a chromosome. FISH techniques can also be used to map individual probes to this cytogenetic map (Trask, 1991). Clones containing fragments of DNA e.g. cosmids and PACs (described in more detail below), can be labelled with a fluorescent dye and hybridised to a chromosomal spread. By viewing under a fluorescent microscope the probes can be visualised and mapped to a specific chromosome (Trask, 1991). Two probes can be distinguished from each other within 1Mb using metaphase FISH (Heiskanen, 1996). If interphase cells are used in addition to metaphase (the chromatin is less condensed in interphase) then probes can be distinguished from each other when they are as close as 50kb (Trask *et al.*, 1989). The order of probes along a chromosome can be assessed by labelling probes in different colours and comparing their localisation to that of a known centromeric or telomeric probe. With the advent of fibre FISH techniques, cytogenetic mapping is becoming ever more sophisticated (Heiskanen, 1996).

Physical mapping

Physical maps of DNA consist of an order of landmarks with distances between them measured in base pairs. The useful physical maps for the 'gene hunter' are of small stretches of DNA encompassing a region known to contain a gene implicated in a particular disease. These original locations may well have been discovered from cytogenetic mapping. Several types of landmark are used in physical mapping:

Sites for restriction enzymes (recognised sequences) can be used as landmarks in physical mapping. When DNA is digested with a given restriction enzyme, a pattern of fragments of varying sizes will be produced that is dependent on the position of these sites throughout the genome. If the enzyme used cuts infrequently, e.g. BssH II, then the fragments produced can be separated by pulse field gel electrophoresis (PFGE) (Schwartz and Cantor, 1984). If the enzyme cuts more frequently, e.g. EcoR I, then the fragments can be separated on an agarose gel by standard gel electrophoresis. The DNA from these gels can be transferred to nylon membranes by Southern blotting (Southern, 1975). Probes from the area of interest can then be hybridised to these filters and the jigsaw pieces can gradually be assembled by repeating the process for a series of restriction enzymes and probes. This type of physical map is referred to as a restriction map. If the area thought to contain the gene of interest is disrupted in

affected patients e.g. by a deletion or translocation, then the pattern of fragments produced by restriction enzyme digest of their DNA will be different i.e. a fragment may be completely absent or may be rearranged and seen as a different size. This further adds to the definition of the map in relation to the location of the gene. The map of the region containing the Wilms tumour gene was initially constructed using long range restriction maps (Call *et al.*, 1990).

Another type of landmark used for physical mapping are sequence tagged sites (STS). These are unique sequences of approximately 300 base pairs (60-1000 bp) which can by amplified with known primers using the polymerase chain reaction (PCR) (Watson *et al.*, 1992). STSs can be mapped relative to each other in terms of both order and spacing within a given region of the genome. These markers can either be used as hybridisation probes or in PCR amplifications. Data regarding STSs, including the primers used to amplify them and their chromosomal location can be found in various public databases e.g. Genome database (http://www.gdb.org/) and the Whitehead Institute for Biomedical Research (http://www-genome.wi.mit.edu/). Similar to the STSs are the expressed sequence tags (ESTs). These are unique sequences isolated from cDNA libraries that can be amplified by known PCR primers, as for the STSs. These markers are of particular interest to the 'positional cloner' because they potentially represent pieces of genes. Hundreds of thousands of ESTs have been generated to date and are available from public databases e.g. dbEST (see http://www.ncbi.nlm.nih.gov/dbEST/index.html). Unfortunately, very few have been positioned on physical maps and so their use is limited in the construction of a localised map. This will gradually change as more of the human genome is sequenced and mapped.

The greatest advantage of these markers (STSs and ESTs) is that they can be used to assemble maps of regions of the genome represented in smaller and smaller clones (contigs). For the identification of coding sequences within a given area, not only does the region need to be well defined but the DNA is much easier to manipulate for further experiments if it is available in short pieces i.e. digested and cloned into vectors such as yeast artificial chromosomes (YACs), P1 derived artificial chromosomes (PACs), bacteriophage derived artificial chromosomes (BACs) and cosmids. YACs are an excellent starting point for the construction of these contigs because they contain inserts of DNA from 100kb - 1.5Mb and, therefore, even a relatively large region can be quickly covered in YAC clones (Schlessinger, 1990). Public human genomic YAC libraries are available (e.g.

Centre D'Etude du Polymorphisme Humaine - CEPH) for screening with STS and EST markers to identify relevant clones. If different clones contain the same markers then they can be assumed to overlap. Techniques are available to clone the end sequences of YACs (Riley *et al.*, 1990). This generates more specific sequence (additional STSs) from the region and further YAC clones can be identified. This method of cloning a genomic region is known as chromosome walking. The size of the YAC inserts can be assessed by PFGE and the distances within the physical map can be accurately defined. The depth of the map contig can be increased by constructing similar contigs in bacteriophage (PAC and BAC) derived clones and cosmids. PACs are derived from the bacteriophage P1 cloning system (Pierce *et al.*, 1992) and contain inserts of DNA between 100-150kb. Methods for obtaining the sequence of the insert ends are available for these clones as well (Wang and Keating, 1994). BACs (bacterial artificial chromosomes) (Shizuya *et al.*, 1992) take slightly larger inserts than PACs e.g. in the region of 300kb, whilst cosmids are circular vectors able to take up to 40kb of insert. PAC libraries are publicly available as for the YACs (Human Genome Mapping Project-HGMP, Cambridge) but cosmid libraries are not widely available and need to be specifically constructed, often from corresponding YAC clones.

When a region is mapped and cloned in this way, candidate gene sequences can be identified. For example, deletion of the long arm of chromosome 5 (5q-) is a recurring abnormality seen in myeloid malignancies and is thought to contain at least one tumour suppressor gene. A physical map of the region was generated in 108 PAC, YAC and BAC clones with 11 ESTs and 97 STSs. Using this highly detailed map the defined minimal region of deletion was greatly reduced and candidate tumour suppressor gene cDNAs were identified (Zhao *et al.*, 1997).

A further technique used in physical mapping is radiation hybrid mapping. This technique is a sophisticated version of somatic cell hybrids. Rearranged and 'broken' chromosomes can now be artificially produced by irradiation of DNA. An individual chromosome from a somatic cell hybrid can be selected and irradiated to artificially produce fragmented DNA with translocations and deletions within the chromosomal structure (Benham *et al.*, 1989). The irradiated cells are then fused again with rodent cells as before so that only a fragment of abnormal human chromosome exists within a cell. A panel of cells representing one chromosome can be produced. Cells containing specific regions of DNA can be identified using STS and EST markers and the DNA can be isolated. The region of DNA containing the Wilms tumour gene was isolated with this technique (Call *et al.*,

1990). Using further mapping and cDNA library screening, the gene was subsequently identified.

Radiation hybrid panels also allow for a powerful form of mapping. By assessing the frequencies with which various STS and EST markers are found together in radiation hybrid cells, statistical packages can be used to assess the distance between them and so assemble a map of a given region. This technique has been extensively used for mapping of many disease regions including that spanning the affected area of chromosome 15q26.1 in Blooms syndrome (Straughen *et al.*, 1996), an autosomal recessive disease characterised by growth retardation, male infertility and immunodeficiency. A similar technique has been used for mapping of chromosome 13 (Hawthorn and Cowell, 1995). The Human Genome Mapping Project are making extensive use of radiation hybrid mapping across the entire genome. This information can be found at web sites such as the CEPH-Genethon integrated map (http://www.cephb.fr:80/)

Genetic mapping

A genetic map reflects the possibility that two markers that are closely spaced on a chromosome will remain together at mitosis i.e. they will not be separated in the process of homologous recombination. The greater the frequency of two markers being separated, the greater the distance between them and vice versa. The unit of measure in a genetic map is the centimorgan (cM). There is some correlation between physical and genetic distances and a cM is taken to be approximately 1Mb. To be informative, a genetic marker must display polymorphic properties i.e. its sequence must be different in the same individual on opposing alleles. If these differences affect a site for a restriction enzyme then digestion of DNA with that enzyme will provide a different length of DNA fragment for each chromosome-restriction fragment length polymorphism (RFLP). These can be detected by Southern blot analysis. An individual is said to be informative for a RFLP if their chromosomes can be differentiated on the bases of different size fragments. RFLPs are inherited in a Mendelian manner and can be traced through families. Linkage maps of the genome have been produced using large sets of families. Much of this work has been done with families from the Mormon Church and the pedigree is very extensive (Watson *et al.*, 1992). DNA from these families makes up the core of the collection of linkage data provided by CEPH.

Mapping of regions of the genome by RFLPs led to isolation of candidate sequences for many genes including the cystic fibrosis gene and the Duchenne muscular dystrophy gene. RFLPs were particularly important in the hunt for the cystic fibrosis gene as no chromosomal abnormality could be detected by other methods (see below).

Microsatellite markers are now superseding RFLPs for genetic mapping. These are short repeat sequences (repeated 5 to 50 times) that are polymorphic and, therefore, often informative for individual chromosomes. They are very numerous (every 30-60kb) and can be amplified by PCR. They can be followed through family pedigrees in the same way as RFLPs.

Ultimately, correlation of information from all three types of map is necessary to build a picture of the human genome and, more importantly in the context of this project, to clearly define the location of a particular disease gene. In all cases of positional cloning, it is essential to keep returning to the patient data to define and redefine the boundaries of the genome that indicate where the relevant gene may lie. As maps become more sophisticated and more markers are generated, these markers can be used to identify more genomic clones and consequently more unique markers. In this way a gradual 'walk' across the genome is performed. Many cDNAs of genes have been identified by chromosome walking e.g. retinoblastoma gene at 13q14.3 (Friend *et al.*, 1986).

When the area that contains a disease gene has been identified and mapped, candidate sequences for the gene need to be identified. There are several different techniques available to do this that will be discussed here.

1.4.5. Identification of candidate gene coding sequences from a mapped region of the human genome

cDNA library screening

cDNA libraries of a given tissue are constructed by reverse transcription of mRNA from that tissue, using an oligo d(T) primer i.e. complimentary to any poly A tail sequence. Most positional cloning techniques will utilise a cDNA library at some stage and this technique remains the mainstay of this type of cloning. The cDNA is cloned into a plasmid or bacteriophage vector and transfected into an *Escherichia coli* strain. The cells are then plated onto appropriate media and filter lifts of the clones are taken. These filters can be

screened with probes to isolate full length clones corresponding to smaller sections of a message. The most efficient probe to use on a cDNA library is another cDNA fragment e.g. an exon. However, genomic fragments from cosmids, phage clones and even YACs have been used successfully (Hochgeschwender and Brennan, 1991). The chronic granulomatosis gene was ultimately isolated using cDNA library screening (Royer-Pokora *et al.*, 1986).

Zoo blots

Many important genes show evolutionary sequence conservation through different species, even though the structure and position of the gene may vary considerably. Protein coding sequences tend to be more conserved than non coding intronic sequences (Hochgeschwender and Brennan, 1991). This sequence conservation forms the basis for finding genes by zoo blot hybridisation. Sections of DNA from a region of interest e.g. cosmid fragments from a physical map, are hybridised to Southern blots containing DNA from other species such as mouse, cow, chicken etc. (zoo blots) and washed at low stringencies. If hybridisation signals are seen in other species then this is suggestive of a conserved coding sequence within the probe used. This probe can then be hybridised to a cDNA library to identify the full coding sequence. Candidate cDNAs for the Duchenne muscular dystrophy gene were isolated using zoo blots (Monaco *et al.*, 1986). This method will obviously not detect any gene that does not show evolutionary conservation.

CpG island identification

Genomic clones can be analysed for the presence of CpG islands. The dinucleotide repeat CG is rare within the vertebrate genome, occurring at approximately one-fifth of the expected frequency (Bird, 1986). Most CpG dinucleotides are methylated, accounting for most of the methylcytosine in the vertebrate genome. However, an unmethylated form of CpG repeat is often associated with the 5' part of a gene and is particularly associated with the promoter region (Bird, 1986). Certain restriction enzymes such as Hpa II recognise unmethylated CpG repeats and will cut these stretches of DNA at multiple intervals producing very small fragments, while they do not cut at methylated sites. For this reason these CpG islands are often termed HTF (Hpa II

tiny fragments) islands. Genomic fragments found to contain a CpG island can be used to screen a cDNA library on the assumption that they may contain the 5' part of a gene.

The human homologue of the Drosophila tailless gene (TLX) was found from the identification of a CpG island (Jackson *et al.*, 1998). The TLX gene lies on chromosome 6 within a region deleted in lymphoid malignancies. A cosmid clone from the map of the region was shown to contain part of a CpG island. Sample sequencing from this cosmid revealed high homologies to the mouse, chicken and Drosophila TLX gene. Further sequencing of the cosmid insert revealed the entire human homologue.

However, many tissue specific genes are not associated with CpG islands and, therefore, cannot be isolated in this way.

Exon trapping

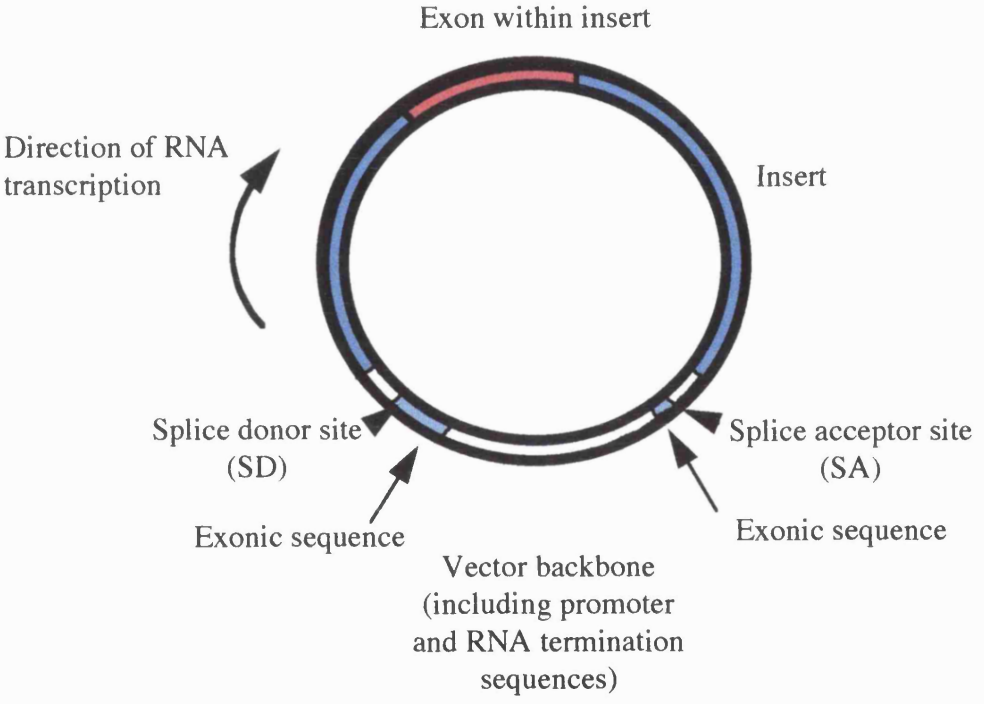
This technique was pioneered to identify rare coding sequences from genomic DNA based on the presence of 5' and 3' splice sites i.e. splice donor and splice acceptor sites, at the borders of exons within the human genome. These sites are recognised by the normal RNA transcription modification machinery within the nucleus of a cell and are utilised in the production of mature mRNA transcripts.

DNA from a region of interest e.g. cloned into YACs, is digested and subcloned into an exon trapping vector. These vectors contain a strong promoter e.g. the rat preproinsulin promoter, and artificial splice donor and acceptor sites. When transfected into mammalian cells, RNA transcription will be driven by the strong promoter from the splice acceptor site to the splice donor site. Any sequences within the insert containing recognisable splice sites will be spliced into the RNA during transcription. More than one exon can be recovered from one genomic insert. By amplifying RNA purified from the transfected cells in an RT PCR reaction using primers from the vector arms, any trapped spliced sequences will be amplified. These can then be cloned and sequenced (see figure 1.7). This system was first described using a plasmid vector, pETV-SD (Duyk *et al.*, 1990). Buckler *et al.*(Buckler *et al.*, 1991) also described a similar system using the plasmid vector pSPL1. A modified phage vector, λ Get, with an automated plasmid subcloning facility has also been reported (Nehls *et al.*, 1994) which

1. DNA clone e.g. YAC, from region of interest

2. Partially digested DNA

3. Individual exon trapping clone transfected into mammalian cells



4. RNA molecule with trapped exon that can be amplified by PCR

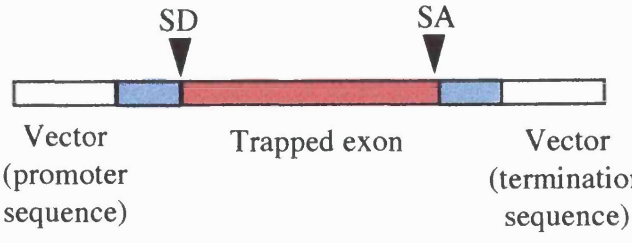


Figure 1.7: Schematic representing the steps involved in exon trapping

allows for cloning of larger fragments of DNA from a region of interest, so covering a larger genomic distance.

Exon trapping techniques have the advantage that expression patterns of a particular RNA are irrelevant to detection of putative exons. However, sequences that are flanked by cryptic splice sites will also be falsely amplified. All amplified products must be hybridised to Northern blots and / or cDNA libraries to confirm their existence. Unfortunately, due to the way in which these sequences are identified, a negative hybridisation does not mean that they do not exist at the RNA level (Hochgeschwender and Brennan, 1991). The transcript that the sequence belongs to may be transcribed in other tissues or at other stages of development, or may be present in too low copy numbers to detect by Northern analysis. A lesser problem with this technique is that any genes that lack introns will not be identified. These genes are thought to be rare in the human genome (Duyk *et al.*, 1990), although the *c-jun* proto-oncogene is an example.

This system will identify genes that are not associated with CpG islands, and genes that are not evolutionarily conserved. Exon trapping has been used to successfully identify exons for the neurofibromatosis 2 tumour suppressor gene (Trofatter *et al.*, 1993) and exons from candidate genes amplified at 12q13.1 in sarcomas and brain tumours (Elkahloun *et al.*, 1997).

cDNA selection

cDNA selection is a technique where cDNAs are selected by hybridisation to genomic fragments (Parimoo *et al.*, 1991; Hamagucchi *et al.*, 1998). YAC DNA from a physical map of a region of interest is immobilised on nylon filter discs and blocked for repetitive sequences. The filters are then hybridised to cDNA fragments from a cDNA library that have been prepared by PCR amplification with known flanking vector sequences. The filters are washed and the hybridised material is amplified using PCR primers complimentary to the vector sequences. The amplified products can be cloned and sequenced. A limitation of this approach is that it requires the chosen cDNA library to represent the coding sequences of the gene of interest and this may be difficult to control for. This system will detect messages without splice sites, those with no evolutionary conservation and those that lack CpG islands. cDNA selection has been successful in identifying a gene implicated in early onset Alzheimer's disease (Sherrington *et al.*, 1995) and in the identification of new genes at 12p (Baens *et al.*, 1995).

RACE PCR

Rapid amplification of cDNA ends (RACE) by PCR is a method for identifying the 3' and 5' regions of cDNA messages which have only partially been characterised (Frohman *et al.*, 1988). It is a PCR based technique where modified sequences are added to the 3' and 5' ends of a cDNA to allow amplification between a known coding sequence and the rest of its cDNA. It is particularly useful for identifying low abundance messages that are difficult to recover from cDNA libraries. As a technique, it has been used widely to isolate full length genes including the cystic fibrosis gene (Riordan *et al.*, 1989). It has also been employed to discover new genes at translocation breakpoints, as it allows extension of a sequence from known sequence into unknown sequence at the RNA level (Willis *et al.*, 1998). This technique is discussed in greater detail in chapter 6.

1.4.6. The cystic fibrosis gene

To illustrate how positional cloning can be successful and how complimentary techniques need to be employed, sometimes on an enormous scale, the example of the cystic fibrosis gene will be briefly described.

The identification of the cystic fibrosis (CF) gene required the combination of many of the techniques described. Unlike other cases of successful positional cloning prior to the identification of the CF gene, no chromosomal rearrangements or deletions could be identified with conventional techniques in affected individuals. The gene was ultimately assigned to chromosome 7, band q31, with linkage analysis studies. Two flanking markers, MET and D7S8, allowed cloning techniques to be employed. Starting from these markers, a combination of chromosome walking and chromosome jumping (a modified form of chromosome walking (Watson *et al.*, 1992)) was used in a bi-directional walk. This enabled investigators to clone and map a region of over 500kb (Rommens *et al.*, 1989). The positioning of markers within this map was done using a combination of linkage analysis, somatic cell hybrid mapping and long-range restriction mapping with PFGE. Candidate coding sequences were isolated by zoo blot hybridisations. These were then systematically used to screen lung and sweat gland epithelium cDNA libraries. Many clones were identified which were then excluded by sequence analysis from being candidates for the CF gene. Exhaustive cDNA

library screening yielded one clone that was shown by sequencing and digestion with Hpa II and Msp I restriction enzymes to contain a CpG island. This clone ultimately corresponded to the CF gene. The full transcript was identified by extending the cDNA using 5'- and 3'- RACE PCR (Riordan *et al.*, 1989). The complete cDNA was 6129 base pairs in length (excluding the poly $(A)^+$ tail). A 3-bp deletion at amino acid position 508 resulting in the loss of a phenylalanine residue was shown to be present in affected individuals in 70% of cases (Kerem *et al.*, 1989). As a result of this massive effort, this 3-bp deletion and a few other mutations identified in the remaining 30% of patients can now be detected by PCR analysis and offered as prenatal diagnostic tool to affected families.

1.4.7. The positional candidate approach

The third approach to cloning a disease gene is to look at known genes that may be good candidates from within an area of interest and to characterise them for mutations. The gene mutated in the ataxia-telangiectasia syndrome is located at 11q23. People with this syndrome suffer from cerebellar ataxia, immune deficiency and occulocutaneous telangiectasia. Affected patients are also prone to lymphoid neoplasms. Among the chromosomal aberrations seen in T-cell prolymphocytic leukaemia (T PLL), deletions of 11q are relatively common. Stilgenbaeur *et al.* (Stilgenbauer *et al.*, 1997) sequenced the ATM gene in a series of patients with T PLL shown to have a heterozygous deletion of 11q21 and found the ATM gene to be mutated on the remaining allele in all cases. These data suggest that this gene functions as a tumour suppressor gene in normal individuals and is involved in the pathogenesis of T PLL.

The gene responsible for Marfan's syndrome was found using a similar approach. Linkage analysis in affected families had shown the gene responsible to be located at 15q15-21.3. A glycoprotein, fibrillin, had been shown to be encoded for by a cDNA from the long arm of chromosome 15. This was a good candidate for Marfan's syndrome, an inherited disorder of connective tissue, and missense mutations were shown to exist in this gene in patients with the disease (Dietz *et al.*, 1991).

This positional candidate approach will become more widely employed as public data from both the Human Genome Mapping Project and EST databases become more extensive. As transcripts from novel genes are identified and mapped, finding disease genes will become a matter of defining the location of the

gene and sequencing candidates from the region, using an 'educated guess' as to the possible functions of the gene in question. It has been suggested that this approach will ultimately supersede the positional cloning approach (Collins, 1995).

1.5. Detailed analysis of the structural abnormalities of chromosome 13 in B-cell CLL

The first report of abnormalities of the long arm of chromosome 13 in association with B-cell CLL was published in 1987 (Fitchett *et al.*, 1987). Both deletions and translocations were seen using conventional cytogenetic banding techniques. One of the early problems with cytogenetic studies in B-cell CLL was the difficulty in obtaining metaphases from malignant B cells. The use of mitogens such as phytohemagglutinin (PHA) were often unsuccessful, and it was not until the use of TPA (12-0-tetradecanoylphorbol-13-acetate) was reported in the late 80's (Oscier *et al.*, 1990) that chromosomal abnormalities became easier to detect, although the numbers of recoverable metaphases remain limited even with modern techniques. As described, large scale studies showed that abnormalities of 13q were the commonest structural abnormalities seen in B-cell CLL with groups reporting an incidence of 13-17% in all patients (Oscier *et al.*, 1990) (Juliusson *et al.*, 1990).

Juliusson *et al.* noted that of the 51 patients found to have abnormalities of chromosome 13, 23 had deletions of band 13q14 and a further 12 had a translocation involving this same band (Juliusson *et al.*, 1990). Three additional patients had deletions of 13q with unidentifiable breakpoints. The retinoblastoma gene (RB 1), had been discovered by positional cloning techniques to lie at 13q14 (Friend *et al.*, 1986). Interestingly, the RB 1 locus had also been shown to be homozygously deleted in a T cell leukaemic cell line (Cheng *et al.*, 1990) although it was unaffected in the other 12 leukaemic cell lines that were tested. Consequently, the RB gene was a good candidate for the tumour suppressor gene involved in this group of patients with B-cell CLL.

1.5.1. The retinoblastoma gene (RB 1) and B-cell CLL

In 1993 several reports suggested that another locus, marked by the microsatellite D13S25, was more likely to denote the critical region of deletion in

B-cell CLL: Brown *et al.* used a combination of somatic cell hybrids from patients with 13q translocations and PFGE, and were able to demonstrate that although RB 1 was deleted in some patients, the telomeric locus D13S25 was more frequently deleted (Brown *et al.*, 1993). They demonstrated heterozygous loss of this marker in 7 cases and homozygous loss in a further 4 cases. One informative patient had a translocation involving chromosomes 11 and 13. This translocation was associated with a deletion on chromosome 13 involving the region at D13S25. PFGE using DNA from this patient showed that a 530kb Not I fragment containing the RB 1 locus was not rearranged. However, an expected 1.3Mb Not I fragment containing the D13S25 locus was rearranged. Brown suggested that the region deleted in B cell malignancy (DBM) did not involve the RB 1 locus but another, more telomeric locus associated with D13S25. The existence of homozygous deletions was highly suggestive that the gene involved in the pathogenesis of B-cell CLL was a tumour suppressor gene.

Liu *et al.* extended these findings (Liu *et al.*, 1993), reporting 30% of their B-cell CLL cases (total 31 patients) to have heterozygous deletion of the RB 1 locus as shown by Southern blotting techniques. They carried out SSCP analysis on these patients to look for mutations in the remaining copy of the gene. Using 6 pairs of primers they were able to analyse the entire gene and no mutations were found. They then went on to look for other candidate regions at 13q14 where the tumour suppressor gene may be located. Probes for the RB 1 gene and the D13S25 locus were used. 14 patients (45%) showed loss of the D13S25 probe, of which 4 showed homozygous deletion. As previously stated, 30% of these had RB 1 heterozygous deletions. All the patients with homozygous deletion of D13S25 had at least one retained copy of the RB 1 gene. Interestingly, several of the patients with abnormalities detected by the D13S25 probe had normal karyotypes.

Hawthorn *et al.* again used somatic cell hybrids to look at the abnormal region of chromosome 13 (Hawthorn *et al.*, 1993). Cell lines were set up from 5 patients with known translocations involving the long arm of chromosome 13. DNA was extracted from these cell lines and tested by PCR analysis for the presence of STS markers covering the region at 13q14.3. Control markers for 13q12 and 13q34 were used. In four cases both RB1 and D13S25 were lost. However, in one case the RB 1 marker was retained but the D13S25 marker was lost, confirming previous reports. A further study from the same group (Chapman *et al.*, 1994) analysed DNA extracted from patients circulating malignant B cells

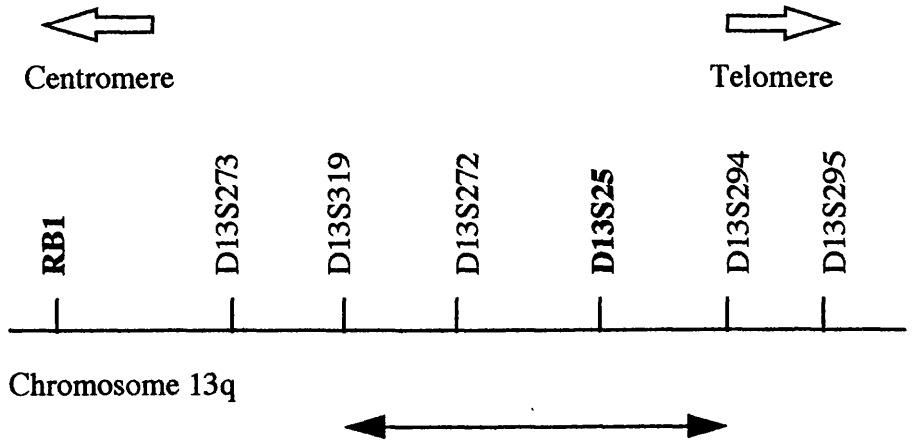
by LOH and Southern blot analysis. Polymorphic markers from RB 1 were used in the LOH studies, showing that whilst 41/68 patients were informative, only 9 (22%) showed evidence of LOH and none showed homozygous loss. No useful polymorphic markers for the D13S25 locus were available at this time (one has since been reported (Ibbotson *et al.*, 1996)) and this region was studied using Southern blot analysis and densitometry techniques. Of the 11 patients shown karyotypically to have 13q abnormalities, 10 had deletion of D13S25, 3 of which showed homozygous loss i.e. one patient did not show loss of this region despite a 13q deletion documented by banding techniques. Significantly, 11/15 of tumours with normal karyotypes also demonstrated allelic loss of the D13S25 probe and this loss was homozygous in 4 cases.

1.5.2. Mapping of the genomic region at D13S25

In June 1994, the Genethon map for each chromosome, including chromosome 13, based on radiation hybrids was published (Gyapay *et al.*, 1994). The chromosome 13 map was further refined by the mapping of additional STSs using radiation hybrid techniques (Shaw *et al.*, 1995). One group with an interest in 13q deletions in CLL compiled a comparison of the genetic and physical maps of 13q to date which served as a useful basis for further deletion mapping at the region of 13q14 (Hawthorn and Cowell, 1995). A YAC contig spanning the region of deletion in B-cell CLL from the RB 1 locus to a marker, D13S31, telomeric of D13S25 was then described (Hawthorn *et al.*, 1995). (see figure 1.8).

Devilder *et al.* mapped a series of microsatellites close to D13S25 and performed deletion analysis on a series of B-cell CLL patients (Devilder *et al.*, 1995). They showed that the D13S25 locus was deleted in all cases where loss of 13q14.3 was demonstrated. They also identified three novel microsatellite markers from the region. The markers D13S319 and D13S272 were positioned centromeric of D13S25, and D13S294 was shown to be telomeric of D13S25. In all cases of D13S25 deletion, D13S294 was also deleted. In addition, D13S319 and D13S272 were deleted in all but one patient. These markers defined the minimally deleted region in their patients. The distance between D13S319 and D13S25 was not known (figure 1.8).

Figure 1.8: Depicting the position of the markers published by Devilder *et al.* and the minimal region of deletion in their patients. The marker D13S319 is referred to as Mgg 15 in publications by Hawthorn *et al.*



Markers that define the minimal region of deletion.
Devilder *et al.* 1995

The marker, D13S319 was shown to be of further importance in the mapping of the minimal region of deletion at 13q (Liu *et al.*, 1995). Liu *et al.* carried out a study looking at the extent of deletions at 13q in B-cell CLL, Non-Hodgkin's lymphoma (NHL) and acute lymphocytic leukaemia (ALL). The analysis was performed by Southern blot using probes for the D13S319 and D13S25 loci, and with PCR primers to amplify the microsatellites at RB 1, D13S319, and D13S25. They were able to estimate the distance between RB 1 and D13S25 as not exceeding 1.4Mb and the distance between D13S319 and D13S25 as not exceeding 680kb. They showed that in all three lymphoid malignancies the D13S319 locus was the most frequently deleted. They found 6 CLL patients (from a total of 75 with B-cell CLL) that had heterozygous loss of D13S319 and no loss at D13S25. They concluded that D13S319 marked the vicinity of the tumour suppressor gene involved in the pathogenesis of B cell malignancies.

By contrast, Bullrich *et al.* observed that D13S25 was more frequently deleted in B-cell CLL than D13S319 (Bullrich *et al.*, 1996). This group assembled a detailed YAC contig of the region at 13q14.3 using STS markers. They used PFGE restriction mapping to re-evaluate the distances involved and tested patient samples for deletion using LOH of microsatellite markers by PCR and Southern

blotting. In addition, they defined many more microsatellite markers between D13S273 and D13S25. With data from 60 patients with B-cell CLL they narrowed the minimal region of deletion in these patients to 550kb, between a novel marker 206XF12 and D13S25 (see figure 1.9). The marker 206XF12 is telomeric of both D13S319 and D13S272. These data regarding the minimal region of deletion have since been confirmed by another group which, following assembly of a detailed physical map of the region, tested the area of deletion in their patients using FISH. They found the deleted region in their patients to lie between 206XF12 and D13S294 (Bouyge-Moreau *et al.*, 1997).

These findings were disputed by Kalachikov *et al.* who showed the minimal region of deletion in their patients to include the region at D13S319 but to exclude D13S25 (Kalachikov *et al.*, 1997). They also published a highly detailed physical map of the region with details of publicly available YAC and PAC clones covering the area. This map allowed the distances at 13q to be clearly defined. They described additional STS markers and numerous novel ESTs from the region. The minimally deleted region in the 57 patients tested by LOH studies was defined as being 300kb, between an STS 173a12-82 and a second STS 138G4t3. Although this region conflicted with that published by Bullrich *et al.*, there was a region of common overlap defined by these two groups that lay in the region between D13S319 and D13S25. Much of the detail of these maps is shown in chapter 3 but a schematic of the regions of deletion is shown below (figure 1.9).

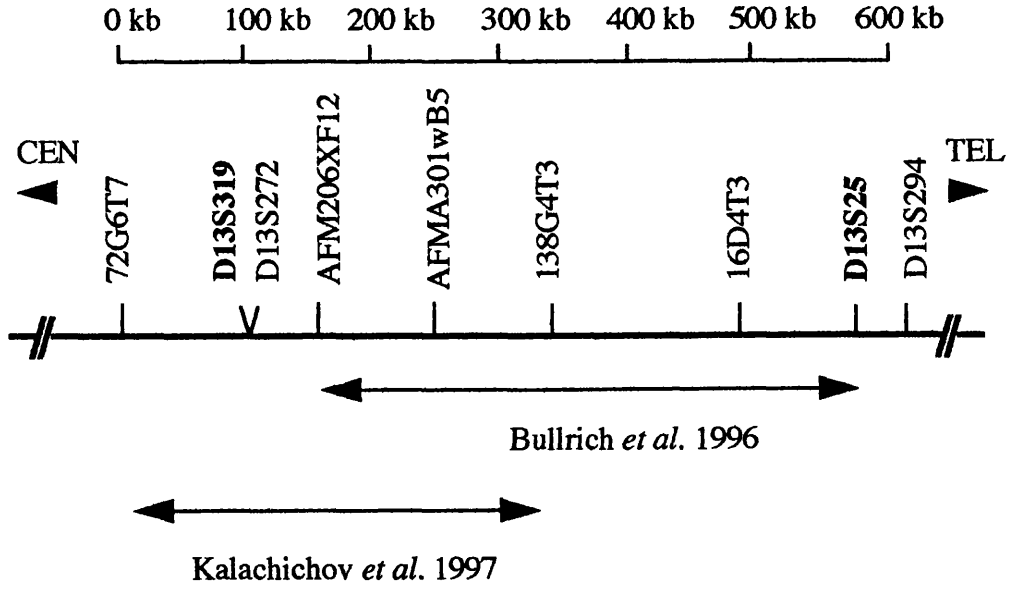


Figure 1.9: The conflicting areas of deletion at 13q14.3 and their region of overlap.

1.5.3. Candidate tumour suppressor genes from the region of deletion

Since 1997, attention has been focused on isolating candidate tumour suppressor genes from the region of deletion at 13q14.3. Many of the ESTs described by Kalachikov *et al.* were isolated using a combination of exon trapping and cDNA selection. All of the ESTs published in this paper were ruled out as belonging to a candidate tumour suppressor gene involved in B-cell CLL (Kalachikov *et al.*, 1997). Another group continued to map the deletions of an ever increasing number of patients with B-cell CLL. They constructed a physical map in cosmids, BACs and PACs spanning a 600kb region encompassing the deleted locus (Corcoran *et al.*, 1998). Corcoran *et al.* localised a series of CpG islands at the D13S319 region and, by identifying novel markers including the dinucleotide repeat 6E3.2, were able to define a 130kb minimally deleted region. This region was located centromeric of D13S319 and, whilst within the region of deletion described by Kalachikov *et al.*, was outside the region reported by Bullrich *et al.* (Bullrich *et al.*, 1996). Further work by the same group used this physical map of the deleted region to construct a restriction map. Using fragments from this map to screen DNA from 206 B-cell CLL patients by Southern blotting the region of deletion was narrowed to 10kb (Liu *et al.*, 1997). Four ESTs were found by sequencing within this area. Two were thought to be intronic sequence on further analysis (Leu 3 and 4), but two were shown to be part of 2 separate novel genes (Leu 1 and 2). The full length transcripts were identified by cDNA library screening and RACE PCR. Despite extensive mutational analysis in 170 patients for both of these genes, no mutations could be demonstrated in any patient with a heterozygous deletion of 13q14.3. This has now been corroborated by a second group (Rondeau *et al.*, 1999). The gene Leu 1 (deleted in leukaemia) is discussed in more detail throughout this work. Due to the location of these genes in the 10kb deleted region seen in their patients, Liu *et al.* state that Leu 1 and Leu 2 remain candidates for the tumour suppressor gene involved in B-cell CLLb (Liu *et al.*, 1997).

A fifth gene, Leu 5, has also been described which lies centromeric of their region of deletion (Kapanadze *et al.*, 1998). It has interesting features in that it shows homology to the zinc finger domain of the tumour suppressor gene BRCA 1. It is, therefore the first gene discovered from this region to show homology to another tumour suppressor gene. However, this gene lies outside the region of

deletion described by most groups so far and is unlikely to be the tumour suppressor gene involved in the pathogenesis of B-cell CLL.

Another group reported screening of 322 cases of B-cell CLL and 30 cases of mantle cell lymphoma (MCL) for deletion of D13S25 using FISH. The incidence of deletion within their patients was 51% in the CLLs and 70% in the MCL (Stilgenbauer *et al.*, 1998). They then performed detailed deletion mapping with probes designed from the area between RB 1 and D13S25. Interestingly, they identified two segments that were both frequently deleted: one at D13S272 (close to D13S319) and another 240kb distal from the first. They argued that the proximal region of deletion is likely to be the crucial one as all aberrations analysed in their series extended into this segment. Stilgenbauer *et al.* also point out that the regions are sufficiently close to potentially be inactivating the same gene if its exons spread over a sufficient genomic distance. They have identified candidate coding sequences from these regions using λ Get exon trapping and are currently analysing them to see whether or not they represent the gene.

At this present time the deleted region at 13q14.3 is heavily mapped and this map continues to be refined with the addition of STSs and, more importantly, ESTs (Bezieau *et al.*, 1998). The candidate location of the tumour suppressor gene involved in B-cell CLL should fall within the mapped region at 13q14.3. At the time of submission of this thesis a short presentation had just been given by Carlo Croce regarding an anti apoptotic gene (CLL 1) identified from 13q14.3 by his group. However, no details were given and no publications were available.

1.6. The project plan

In this introduction I have discussed briefly how cancer is thought to arise from a single cell due to the accumulation of genetic abnormalities. I have also discussed the importance of tumour suppressor genes in the emergence of a malignant phenotype in many tumours.

Whilst much is now known about the clinical phenotype of B-cell CLL, the underlying genetic defects remain obscure. Deletion of the long arm of chromosome 13 at band 14.3 has been shown to be the most frequent genetic abnormality in B-cell CLL and this region has been the focus of attention of many groups. The aim of this project was to define the minimally deleted region at 13q14.3 in our patients and to identify and characterise candidate tumour suppressor genes from within this region using positional cloning techniques.

Chapter 2. Materials and Methods

2.1. Preparation of DNA from transformed cells

2.1.1. Small scale plasmid preparations (minipreps); method 1

1.5ml aliquots of 2 x TY plus ampicillin (100μg/ml) were inoculated with a single bacterial colony containing the desired plasmid construct. Cocktail sticks were used to inoculate the cultures. These were incubated at 37°C overnight in a shaking incubator. The 1.5ml culture was transferred into Eppendorf tubes and spun at 13000 rpm in a MSE Microcentaur microfuge for 5 minutes to pellet the cells. Stocks were kept by streaking each cocktail stick on 2 x TY / ampicillin plates, which were then grown at 37°C overnight and stored at 4°C.

The cell pellets were resuspended in ice cold GTE and were then left at room temperature for 5 minutes. 200μl of freshly prepared 0.2M NaOH/1% SDS was then added to lyse the cells. The tubes were rapidly inverted and placed on ice for 10 minutes. The bacterial DNA was precipitated by adding 250μl of 3M Na Acetate pH 4.8. The tubes were shaken and incubated on ice for 15 minutes. The cell debris and *Escherichia coli* DNA was then pelleted by spinning in a microfuge for 5 minutes. The supernatant was transferred to a fresh tube and the plasmid DNA was precipitated with 2 volumes of 100% cold ethanol. The tubes were left at -20°C for 10 minutes to maximise the amount of plasmid DNA precipitated. The DNA was then pelleted by centrifugation, washed with 70% ethanol, dried in a vacuum dessicator for 5-7 minutes and resuspended in 30-50μl of TE.

5μl were analysed by restriction endonuclease digestion with the addition of 0.5μl DNAse free RNAse (10mg/ml). The plasmid inserts were then characterised by agarose gel electrophoresis and visualised by ethidium bromide staining on a UV light transilluminator.

2.1.2. Small scale plasmid preparations (minipreps); method 2

The miniprep cultures were set up as for method 1 in 1.5ml cultures with appropriate antibiotics. The following day the pellets were collected by centrifugation and stocks were made as already described. To prepare the DNA from the cell pellet, a Qiagen spin column miniprep kit was used. The

manufacturers directions were followed exactly, with the DNA being eluted in 30-50µl of TE. The plasmid inserts were analysed as described in 2.2.1.

2.1.3. Large scale plasmid preparations (maxi preps)

100ml of 2 x TY (LB for cosmid preps and PAC preps) with 100µg/ml ampicillin was inoculated with a single bacterial colony containing the plasmid (or cosmid) of interest and was grown overnight in a shaking incubator at 37°C. The following morning 600µl of the culture was used to prepare a 15% glycerol stock that was then stored at -70°C. The culture was transferred into two 50ml Falcon tubes and spun at 3000 rpm in a MSE Mistral 2000 centrifuge for 15 minutes. 10ml of cold GTE was added to the pelleted cells and incubated at room temperature for 10 minutes. The cells were then lysed by adding 20ml of freshly made 0.2M NaOH / 1% SDS to each tube. The contents of the tube were quickly mixed and incubated at room temperature for a further 5 minutes. To precipitate the *Escherichia coli* DNA, 10ml of 3M Na Acetate pH 4.8 was added to each tube and the tubes were shaken vigorously before placing on ice for 15-60 minutes. The tubes were then spun as before and the supernatant was strained through gauze into a fresh Falcon tube. The DNA was precipitated by the addition of an equal volume of isopropanol. After incubating on ice for 10 minutes the DNA was pelleted by centrifugation as before. The pellet was dissolved in 2ml TE and 2ml of 5M LiCl was then added to precipitate the *Escherichia coli* DNA. After incubating on ice for 5 minutes the tubes were spun again and the supernatant was collected and transferred to a fresh Falcon tube. 2 volumes of ice cold 100% ethanol were added to reprecipitate the DNA. The pellet was collected by centrifugation after incubating on ice for 5 minutes. The pellet was resuspended in 2ml of TE and incubated with 4µl of 10mg/ml DNAse free RNAse, to destroy the *Escherichia coli* RNA, for 30 minutes at 37°C.

1ml of 2.5M NaCl, 20% PEG was added to collect the plasmid DNA from solution and the reaction was incubated on ice for 10 minutes. The DNA was then pelleted at maximum speed in a Microcentaur microfuge for 5 minutes, having been transferred to 1.5ml Eppendorf tubes. All excess PEG was removed and the pellets were resuspended in 500µl TE. 2 PCIA extractions and 1 chloroform extraction were performed. The plasmid DNA was precipitated with 2 volumes of 100% ethanol and 10% 3M Na Acetate at -20°C for 20 minutes. The precipitate was collected by centrifugation, washed with 70% ethanol and then dried in a

vacuum dessicator for 5-10 minutes. The pellet was resuspended in 100-200µl of TE depending on its size and the DNA yield was determined by spectrophotometry at OD_{260}nm.

2.1.4. Large scale PAC DNA preparations

A single bacterial colony was inoculated into 50ml 2 x TY medium containing 25µg /µl kanamycin and grown overnight in a shaking incubator at 37°C. A 15% glycerol stock was made the following day with 600µl of the culture and stored at -70°C. The cells were spun down at 3000 rpm for 15 minutes in a Beckman centrifuge. The DNA was then extracted following the method for plasmid DNA maxipreps 2.1.3.

2.1.5. Preparation of DNA from transformed yeast cells (YAC DNA maxiprep)

5ml of AHC media was inoculated with a single yeast colony and grown to stationary phase at 30°C in a shaking incubator (2-3 days). 400ml of AHC was then inoculated with 1ml of the stationary phase liquid culture and was grown at 30°C for a further 2-3 days. The culture was then centrifuged at 2500 rpm for 15 minutes in a Centrikon H-401B centrifuge to collect the yeast. The pellet was washed in 40ml of water and then spun again in a 50ml Falcon tube. The wet pellet was weighed. It was essential that the pellet weighed at least 3 grams at this stage. The pelleted cells were suspended in 3.5ml SCE containing 85µl of 2M DTT and 20mg Zymolase (Yeast lytic enzyme - ICN 100,000 U/GM). The mixture was incubated at 37°C, mixing every 10 minutes, for 1 hour. The spheroplast suspension was then directly added to 7ml of lysis buffer containing 100mg/ml Proteinase K (Boehringer Manheim). The lysate was then heated at 65°C for 15 minutes before being cooled quickly to room temperature in water. To collect the YAC DNA, the lysed yeast was spun through a sucrose gradient. This was made as follows; 11ml of 20% sucrose solution was placed in a 36ml Beckman SW 28 rotor tube, 11ml of 15% sucrose solution was then laid over the top and finally 3ml of a 50% sucrose solution was then laid at the bottom of the tube. The lysate was then poured on top of the gradient and spun for 3 hours at 20°C at 26K rpm in a SW 28 rotor. The viscous DNA was visible close to the bottom of the gradient after centrifugation. This viscous material was transferred

into a dialysis bag and dialysed overnight at 4°C against TE buffer. After 16 hours, the clear, viscous solution from inside the dialysis bag was transferred to a 50ml Falcon tube and an equal volume of butanol was added. The aqueous phase was collected by centrifugation at 2000 rpm for 5 minutes. This step was repeated until the volume was reduced to less than 5ml. The YAC DNA was then precipitated with 2 volumes of cold ethanol and 10% Na Acetate. The DNA was washed with 70% ethanol and air dried for approximately 1 hour. 50-200µl of TE was used to resuspend the pellet, depending on its size, and the DNA yield was determined by spectrophotometry at OD_{260}nm.

2.1.6. Phage DNA preparation

A phage plaque was picked and placed in a 1.5ml Eppendorf tube containing 300µl of λ diluent. After 30 minutes at room temperature 200µl of the diluent was added to a 1 in 50 dilution of Q358 cells which had been grown overnight at 37°C in 2 x TY plus 10mM $MgSO_4$ and 0.2% maltose. The cells were then returned to a shaking incubator at 37°C and grown overnight. The following day the culture was spun at 1000rpm for 10 minutes in a MSE Mistral 2000 centrifuge. The supernatant was removed and stored at 4°C. This was the phage high titre stock (HTS).

1ml of HTS was added to 100ml of a 1 in 100 dilution of Q358 cells grown to static phase at 37°C overnight in 2 x TY supplemented with 10mM $MgSO_4$ and 0.2% maltose. The culture was transferred to a 500ml flask and grown at 37°C until all the bacterial cells were lysed. This took approximately 6 hours. 5.9g of NaCl was then added to the culture and the flask was incubated on ice for 90 minutes. The flask contents were then equally distributed between four 50ml Falcon tubes which were spun at 3000rpm for 10 minutes. The supernatant was transferred to two 50ml Falcon tubes and 10% PEG 6000 was added. Once the PEG was dissolved, the tubes were incubated on ice for 60 minutes. The tubes were then spun again at 3000rpm for 10 minutes and the pellets were dissolved in 5ml of λ diluent. The contents of both tubes were transferred to a fresh tube and an equal volume of chloroform was added. The aqueous phase was collected by further centrifugation at 3000rpm for 10 minutes. 10% PEG 6000 was added to the aqueous phase and the mixture was incubated on ice for 2 hours. A pellet was collected as before and dissolved in 400µl of λ diluent. The solution was transferred to a 1.5ml Eppendorf and 25µl of DNase (1mg/ml) and 10µl of RNase

(10mg/ml) was then added. The tube was incubated in a water bath at 37°C for 30 minutes. An equal volume of chloroform was then added, the aqueous phase was collected following centrifugation at maximum speed in a Microcentaur microfuge for 5 minutes and was transferred to a fresh tube. 20µl of proteinase K (Boehringer Mannheim 5mg/ml) and 0.5% SDS was then added and the tube was incubated in a water bath at 65°C for 60 minutes. The aqueous phase was extracted as before with phenol and then a second time with PCIA. The phage DNA was then precipitated with 100% ethanol at -20°C and 10% sodium acetate. The DNA pellet was then washed and resuspended as for the plasmid DNA in section 2.2.3.

2.2. Restriction endonuclease digestion of DNA

For mapping and analysis of plasmid, cosmid, PAC and YAC DNA with restriction endonucleases, 500ng - 3µg of DNA were digested in a total volume of 30 - 50µl, with a 5-10 fold excess of enzyme. All digests were set up in 1 x 'Carlo's' restriction buffer with 1mM DTT and 3mM spermidine. Digests were incubated at 37°C for a minimum of 3 hours. When specific fragments were required for probe preparation or subcloning, the digests were scaled up accordingly with the volume of enzyme added not exceeding 10% of the total volume.

For preparation of genomic DNA Southern blots, 10µg of DNA were digested in a 50µl volume with a 10 fold excess of enzyme for at least 5 hours at 37°C unless otherwise stated.

2.3. Agarose gel electrophoresis of DNA

All DNA samples were mixed with 0.2 volumes of loading dye before electrophoresis. Most restriction digests of cosmid, plasmid, PAC and YAC DNA were resolved in 0.8% agarose gels containing 1µg/ml ethidium bromide. Genomic DNA digested for Southern blot analysis was also resolved in 0.8% gels with ethidium bromide. For the high molecular weight genomic DNA digests and any digests required for accurate mapping, 1 x AGB gels (24.5 cm) were run at 40 volts for 18-24 hours in a Flowgen tank. Other digests were run on 0.5 x TBE (12.5 cm) gels at 100 volts for 1-1.5 hours in a Hybaid tank. Small PCR products and other small digestion fragments were resolved on higher concentration agarose

gels i.e. 1-2% gels. The size markers used were either λ bacteriophage cut with Hind III (NBL Gene Sciences) or ϕX 174 cut with Hae III(GibcoBRL).

2.3.1. Pulse Field Gel Electrophoresis (PFGE)

This method of electrophoresis to obtain separation of DNA fragments larger than 20kb e.g. undigested YAC DNA or genomic DNA digested with rare cutting restriction enzyme, was first described by Schwartz *et al* (Schwartz and Cantor, 1984). In this study PFGE was primarily used to analyse YAC clones. 5ml cultures of yeast were grown in AHC plus 2% glucose for 3 days at 30°C. The cells were then pelleted by centrifugation at 2000 rpm for 5 minutes. The pellet was washed twice in 0.05M EDTA (pH 8.0) before being resuspended in 200µl of solution A (5ml SCE, 170µl 1mM DTT, 5mg Zymolase enzyme, 100,000 U/GM). The mixture was then incubated at 37°C for 15 minutes.

200µl of 1% low melting point agarose (BioRad) in 0.125M EDTA, pH 8.0 was added to the cells and maintained at 50°C. Blocks for electrophoresis were prepared by pouring the molten agarose plus lysed cells into moulds. When the blocks were set they were removed from the moulds and 4ml of solution B (22.2ml 0.5M EDTA, pH 9.0, 0.24ml 1 M Tris, pH 8.0 and 1.2ml 1M DTT) was added. The plugs were then incubated at 37°C overnight.

The following day the blocks were removed from solution B and added to 4ml of solution C (21.6ml 0.5 M EDTA, 2.4ml 10% sarcosyl, 0.24ml 1M Tris, pH 8.0 and 24mg Proteinase K). The blocks were incubated in solution C at 50°C overnight. The blocks were then stored at 4°C in 0.5M EDTA with 30ml 10mg/ml RNase until use.

For electrophoresis, each YAC block was added to 200µl of TE with 30µl of loading dye in a 1.5ml Eppendorf and the dye was allowed to diffuse into the agarose. The blocks were then loaded into preformed wells of a 1% agarose 0.5 X TBE 10cm x 10cm gel. A YAC size marker was also loaded onto the gel (New England Biolabs, catalog no. 343). The gel was electrophoresed for 72 hours in a BioRad pulse field gel tank at 90 volts with a current alternating every 120 seconds. The buffer was continuously cooled to 14°C via a pump and circulating refrigerated water bath. Following electrophoresis, the gel was stained with ethidium bromide by soaking the gel in 0.5µg/ml ethidium bromide in 0.5 x TBE for 30 minutes. The DNA was then visualised under UV light and the YAC size

marker was marked with ink, using a 19 gauge needle. The gel was then Southern blotted and the filters hybridised as for other nylon filters.

2.4. Southern Blotting

After electrophoresis, a photograph of the gel was taken and the size markers were highlighted with ink, using a 19 gauge needle. The gel was then placed in depurination solution with gentle shaking for 15-20 minutes. Depurination of sites within the DNA allows cleavage during denaturation, fragmenting the longer pieces of DNA and enabling more efficient transfer to the membrane of the high molecular weight DNA. This step was omitted if all fragment sizes were below 10kb. The gel was then rinsed in distilled water before soaking in denaturing solution for 30 minutes. Following this, the gel was transferred into neutralising solution for a further 30 minutes. The gel was then capillary blotted onto Hybond nylon filters (Amersham Life Sciences) overnight (1 hour for small plasmid blots). The DNA was fixed to the filters by UV fixation (Stratalinker) and / or baking at 80°C in a vacuum oven for 2 hours. The filters were rinsed in 3 x SSC before hybridisation.

2.5. Preparation of DNA probes

Following restriction endonuclease digestion of cosmid or plasmid clones containing fragments to be used as probes, the fragments were resolved on agarose gels. The fragment of interest was excised under UV light with a scalpel blade. The DNA was then purified from the agarose slice using a Jetsorb DNA extraction kit (Genomed). The manufacturers instructions were followed and the DNA was eluted in a final volume of 40µl of TE. Probes to be used on cDNA libraries were double purified by electrophoresing the DNA for a second time, excising the band and recovering the DNA as before. This was done to remove all traces of vector DNA that might cross hybridise to the library vector.

1µl of probe DNA (approximately 20-40ng of DNA) was diluted to 48µl with distilled water and boiled for 5 minutes before placing on ice. The 48µl was then collected by brief centrifugation and added to a pre-prepared Rediprime (random prime labelling, Amersham Life Sciences) tube for labelling DNA. 2µl of [α -^{32}P]dCTP (100 µCi /µl) was added and the reaction was incubated at 37°C for 30-60 minutes. The reaction was diluted with water to 100µl and

unincorporated nucleotides were separated from the DNA probe by centrifugation through a G50 Sephadex column. This was prepared in a 1ml syringe plugged with glass wool and equilibrated with distilled water. The labelled probe was then diluted again to 500µl and boiled for 5 minutes to denature the DNA before being added directly to the hybridisation solution.

For probes containing repeat sequences the probe would be competed prior to hybridisation as follows. After labelling of the probe, 200µl of placental DNA (1µg/µl - Sigma), 300µl of 20 x SSC and 400µl distilled water were added. The 1ml mixture was then boiled for 5 minutes before incubating at 65°C for 2 hours to allow repeat sequences to anneal. The probe was then added directly to the hybridisation mixture.

2.6. Kinase end labelling of oligonucleotides

Primer sequences of 25-40bp were occasionally used as probes for hybridisation. These were end labelled using T4 kinase and ^{32}P γ–ATP. A 20µl reaction was set up to label 100ng of oligonucleotide using 2µl of 10 x kinase buffer, 2µl of ^{32}P γ–ATP (100µCi /µl) and 1µl of T4 kinase (New England Biolabs 10,000U/ml). This mixture was incubated at 37°C for one hour. The reaction was diluted to 100µl with distilled water and unicorporated nucleotides were separated by centrifugation through a G50 sephadex column. The labelled probe was then added directly to the hybridisation solution. Hybridisations were carried out at room temperature overnight in 6 x SSC, following a 2 hour prehybridisation of the filter in Church buffer at 65°C.

The following day, washes were in 6 x SSC / 0.1% SDS. Washing was started at room temperature and the temperature was gradually increased. The final washing temperature was determined by the melting temperature (Tm) of the oligonucleotide minus 12°C (Sambrook *et al.*, 1989). The signal was detected by autoradiography as described in section 2.7.

2.7. Filter hybridisation

Filters were prehybridised at 65°C for 1-2 hours in hybridisation solution. Hybridisation of cDNA libraries and Southern blots of genomic DNA were hybridised in 'Cambridge' buffer. All other filters were hybridised in Church buffer. Labelled probe was then added to the hybridisation solution and incubated

with the filters. All YAC filters, cDNA library filters and genomic Southern blots were hybridised overnight. Filter lifts from transfected competent cells and plasmid / cosmid filters were hybridised for 4 hours.

The stringency of the washes depended on the hybridisation that had been performed. Filters that had been hybrdised in 'Cambridge' buffer were washed at 65°C for at least 40 minutes in SSC, up to stringencies of 0.1 x with 0.1% SDS. Filters that had been hybridised in Church buffer were washed in 3 x SSC / 0.1% SDS, initially at room temperature for 10 minutes and then at 65°C for 10 minutes. When a cross species hybridisation was being performed i.e. a human probe hybridised to a mouse genomic DNA filter, lower stringencies of wash were used, typically 6 x SSC at 65°C. If, following washing, filters showed residual background radioactivity i.e. above 15-20 counts per second as detected by a Geiger counter, they would be washed further at a higher stringency before being exposed.

Filters were exposed to Fuji Medical X ray autoradiography film (RX) at -70°C. The time of exposure varied from 2 hours to 14 days depending on the intensity of the signal. The position of positive signals could be identified using orientation marks and the size of hybridisation bands were assessed by comparison to marked DNA molecular markers.

2.8. Cloning of DNA into plasmid or phagemid vectors

2.8.1. Vector preparation

1µg of uncut vector e.g. Bluescript plasmid or M13mp18 phagemid (Boehringer Mannheim) was digested in a 50µl digest with the appropriate restriction nuclease enzyme/s required. After 3 hours incubation at 37°C a 5 µl aliquot was electrophoresed on a 0.8% agarose gel with 50ng of uncut vector in parallel with a molecular weight marker, to check for complete digestion of the vector. The reaction was then placed at 65°C for 10 minutes to inactivate the enzymes before being placed briefly on ice. To dephosphorylate the vector ends, 2-4µl of calf intestinal alkaline phosphatase (Boehringer Mannheim 1U/µl) was added and the reaction was incubated again at 37°C for 1 hour. The reaction was then diluted with TE to 100µl and mixed with 100µl of PCIA. The aqueous phase was collected by centrifugation and was centrifuged through a G50 Sephadex, column that had been equilibrated with TE, to remove residual phenol. The vector

was stored at -20°C and 2µl (approximately 20ng) of vector was used for each ligation reaction.

2.8.2. Insert preparation

These methods are for preparing inserts with overhanging (sticky) ends. For preparation and cloning of blunt ended fragments, see 2.8.4.

Depending on the size of the fragment to be cloned, 5-20µg of DNA was digested with 5-10 fold excess of appropriate restriction enzyme under optimal conditions. The digested products were run on an ethidium bromide agarose gel of a concentration appropriate to separate the fragment of interest. The band to be cloned was then excised using a scalpel under UV light and the DNA was purified using a Jetsorb (Genomed) kit. The resulting pure fragments were eluted in 30-40µl of TE.

If the fragment to be cloned was a PCR product, then the PCR primers were designed with restriction enzyme sites at the 5' end. 5µl of the PCR was analysed by electrophoresis on an agarose ethidium bromide gel. The rest of the PCR product was then purified using a Qiagen PCR clean up kit. The DNA to be cloned was eluted in 40µl of sterile water and the total DNA was digested in a volume of 50µl with the appropriate restriction enzymes under optimal conditions. 15-20µl of the digested insert was then electrophoresed on an agarose gel (at the appropriate concentration), and the DNA band to be cloned was excised under UV light and purified as previously described.

2.8.3. Ligation

Two ligation reactions were set up for each insert, usually with 1-3 molar excess of insert DNA and 2µl (10µg/µl) of vector, plus a control ligation with vector only. Ligations were carried out in a total volume of 20µl using 200-400U of T4 DNA ligase (New England Biolabs 4000,000 U/ml) in 1x ligase buffer (New England Biolabs) and 1mM ATP. The ligations were incubated at 14°C overnight. The ligations were then stored at -20°C until required.

2.8.4. Sonication of DNA and shotgun cloning into M13mp18

To facilitate sequencing of large Bluescript DNA inserts, the clones were subcloned into the vector M13mp18 following sonication as follows:

The plasmid DNA was prepared as described in section 2.1.3. 15µg of DNA was diluted to 100µl with TE in a 500µl Eppendorf tube and sonicated using an Ultrasonic Processor on full setting for 40 seconds, before placing on ice for 1 minute. This cycle was repeated 5 times. The size of the fragments was estimated by running 5µl of the sonicated DNA in a 1.5% agarose gel with ϕX 174-Hae III marker. If the fragments were larger than 600 bp then further cycles of sonication were performed and the sample was re-analysed by electrophoresis.

The ends of the fragmented DNA were filled in prior to blunt cloning. 15µl of the sonicated DNA was added to the following:

3µl 10mM dNTPs
3µl 10 x 'Carlo's' buffer
3µl 15mM DTT
2µl Klenow (sequencing grade
Boerhinger Manheim 5U/µl)
4µl distilled water

The reaction was incubated at 37°C for 30 minutes before heat inactivation of the Klenow at 65°C for 10 minutes. The sonicated DNA was then ligated into an M13mp18 vector prepared by digestion with the restriction enzyme Sma I at 25°C for 5 hours. The following day the ligations were transfected into competent cells. Positive colonies for sequencing were identified by hybridisation to the original plasmid insert DNA that had been radiolabelled. By sequencing a series of smaller overlapping clones the full sequence of the larger fragment was assembled using the DNAStar software program.

2.9. Transformation of bacterial cells with recombinant clones

2.9.1. Preparation of Hannahan's competent cells

A single colony of the TG1 strain of *Escherichia coli* cells was inoculated in 5ml of SOB medium containing 2% glucose and 10mM $MgCl_2$. This was then grown overnight to stationary phase in a shaking incubator at 37°C. 500µl of the overnight culture was diluted in 50ml of SOB medium in a sterile 250ml flask.

500µl of 20% glucose and 500µl of 1M $MgCl_2$ were also added and the flask was returned to the 37°C incubator. The culture was then grown to an OD_{600nm} of 0.5-0.55. The cells were placed on ice for 15 minutes and then the culture was transferred to a 50ml Falcon tube. The cells were pelleted by centrifugation at 1600 rpm for 10 minutes. The pellet was then resuspended in 18ml of cold TFB and left on ice for 10 minutes. The cells were then spun as before and the pellet resuspended in 4ml of TFB with 140µl of DnD. The cells were incubated on ice for 10 minutes before a second 140µl of DnD was added. After a further 15 minutes incubation on ice the cells were ready for transfection.

2.9.2. Transfection of competent cells

200µl of competent cells were then mixed with 8µl of ligation in a 1.5ml Eppendorf, incubated on ice for 45 minutes and then heat shocked in a 42°C water bath for 2 minutes.

For Bluescript transfections, 400µl of 2 x TY were added and the cells were incubated at 37°C for 30 minutes to allow expression of the ampicillin resistance gene prior to plating. The whole culture was then plated onto 2 x TY agar plates supplemented with ampicillin (100µg/ml). For M13 ligations, the transfected cells were transferred to a 5ml glass tube prior to adding 3ml of 2 x TY top agar at 50°C. The cells plus agar were quickly poured onto 2 x TY agar plates. The plates were then incubated overnight at 37°C.

2.9.3. In situ colony hybridisation

In order to select recombinants containing the desired insert, plaques or colonies were screened by in situ hybridisation as follows. Lifts were taken from each plate onto Whatman 1MM paper filters. The filters were laid over the plate surface for 1-2 minutes. Orientation marks were made using a 19 gauge needle and black ink. The filters were then carefully removed from the plates. For M13 transfections the DNA was fixed to the filters by UV fixation and used directly.

For Bluescript colonies, the filters were placed on 3MM Whatman paper soaked in 2 x SSC / 5 % SDS for 2 minutes to lyse the cells. DNA was then transferred and fixed to the filters by microwaving on the highest setting for 2.5 minutes. Further fixing was carried out by UV cross linking. After rinsing the filters in 3 x SSC, the filters were ready for prehybridisation.

The filters were then hybridised to the appropriate labelled DNA probe.

2.10. Manual Sequencing of single stranded plasmid DNA

2.10.1. Single stranded plasmid DNA preparation (ssDNA)

Colonies of interest for sequencing of plasmid or phage DNA were picked using a sterile cocktail stick into a glass tube containing 1.5ml of 2 x TY. If plasmid DNA was to be rescued, ampicillin at a concentration of 100µg/ml was added. No antibiotic was required for the M13 cultures. These cultures were then incubated at 37°C in a shaking incubator overnight.

The following morning new glass tubes containing 1.5ml of 2 x TY (plus antibiotic if required) were prepared. A 1:100 dilution from the overnight culture was then set up the following day. The tubes were then returned to the 37°C incubator. The M13 cultures were grown for 5.5 hours. The plasmid cultures were grown for 2 hours and then 3µl of Helper Phage (VCS-M13, Stratagene catalog no. 200251-81) was added to each tube to rescue the single stranded copy of the DNA. The cultures were then incubated for another 5.5 hours at 37°C with vigorous shaking.

The cultures were then transferred to 1.5ml Eppendorf tubes and spun for 5 minutes at maximum speed in a MSE Microcentaur microfuge. The supernatant was transferred to a new 1.5ml tube containing 200µl of 20% PEG 2.5M NaCl. The tubes were then left at room temperature for a minimum of 10 minutes before spinning again at maximum speed for 5 minutes. The pellet was carefully dissolved in 100µl of TE and 200µl of PCIA was added. The mixture was vortexed for 5 seconds before spinning at 13000 rpm for 5 minutes to collect the aqueous phase. The aqueous phase was transferred to a fresh tube and the previous step was repeated. 100µl of CIA was then added to the aqueous phase to remove residual phenol. The two layers were mixed by vortexing and the tubes were then spun again. The top layer was transferred to a new tube and 10% Na-Acetate and 250µl 100% ethanol were added to precipitate the single stranded DNA (ssDNA). The DNA was incubated on ice for at least 20 minutes before being recovered by centrifugation at 13000 rpm for 5 minutes. The pellet was washed with 70% ethanol and then air dried. Each pellet was then resuspended in 30µl TE and stored at -20°C until required.

2.10.2. Manual DNA sequencing

The dideoxy chain termination method for manual sequencing was originally described by Sanger *et al.* (Sanger *et al.*, 1977). All sequencing reactions were performed in U bottom microtitre plates. 2µl of ssDNA from each clone were transferred into 4 vertical wells. 2µl of the following master mix were then added to each well; 1µl reaction buffer, 1µl M13 -40 primer and 6µl water. The plates were then spun briefly in a MSE Mistral 2000 centrifuge to collect the reaction mix at the bottom of the well, covered in Saran wrap and incubated at 55°C for 30 minutes to allow the primer to anneal to the template. During this time a mix was made up for each clone as follows;

0.4µl 7.5µM dNTP
0.4µl 100mM DTT
6.5µl water
35S-dATP 0.5 µl
0.25µl sequenase enzyme (Amersham Life Sciences 13U/ml)

After incubation at 55°C, 2µl of the above solution were transferred to each well. The plates were spun again and incubated at room temperature for 10 minutes. 2µl of one of four stop nucleotide mixes (Amersham Life Sciences 250ul) i.e. T, C, G or A were added to each of the four wells. The plates were briefly spun as before and incubated at 37°C for 6 minutes. The reactions were terminated by adding 4µl of formamide stop solution to each well. The reactions were incubated at 80°C for 20 minutes to denature the DNA, prior to loading the sequencing gel.

M13 -40 primer sequence
5'- GTTTTCCCAGTCACGAC -3'

2.10.3. Acrylamide gel electrophoresis for sequencing

Sequencing electrophoresis was performed using 40 x 50 x 0.4cm gradient gels. Before pouring the gel, the plates were siliconised with Repelcote. The plates were separated by 0.4mm spacers and fastened together with tape and bulldog clips. 30µl of 10% ammonium persulphate solution (APS) and 13µl of Temed (Amresco) was added to 15ml of 2.5 x TBE acrylamide gel mix and the reaction

was poured into the plates using a 50ml syringe. This was followed by the second mix of 50ml 0.5 x TBE acrylamide gel mix, 100µl APS and 65µl Temed. A comb to form the wells was placed in position and the gel was allowed to set for 30 minutes. The tape and bulldog clips were then removed and the gel was then placed in a Gibco BRL S2 gel tank, 1 litre of 0.5 x TBE was used as buffer and 4µl of each sample reaction loaded into each well. The gel was run at 30 volts for 3.5-4 hours. The plates were then disassembled and the back plate was removed. The acrylamide gel was transferred to 3MM Whatman paper and covered with Saran wrap before being dried under vacuum at 80°C on a gel dryer for 2 hours. The dried gel was then exposed overnight at room temperature to Fuji Medical autoradiography film. Sequences were analysed using DNAStar computer software.

2.11. Automated sequencing of double stranded plasmid DNA

This technology allows direct sequencing of a plasmid insert from double stranded DNA. The machine used during the latter part of this project for automated sequencing was an ABI Prism 377 (Perkin Elmer). The DNA template was labelled with 4 fluorescent dyes, 1 for each nucleotide, using a PCR primer extension method. The fragments were then electrophoresed through a denaturing acrylamide gel to separate them according to size. At the bottom of the gel a laser beam continuously scans the gel exciting the fluorescent dyes attached to each fragment. The dyes emits light at a specific wavelength and the light is collected and separated by a spectrograph onto a cooled, charge coupled device. The computer software then collects these data and stores them for processing as readable sequence.

2.11.1 Reactions for automated sequencing

All plasmid DNA to be sequenced was prepared from 1.5ml overnight cultures using the Qiagen spin column miniprep kit as already described. The DNA was quantitated by spectrophotometry at OD_{260nm} and 400ng were used for each reaction. The DNA was placed in a 250µl tube and was made up to 8µl with sterile water. The following was then added;

1µl Big Dye (Perkin Elmer)
7µl 2.5x sequencing buffer

4µl primer at 0.8 pmol/µl

The Big Dye contains a set of 4 fluorescently labelled dye terminators as well as AmpliTaq DNA Polymerase. Although any primer can be used in the sequencing reaction, the M13 -20 and T3 primers were frequently used;

M13 -20 primer

5'- GTAAAACGACGGCCAGTG -3'

M13 T3 primer

5'- AATTAACCCTCACTAAAG -3'

The contents of the tube were mixed by gentle pipetting before being placed in a Perkin Elmer Gene Amp PCR system. The PCR profile used for sequencing reactions was as follows:

25 cycles at:	96°C	10 seconds
	50°C	5 seconds
	60°C	4 minutes

The DNA was then precipitated in 50µl of 100% ethanol plus 2µl 3M sodium acetate on ice for 20 minutes. The DNA was pelleted by centrifugation at 13000 rpm in a MSE Microcentaur microfuge for 30 minutes. The pellet was washed in 70% ethanol, briefly vortexed and re-collected by spinning in a microfuge at 13000 rpm for a further 5 minutes. The DNA pellet was vacuum dried for 5 minutes before being resuspended in 8µl of formamide loading buffer (5 parts deionised formamide - Amresco, and 1 part loading dye - Perkin Elmer).

Just prior to loading the sequencing gel, the DNA was denatured by incubation at 98°C for 2 minutes. The denatured DNA was then kept on ice until loading.

2.11.2. Electrophoresis using an ABI 377 automated sequencer

All methods used for the ABI sequencer (Perkin Elmer) were in accordance with manufacturers instructions. The plates were assembled using Perkin Elmer apparatus with 0.2mm spacers. The gel mix was then made up with 18g urea dissolved in 25ml double distilled water. 5g of mixed-bed ion exchange resin (AG501-X8 Resin, Bio Rad) was then added. The solution was vacuum filtered to

remove the resin and to degas the gel mix. 5ml of filtered 10 x TBE and 5.2ml of the Long Ranger Acrylamide (FMC Bio Products) were then added and the mix was made up to a final volume of 50ml with double distilled water. Prior to pouring, 250µl of 10% APS and 25µl of Temed (Amresco) were added. The gel was allowed to set for 2 hours before use.

2.12. cDNA library screening

cDNA libraries from different tissues were utilised during this study and these are described in detail in the relevant chapters. Both plasmid and bacteriophage cDNA libraries were screened.

2.12.1. Screening plasmid cDNA libraries

The titre of each library was first determined by preparing a 10^{-2} dilution of the frozen stock and plating out serial dilutions on LB agar plates containing the appropriate antibiotic selection as required. These plates were incubated at 37°C overnight. The number of colonies obtained for each dilution were counted and the volume of the original 10^{-2} dilution required to yield 50 000 colonies (optimal number for a 132mm plate) was calculated. The appropriate volume of the 10^{-2} was then spread onto 12-15 132mm Hybond membranes (Amersham Life Sciences) which had been laid on the surface of LB agar plates with antibiotics. These master plates were then grown at 37°C overnight and used the next day to prepare replica plates. One replica filter was prepared from each master. The first master was removed from its agar plate and placed on two pieces of Whatman 3MM paper. Fresh Hybond membranes were wetted on new LB agar plates (plus antibiotics). The prewetted replica was then placed on top of the master, 2 more pieces of Whatman 3MM were laid over the top and the two filters were pressed together. The master was marked with a pattern of spots using a marker pen and the same pattern was also marked on the replica plate. The filters were pulled apart and each was placed on fresh LB agar plates. This process was repeated with all the master plates. All plates were incubated at 37°C for 4 hours. The master plates were then put at 4°C whilst the replicas were left at room temperature overnight to allow further colony growth.

Once the colonies had regrown, the replicas were processed. Each filter was placed colony side up on a piece of Whatman 3MM soaked in denaturing

solution for 10 minutes. The filters were blotted on Whatman 3MM and then placed in neutralising solution for two minutes. They were then blotted again and washed in 3 x SSC. DNA was fixed onto the filter by UV cross-linking and then by baking at 80°C for a minimum of 2 hours. The filters were scrubbed with a tissue whilst soaking in 3 x SSC to remove all residual cellular debris before being hybridised in 'Cambridge' buffer at 65°C overnight to the probe of interest (all probes used for cDNA library screening were double purified as described in section 2.5).

After hybridisation the presence of a positive hybridisation signal on the replica filters, as revealed by autoradiography, was used to identify the positive colony on the master plate. A 5mm area around the colony was scraped off using a sterile inoculating loop into 1ml of LB media containing antibiotic. This was then mixed and 10 fold dilutions were made which were plated onto 82mm LB agar plates plus antibiotic. The stock tube was stored at 4°C and the plates were incubated at 37°C overnight. The next day lifts were taken from those plates with well separated colonies. A Hybond membrane was laid over the colonies and orientation marks were made on the filter and the plate using a 19 gauge needle and ink. The secondary filter was then processed in exactly the same way as the primary replica filter had been. The filters were hybridised again under the same conditions with the probe of interest. The plates were incubated at 37°C for 4 hours and then stored at 4°C until required.

Single positive colonies from secondary screening were then picked into 50ml LB media plus antibiotic and DNA from the cDNA clone was prepared using the plasmid maxiprep DNA protocol (section 2.1.3). The insert was characterised by restriction endonuclease digestion and electrophoresis on an agarose gel stained by ethidium bromide.

2.12.2. Screening bacteriophage cDNA libraries

The methods for bacteriophage vector libraries differ in several ways from plasmid vector library screening. 50ml of LB media supplemented with 0.2% maltose and 10mM $MgSO_4$ was inoculated with a single colony of *Escherichia coli* XL1-blue and the cultures grown overnight at 37°C. The following day the cells were pelleted by centrifugation for 10 minutes at 2000rpm. The pellet was then resuspended in 10mM $MgSO_4$ to an OD_{600nm} of 0.5. Dilutions of the bacteriophage library were made in SM buffer. 1ml of each dilution was added to

200µl of cells and incubated at 37°C for 15 minutes. 3ml of NZY top agar was added (at 50°C) and the agar and cells were poured onto NZY plates. The next day the titre of the library could be calculated. The library was plated on large (132mm) NZY plates to achieve approximately 50 000 plaques per plate. For each plate 600µl of XL1-blue cells in $MgSO_4$ at an OD_{600nm} 0.5 were incubated with the appropriate amount of lambda phage for 15 minutes before plating in 10ml of NZY top agar. After incubation for 6-8 hours at 37°C, filter lifts were taken directly from the master plates. Hybond filters were placed onto the master plates and left for 2 minutes. Orientation marks were made using a needle and ink. The filters were then removed and placed face up in denaturing solution for 2 minutes before transferring to neutralising solution for 5 minutes. They were then rinsed in 0.2M Tris.HCl pH 7.5 / 2 x SSC. The filters were then baked for 2 hours at 80°C to allow DNA fixation. Hybridisation was performed as for plasmid library hybridisations. Positive colonies were picked into 1ml of SM buffer and vortexed. Secondary and tertiary screening of candidate positive plaques were set up by titrating the buffer and phage DNA against XL1-blue cells as before.

2.13. RNA extraction

This method for RNA extraction was originally described by Chomzynski and Sacchi, 1987 (Chomczynski and Sacchi, 1987). The following protocol is for RNA extraction from 1.5 x 10^7 cells. RNA was isolated using RNAse free reagents, glass and plastic ware.

The cells were initially pelleted, 500µl of GITC were added and the resulting suspension was homogenized with a needle (0.8 x 40mm). 50µl of Na acetate pH 5.0 was then added and the mixture was vortexed for 5 seconds. 500µl of acid phenol (Sigma 108-95-2) were added and the mixture vortexed for 10 seconds. 100µl of CIA were added and the tube was again vortexed. The tube was then incubated on ice for 20 minutes before being centrifuged at 4°C, 13000 rpm for 20 minutes in an Eppendorf Centrifuge 5402. The aqueous phase was removed into a fresh 1.5 ml Eppendorf and 500µl PCIA were added and the Eppendorf was vortexed again for 10 seconds. After incubation for 5 minutes on ice, the aqueous phase was collected as before. The aqueous phase was then transferred to a clean tube and 500µl of CIA were added. The tube was again incubated on ice for 5 minutes before the aqueous phase was collected as above. The RNA was then precipitated in 2 volumes of 100% ethanol at -20°C for at least two hours.

The RNA was collected by centrifugation at 4°C, 13000rpm for 20 minutes before being washed with 70% ethanol. The pellet was collected by centrifugation as before and air dried. The RNA was dissolved in 10µl of DEPC water and stored at -70°C prior to use.

2.14. cDNA Preparation by Reverse Transcriptase

To make cDNA, 2µg of RNA was transferred into a sterile Eppendorf with 1µl of RNasin (Promega 200 U/µl). The volume was then made up to 20µl with DEPC water and the tube incubated at 65°C for 5 minutes before being cooled on ice for a further 5 minutes. The following was then added:

1µl RNasin (Promega 200U/µl)
10µl 5x Reverse Transciptase buffer (Gibco BRL)
5µl 5mM dNTPs
0.5µl 0.1 M DTT (Gibco BRL)
1µl pd(N)6 (Pharmacia 0.5 µg/µl)*
1µl MMLV Reverse Transcriptase (Gibco BRL 200U/µl))

The reaction was then incubated at 37°C for 1 hour and then heat inactivated at 65°C for 10 minutes. 2 µl of cDNA was then used for each subsequent PCR reaction. * any alternative antisense primer can be substituted here.

2.15. Polymerase Chain Reaction (PCR)

The polymerase chain reaction (PCR) was performed throughout this project using many different DNA templates and primers. All reactions were carried out in 50µl volumes with 1 x PCR buffer (Promega) plus $MgCl_2$. Unless otherwise stated the Taq Polymerase used was Promega and 0.5-1µl was used per reaction. 10% of 2mM dNTPs were added with 250ng of each primer. The volume was then made up with the DNA template and sterile water. A drop of mineral oil (Sigma 8042-47-5) was laid over the reaction to prevent evaporation.

A standard program was used for all PCRs unless otherwise stated, with the annealing temperature being altered depending on the Tm of the primers used;

Hot start	94°C	5 minutes	
			1 cycle
Denaturation	94°C	1 minute	

Annealing	variable	1 minute	
Extension	72°C	2 minutes	
			35 cycles
Extension	72°C	10 minutes	
			1 cycle

PCR products were characterised by electrophoresis on an agarose gel containing 1μg/ml of ethidium bromide. The agarose percentage varied depending on the expected size of the PCR products.

Chapter 3. Physical mapping of the critical region at 13q14.3 deleted in B-cell CLL

3.1. Introduction

In order to begin a positional cloning project, a map of the critical region must be assembled. As discussed in detail in chapter 1, analysis of the deleted region at 13q14.3 by other groups had identified two markers, D13S319 and D13S25, as being more frequently deleted than the retinoblastoma gene (RB 1) (Brown *et al.*, 1993; Liu *et al.*, 1993; Liu *et al.*, 1995). Both markers were located telomeric of the RB 1 gene. This chapter describes how a map of the region at 13q14.3, spanning these two markers deleted in B-cell CLL, was constructed.

3.1.1. Preliminary data

Prior to my starting work on this project, some initial data had been collected within our lab. An ICI YAC genomic library (available from the Human Genome Mapping Project in Cambridge, UK) was screened by PCR using primers for both D13S25 and D13S319 (see below). 2 YAC clones containing the D13S319 marker and 3 YAC clones containing the D13S25 marker were identified (table 3.1). In addition, a series of CEPH (Centre Européen pour l'Etude du Polymorphisme Humain) clones were known to be located within the deleted region at 13q14.3 (Gyapay *et al.*, 1994). Five of these clones were initially obtained and screened by PCR with the primers for D13S25 and D13S319. YAC 933e9 (table 3.1) was shown to contain both markers and was chosen for further study. All the YAC clones described were shown by fluorescent in situ hybridisation to be non chimeric clones. The YACs were sized by PFGE (table 3.1 and figure 3.2).

Primers for amplification of microsatellite markers D13S25 and D13S319;

D13S25 F

5'- GATTCTAATCAGATAGACTG -3'

D13S25 R

5'- AAGCTTCTGTTTGTTCAAGA -3'

D13S319 F

5'- CGAGCTGGAGTCCATCGTAT -3'

D13S319 R

5'- CGTCGCTGCAGATCAAAGGA -3'

Name of YAC	Source	Marker identified	Approximate size
YAC 11AF3	ICI library	D13S319	250 kb
YAC 37AC6	ICI library	D13S319	250 kb
YAC 22FA4	ICI library	D13S25	500 kb
YAC 32EH6	ICI library	D13S25	350 kb
YAC 39FC12	ICI library	D13S25	300 kb
YAC 933e9	CEPH library	D13S25 and D13S319	1.5 Mb

Table 3.1: YAC clones isolated from the genomic region at 13q14.3

3.1.2. New mapping data from other groups

The construction of the physical map at 13q14.3 was refined throughout this project, in part due to the publication of two important papers that included detailed physical maps of the deleted region at 13q14.3. The first described several microsatellite markers from the region at 13q14.3 (Bullrich *et al.*, 1996). Primer sequences for amplification of these microsatellite markers were publicly available from Whitehead Data Base (http://www-genome.wi.mit.edu/) and the Genome Data Base (http://gdbwww.gdb.org/) web sites (see table 3.2).

The second publication described a contig of overlapping PAC clones covering the candidate regions of deletion at 13q14.3 (described in chapter 1) (Kalachikov *et al.*, 1997). Some of these PACs were available from the Human Genome Mapping Project (HGMP) at the Medical research Council (MRC) in Cambridge, UK. The authors also detailed many EST and STS markers mapping to the 13q14.3 region. Some of these were used in this project and are shown in table 3.3.

3.2. Methods

3.2.1. Probes

D13S25 - A probe for D13S25 was available as a Bluescript clone of a 2kb Hind III genomic fragment containing the D13S25 marker as previously described (Jabbar *et al.*, 1995).

D13S319 - A probe for D13S319 was available as a Bluescript clone containing a 4kb EcoR I repeat free genomic fragment obtained from a cosmid, Mgg 15, containing the D13S319 marker (kindly provided by Liu *et al.* (Liu *et al.*, 1995)).

pp1.6 - A cosmid library of YAC 22FA4 was constructed by Dr Panayiotidis. The cosmid clones were analysed for the presence of HTF islands by double restriction enzyme digestion using EcoR I, Hind III and rare cutters such as Sal I, BssH II and Not I. One cosmid was found to contain Sal I sites and a repeat free fragment (pp1.6) was identified from this cosmid, subcloned into a Bluescript vector and used as a probe for hybridisation to genomic clones.

3.2.2. PCR amplifications of microsatellite markers from 13q14.3

Name	Sequence 5' - 3'	Annealing temp.
D13S273 F	CTGNGGCAAAAACAACTCTT	56°C
D13S273 R	ATCTGTATGTCCTCCTTTCAATG	56°C
D13S272 F	ATACAGACTTCCCAGTGGCT	58°C
D13S272 R	AGCTATTAAAGTTCCCTGGATAAAT	58°C
D13S294 F	CCCAGTGAGCAGCCTCTAAA	56°C
D13S294 R	AACAGAAATCAGGCCAGTGT	56°C
206XF12 F	TAGGAATTCGGTATGACCCAAGGATTAC	60°C
206XF12 R	TAGAAGCTTGCTGCTCCAGGCTCATTA	60°C
AFMA301WB5 F	CACATCTTTGCTCTAATC	48°C
AFMA301WB5 R	TTCAAACTGAGTGACTGC	48°C

Table 3.2: Sequences of the primers and the annealing temperatures used to amplify microsatellite markers from 13q14.3.

Microsatellite markers were mapped to genomic clones from 13q14.3 by PCR. The primer sequences and the annealing temperatures used for each primer pair are shown in table 3.2. The DNA templates for these PCRs were the 13q14.3 YAC and PAC clones described.

3.2.3. PCR amplifications of STS and EST markers from 13q14.3

Dr Foroni mapped all of the STS and EST markers shown in table 3.3 by PCR. The annealing temperatures for each reaction are shown. The DNA templates for these PCRs were the 13q14.3 PAC and YAC clones described.

Sequence type	Name	Sequence	Annealing temp.
STS	72G6T7 F	TGATATTACGCATGGCACA	50°C
	72G6T7 R	TGAGTTTACAGTGAGCCA	50°C
STS	47G5T3 F	ATTTTCCTGGAGGAATGG	50°C
	47G5T3 R	TTGCTCTGGTACTTTGAC	50°C
STS	49G7T7 F	TCTCAGTAACAAGACCAC	50°C
	49GT7 R	ACCCTGTGATGTTCATAG	50°C
STS	138G4T3 F	CCCGGTCACAATGGTATG	52°C
	138G4T3 R	AGTACCTGCATCGCCTTTG	52°C
EST	cDNA-1H4 F	TGACAGTCACACGGCTTC	52°C
	cDNA-1H4 R	AATGGAGACTCAGGAAAGC	52°C
STS	16D4T3 F	TGTAGTCAAGTTTATGACCG	52°C
	16D4T3 R	GAACTAATCCAGACAAAGAG	52°C
STS	16D4T7 F	GTAGGAAGGCCAGGGCTC	56°C
	16D4T7 R	GCTCTGGCAGTGATCTGTG	56°C

Table 3.3: Primers used to amplify six STSs and one EST from 13q14.3.

3.2.4. Mapping of PAC clones

Of the nine PAC clones described by Kalachikov *et al.*, six were available from the HGMP *(Kalachikov et al., 1997)*. The PACs were initially grown on LB agar plates, supplemented with kanamycin (50μg/ml), at 37°C overnight. DNA for each PAC clone was then prepared from an overnight 50ml culture in LB media

plus kanamycin by the adapted plasmid DNA maxiprep method (section 2.1.4). A 5µg aliquot of each PAC clone DNA was digested in two separate 50µl digests with the restriction enzymes Hind III and EcoR I. The digested DNA was then electrophoresed on a large 0.8% agarose gel in 1 x AGB buffer. The DNA was visualised by ethidium bromide over a UV light transilluminator before being transferred to Hybond-nylon filters by Southern blotting. The filter was then hybridised to probes from 13q14.3 (section 3.2.1).

1µg of DNA from each PAC was used in a PCR reaction to amplify each of the microsatellite, STS and EST markers described (3.2.2). PAC clones were assumed to overlap if they were positive for the same markers.

3.2.5. Mapping of YAC clones

Mapping of YACs 22FA4, 32EH6, 39FC12 and 933e9 was achieved by a combination of PFGE (using agarose plugs of individual YACs), Southern blotting of digested high molecular weight DNA using restriction enzymes such as EcoR I and Hind III, and PCR amplification of the microsatellite and EST / STS markers shown in tables 3.2 and 3.3. Filters from both the PFGE and Southern blotting experiments were hybridised to probes for 13q14.3 as described in 3.2.1.

PCR amplifications were performed using YAC DNA embedded in agarose plugs. The agarose plugs were equilibrated in at least two changes of TE over 3 hours prior to use. The plugs were then melted at 65°C and 10µl of YAC DNA in molten agarose was used in a 100µl PCR reaction. As for the PACs, clones were assumed to overlap if they were positive for the same marker.

3.3. Results

A physical map of the region was constructed as shown in figure 3.1. The distances in the map were calculated from the known lengths of the clones used; the YAC clones were sized by PFGE (figure 3.2) and the sizes of the PAC clones were taken from the publication by Kalachikov *et al.* (Kalachikov *et al.*, 1997). A complete contig of the region was assembled in YAC clones. The PAC contig, however, was interrupted between PAC 246m14 and PAC 55o6.

The YACs positive for D13S25 and D13S319 by PCR were verified as positive by hybridisation of these probes to filters of YAC DNA PFGE blots (figure 3.2). Similar hybridisations to filters of PAC DNA showed PAC 269f22 to

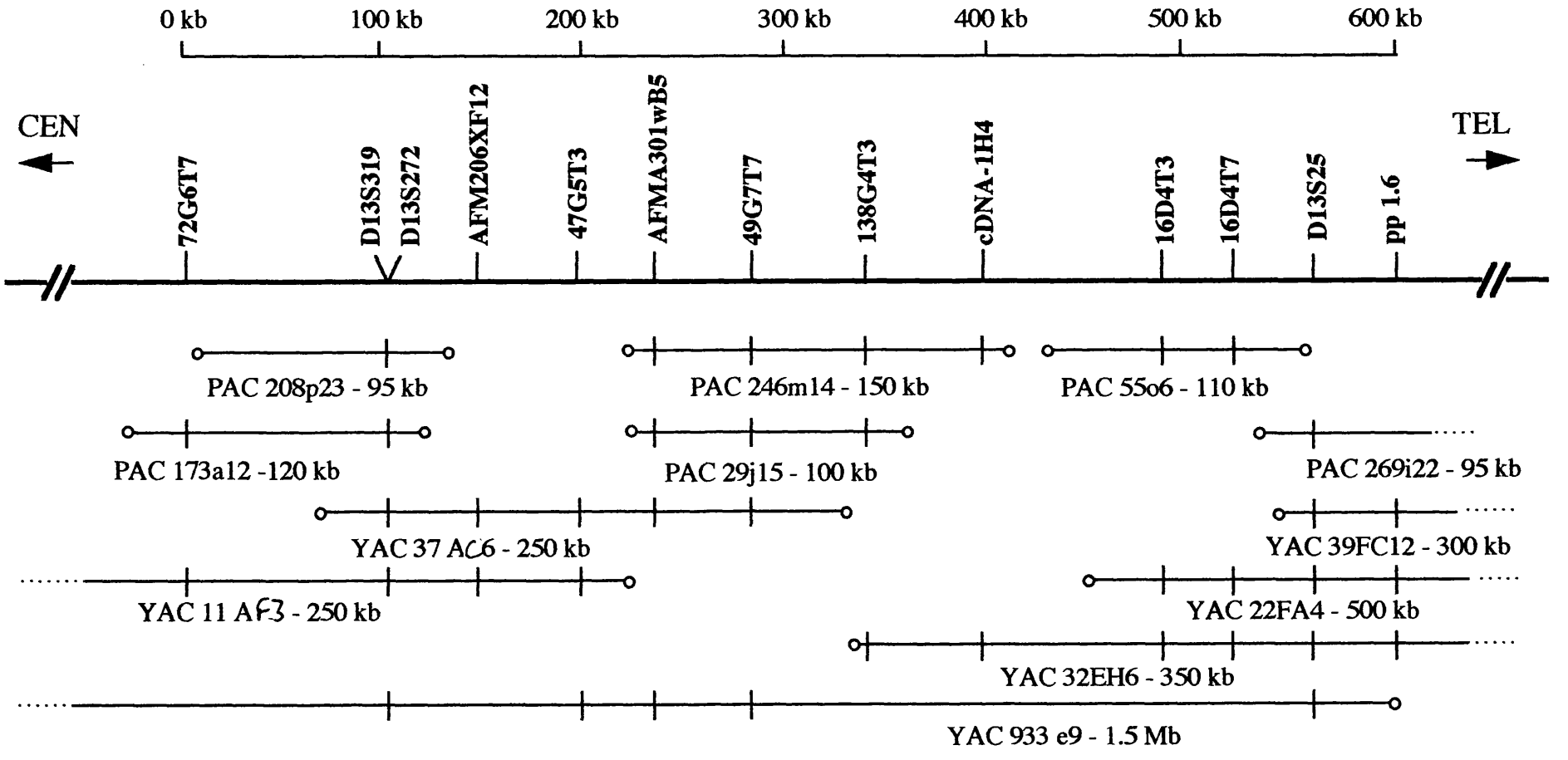
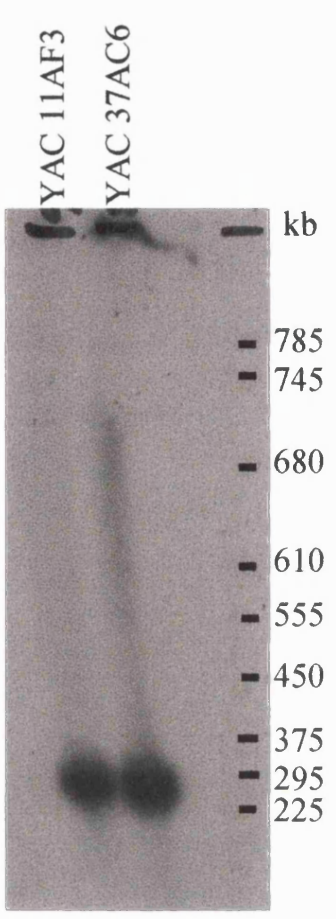


Figure 3.1: A physical map of the region of chromosome 13q14.3 known to be deleted in B-cell CLL. All YAC clones were identified by our group during this study. The scale (kb) was determined from the known sizes of the YAC and PAC clones used to produce the contig.



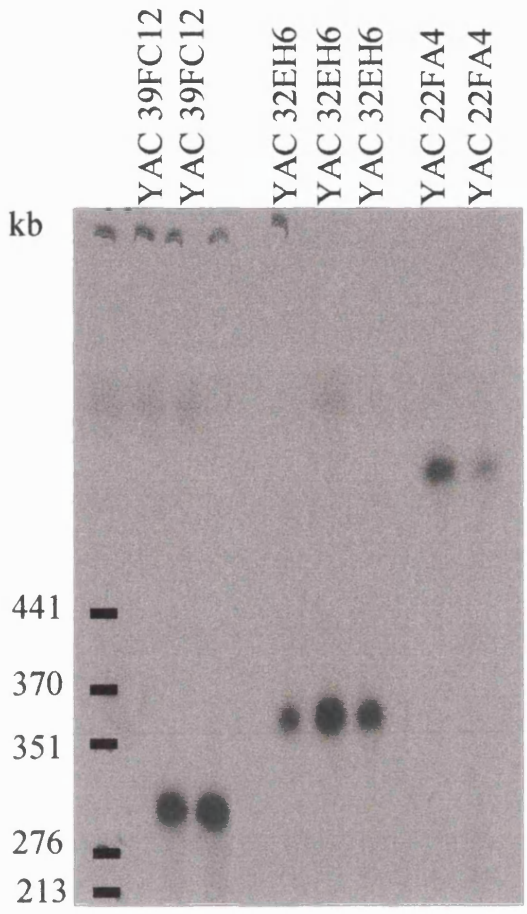


Figure 3.2A

Figure 3.2B

Figure 3.2A: PFGE of YACs 11AF3 and 37AC6 hybridised to the D13S319 probe in Church buffer at 65˚C overnight. The filter was washed at a final stringency of 3 x SSC/ 0.1% SDS at 65˚C and exposed to Fuji Medical autoradiography film overnight. Both YACs are approximately 250kb in size and are positive for the probe.

Figure 3.2B: PFGE of YACs 22FAF4, 32EH6 and 39FC12 hybridised to the D13S25 probe in Church buffer at 65˚C overnight. The filter was washed at a final stringency of 3 xSSC / 0.1% SDS at 65˚C and exposed to Fuji Medical autoradiography film for 12 hours. All 3 YACs are positive for the D13S25 probe. YAC 22FA9 is approximately 500kb. YACs 32EH6 and 39FC12 are 350kb and 300kb respectively.

be positive for D13S25 and PACs 208p23, and 173a12 to be positive for D13S319. YACs 22FA4, 32EH6 and 39FC12 were positive for the probe pp1.6. YACs 933e9, 11AF3 and 37AC6 were negative for this probe, suggesting that this STS lies telomeric of D13S25 (figure 3.1).

The results of the mapping of microsatellite, STS and EST markers can be seen in figure 3.1. The primers for D13S294 repeatedly failed to work despite systematically lowering the annealing temperature of the reaction to 50°C and titrating the reaction with varying concentrations of magnesium chloride, ranging from 1.5mM to 4.5mM. This marker was, therefore, not included in our map.

From this physical map it was possible to estimate the genomic distance between D13S319 and D13S25 as less than 600 kb. This was calculated by adding the known lengths of YAC 37AC6 and YAC 32EH6. As YAC 32EH6 contains the STS pp1.6 which is telomeric of D13S25, the real genomic distance between D13S319 and D13S25 is likely to be less than this estimate.

3.4. Conclusions

By using a combination of techniques, a comprehensive physical map of the region at 13q14.3, deleted in B-cell CLL, was constructed. This map included a contig of 6 YACs and 6 PACs giving complete coverage of the area. In only one region (between PACs 246m14 and 55o6) was there interruption of the PAC contig as described, and this region was consequently only covered by YAC clones. Thirteen unique markers were positioned within this map, greatly improving its accuracy.

From the assembly of this map the physical distance between D13S25 and D13S319 was estimated at less than 600kb. Bullrich *et al.* estimated the distance between 206XF12 and D13S25 as 550kb (Bullrich *et al.*, 1996). Our data are in keeping with this figure. Kalachikov *et al.* estimated the distance between D13S319 and D13S25 at 380kb (Kalachikov *et al.*, 1997). Due to the multiple layers of clones comprising the contig described in their publication, this value is likely to be more accurate. Unfortunately, 3 of the PAC clones described in that paper were unavailable to us and we were unable to map the genomic region at 13q14.3 in an uninterrupted PAC contig. As a consequence we could not verify their data.

Some of the distances in our map vary from those published by Kalachikov *et al.* This is due to YAC 37AC6 being sized at 250kb by PFGE (fig.

3.2A). Consequently, PACs 246m14 and 55o6 are closer together in our physical map than in the other published data. Kalachikov *et al.* estimated the distance between D13S319 and 301wB5 at 100kb. In our map this distance is 150kb. The distance between 206XF12 and cDNA-1H4 is also increased on our map. Despite the differences in calculated distances between markers on these physical maps, the order of the markers at 13q14.3 and the position of the clones remains constant between them.

As with any positional cloning project, the construction of a map is essential for the analysis of further work. This map was useful in establishing the accurate deletion boundaries in the analyses of B-cell CLL tumour samples (chapter 7). It was also essential in the mapping of candidate tumour suppressor gene clones, isolated from the region of deletion (chapters 4, 5 and 6). Consequently, it will be referred to on many occasions throughout this thesis.

Chapter 4. Identification of a candidate tumour suppressor gene cDNA from the deleted region at 13q14.3

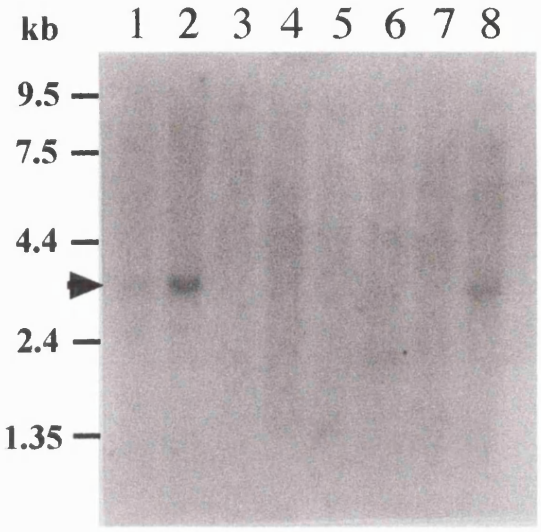
4.1. Introduction

Following the construction of a physical map of the deleted region at 13q14.3, the next aim was to isolate candidate tumour suppressor gene sequences from this region. Many different techniques were available to do this, as discussed in chapter 1. A combination of exon trapping and cDNA library screening was employed and a candidate tumour suppressor gene cDNA was identified.

4.1.1. Preliminary data

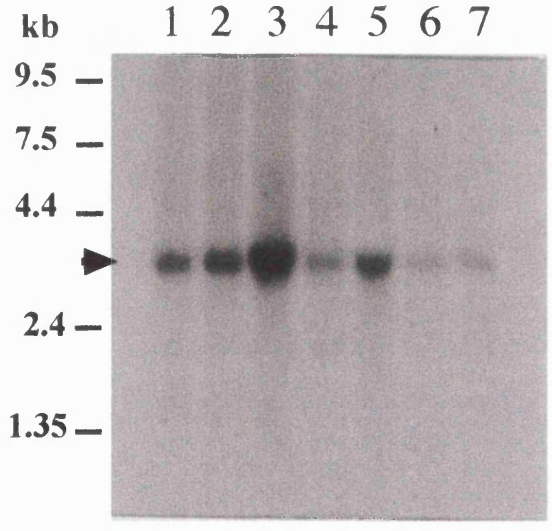
YAC 933e9 was 1.5Mb in length and spanned the genomic region at 13q14.3 between D13S25 and D13S319 (figure 3.1). Dr Panayiotidis from our laboratory constructed a λ Get phage library of YAC 933e9 and used λ Get exon trapping techniques, as described by Nehls *et al.* (Nehls *et al.*, 1994), to identify a 199bp putative exon, TA 6.35. When labelled and used as a probe, this sequence identified a 6kb EcoR I genomic fragment on Southern blot. When hybridised to filters containing DNA from B-cell CLL patients known to have homozygous deletion of D13S25, no signal could be detected with the TA 6.35 probe. This demonstrated that the TA 6.35 sequence was within the deleted genomic region in these patients. The same probe was then hybridised to a multi tissue Northern blot (Clontech). A 3.5kb mRNA message was identified in peripheral leucocytes, adult spleen and adult thymus. The TA 6.35 probe was then hybridised to a immune tissue Northern blot (Clontech) and the same message was identified in all represented tissues (figure 4.1).

This chapter describes a set of experiments that characterised and mapped this putative exon, TA 6.35. In an effort to identify the 3.5kb message seen on Northern blot analysis, a series of cDNA clones from the region at 13q14.3 were isolated and characterised. These are also described within this chapter.



1. Spleen
2. Thymus
3. Prostate
4. Testis
5. Ovary
6. Small intestine
7. Colon
8. PB leucocytes

Figure 4.1 A: Clontech multi tissue Northern blot hybridised to the probe TA 6.35. Final washing stringency 0.2 x SSC/ 0.2% SDS before exposure to Fuji Medical autoradiography film for 48 hours. A 3.5 kb message can be seen in lanes 1, 2 and 8.



1. Spleen
2. Lymph node
3. Thymus
4. Appendix
5. PB leucocytes
6. Bone marrow
7. Fetal liver

Figure 4.1 B: Immune tissue Northern hybridised to the probe TA 6.35. Final washing stringency 0.2 x SSC/0.2% SDS before exposure to Fuji Medical autoradiography film for 48 hours. A 3.5 kb message can be seen in all lanes.

4.2. Mapping of TA 6.35

In order to amplify the sequence of TA 6.35, sense and antisense primers were designed as shown below. Using a standard 35 cycle PCR reaction with an annealing temperature of 65°C, the 199bp sequence could be amplified from genomic DNA.

TA 6.35 F

5'- GGTCTGATGGCCGGGAGGACA -3'

TA 6.35 R

5'- CTGGTCGTGCAACAGTGTAC -3'

This PCR was repeated on genomic clones from 13q14.3. YACs 22FA4 and 32 EH6 were both positive for TA 6.35. All other YACs were negative. This result was then confirmed by hybridising the TA 6.35 probe to filters from PFGE blots of the 13q14 YACs. Only PAC 55o6 was positive for TA 6.35 by PCR. This was also confirmed by hybridisation to a filter of PAC DNA digested with both Hind III and EcoR I. TA 6.35 hybridised to a 6kb EcoR I fragment and a 5.5kb Hind III fragment with PAC 55o6.

4.3. Cloning and sequencing of the genomic 6kb EcoR I fragment with the identification of intron - exon borders of TA 6.35.

In order to clone the 6kb EcoR I fragment identified by TA 6.35 on a genomic Southern blot, a genomic library (the SHAA library) was screened with the TA 6.35 probe. The SHAA library is a human genomic bacteriophage library constructed in a BamH I λ 2001 vector, the insert DNA being partially digested with Sau 3AI (a kind gift from Dr T.H. Rabbitts, MRC Cambridge). The library was propagated in *Escherichia coli* Q358 cells and was screened as for a bacteriophage cDNA library.

Three clones positive for TA 6.35 were identified. The clones were then picked and DNA from each was prepared (section 2.1.6). Following digestion with EcoR I, electrophoresis and hybridisation to TA 6.35, all clones were shown to contain the expected 6kb EcoR I fragment. The 6kb insert was then purified from an agarose gel slice and subcloned into a Bluescript (KS+) vector. Using primers derived from TA 6.35, direct sequencing of the 6kb EcoR I fragment was

performed and the intron-exon boundaries of TA 6.35 were identified as shown below.

5' Intron	TA 6.35 exon	3' Intron
gtttatag	GAGGAGG.....CGACCAG	gttggcac

TA 6.35 was located from position 720 to 919 of the 6kb fragment. The intron borders were flanked with a 5' ag and a 3' gt, as expected for mammalian introns (Lewin, 1994), suggesting that the sequence of this putative exon was complete.

4.4. cDNA library screening with TA 6.35

The putative exon TA 6.35 had been shown to be contained within the genomic area of deletion at 13q14.3 and had been mapped to the telomeric part of this region, i.e. on PAC 55o6 (figure 3.1). It had also been demonstrated to identify a lymphoid specific 3.5kb mRNA by Northern blot analysis. In order to identify the full 3.5kb message, the probe for TA 6.35 was used to screen a cDNA library. Dr Panayiotidis screened 1 million colonies from a peripheral leucocyte cDNA library (Clontech), constructed in the phage vector rpD2, and obtained 3 positive clones. These were isolated and transfected into *Escherichia coli* BNN 132 cells. These cells contain the enzyme recombinase which recombines the transfected linear phage into a circular, multi copy plasmid conveying ampicillin resistance. The BNN 132 cells containing the plasmid were selected on a 2 x TY media supplemented with ampicillin and transferred to a large overnight 2 x TY culture with ampicillin. The plasmid DNA was then purified using the maxiprep method (section 2.1.3). To characterise the size of insert, the clones were digested with Xba I and BamH I and an aliquot was electrophoresed on a 1% agarose 0.5 x TBE ethidium bromide gel. For ease of manipulation and sequencing, the inserts were purified from the agarose gel and sub cloned into a Bluescript plasmid vector digested with Xba I and BamH I.

Three clones were initially identified. Clone 1 had an insert of 1.7kb. Clones 2 and 5 both had inserts of 1.2kb. Clones 1 and 2 were fully sequenced. The sequence of cDNA clone 1 is shown in figure 4.2. Clone 5 was not fully sequenced as it was shown to be identical to clone 2 by sequencing the ends of the insert. All three clones were polyadenylated. Clones 2 and 5 were contained

```
   1
   GGATCCGCGG TGAACTTAGA GGTTGTGGGG CCGAGGGGTC GTCTTATAGC
  51
   TACCAGCCCA CAGGCATTTA GTCTACGTTG GAGGTAAACA AATACGGGTC
 101
   CTGCTTAGGA GAAAAGAAAA ACGTCTTACA GCCAGTGTCT AAACTCCAAA
 151
   CAACGGAATG TATCAATGAG ACCTAGTATA TGGATACACG TGCATTTAAA
 201
   ACCGCCCTGC CGGCTTGTAG AGACTTTTGC CGTTCTCCAG CGCTTTACAG
 251
   GGGTTATCGC ACTTAAGCCT CGGAACAACT TTACCAGATG AGGACACCTG
 301
   AGGTTCAGAT TAAGAAATCT GCCCCAAAGT CTTAGAACTG GTATTCTCCA
 351
   CTGGTTGAAT GCAAATGGAA GCCAGAGGAA AAGGGAACCC TTGACACAGT
 401
   TTTCATATAG ATTGGCCTCC AGGGCAGACA GCAGGAGTGG GGTTTAATCT
 451
   GGACTGAATT GGAGGAAAGG AACTGGCACC AGAGTGGAGG CTGGAAGGAT
 501
   GGTGGAACTG ACACTGGGAC AGAACACAGA AAACAGAATC TAATGCTAGA
 551
   GGAGGTCTGA TGGCCGGGAG GACAGGAAGA AGGAACCCAC AATGAAGCAA
 601
   AAGTCGCGTT TAACTCTGTC CTCACCATTC CTCTCCTGGA CCTGTCCTAA
 651
   ACTGGAAGCT GGGGGAGGAT GGAGTGCTGG AATGCTGCAG AAGTCTCATG
 701
   GGAGGGCGCC TGGTGTTCAC CACATGTGTA CACTGTTGCA CGACCAGATG
 751
   GAAATTTGAA TGCTGGGGCT CAGTGGAACA CCTGAAGGAA GCTGGTGAGG
 801
   CAGATGTCCC GCTATGCCTT CCAAGTGTCA TGCAACCAGT TCATGCCTGT
 851
   CAGCATCTCT GCCGGTGGAC TCTTGAACCT TCAGAAAGGG CAGAAAGGAA
 901
   GTTTACTTCT GCCTCTACAA TGGATTTGCC TACTTTAGAC ATGTTTTCAA
 951
   GATTTATCCA CGTCATGGCA TATACCAGAA CTTCATTTCT TTTATGATTG
1001
   AATAATATTA CATGGTTTTA ATATACTATA TTTTGTTTAT CTGTTCATCA
```

```
1051
   ATTAATAAAT CTTTGGGTTG TTTTCAATTT TAGGCTATAA TAATTAACGC
1101
   TGTTAGGAAT ATTTACTTGT AAGTTTTTGT TATATGTTTT TAATTCTCCT
1151
   GGGTAGATAC TTAGGGATAA AATTGCTGAA TTAGATTGTA ACTTTATGTT
1201
   TAGAAACTGC CCAACTGTTT TCCAAAGTAA CTATCATTTT AATATTTCAC
1251
   CAGCATTGTA TGACAGTTCT GGTCTTTCCT TATTTAAACT AATACTTAAT
1301
   TATTGCCTTT TATTGTAGCT ATCCTAGTAG ATATGAAGTG TTATCTCATT
1351
   CATTATGGTT TTGATTTGCA TTCTCTAATG ACTAATAATA TCAAGCATCT
1401
   TTTCATGTGT ATACAAGATA TTTGTATATC TTCTTTAGAG AAATGTCTAC
1451
   TCAAATCCTG TGCCTATGTT TTAATTGGAT AATTTATGTT TTTATTGTTG
1501
   AGTTGTAATA GTTTTTAAAT ATATTCTGGA TCCTAGACCT TTAACAGCAC
1551
   ATGATTTGTA AATATTTCTC CCATTCTGTG GGTTGGCTTT TCACTTCTTT
1601
   ATATTATCTT TTGAAGCACA AGTTTTTTAT TTTGATAAAG TCCAGTTTAT
1651
   ATAAAAAAAA AAAAAAAAAA A
```

Figure 4.2: The sequence of cDNA clone 1. The sequence of TA 6.35 is highlighted in bold. cDNA clone 1 terminates in a poly A tail.

within clone 1 except for an additional 80 bp sequence immediately 3' of the TA 6.35 sequence. The schematic in figure 4.3 shows the relationship of these clones to each other and to the original probe used to identify them. By comparing these sequences with the genomic sequence, the extra 80 bp sequence contained in clones 2 and 5 was shown to be contiguous with the genomic DNA sequence. This 80bp sequence began with 'gt' and ended with 'ag' suggesting that this was a small intron between TA 6.35 and the following exon.

4.5. Genomic organisation of cDNA clone 1

4.5.1. The design of a 5' cDNA probe from cDNA clone 1

In order to complete the isolation and characterisation of the full length cDNA identified by TA 6.35, a 5' probe was designed. The 5' site used in the original rpD2 vector for the cloning of cDNA inserts was BamH I. This was preserved in cDNA clone 1 when it was subcloned into a Bluescript vector. A Pst I site at position 685 of cDNA clone 1 was identified from the sequence using the DNAStar software package. By digesting the Bluescript cDNA clone 1 with Pst I and BamH I, a 685 bp fragment was released containing the 5' sequence of the clone including TA 6.35 (see figure 4.4). This probe was called the 700bp BamH I Pst I probe.

4.5.2 Hybridisation of cDNA clone 1 probes to genomic Southern blots

The full length insert of clone 1 was purified from an agarose gel slice and labelled by random priming for hybridisation to a genomic Southern blot. Due to the repetitive nature of the sequence this probe, competition with placental DNA was used. The hybridisation revealed 6 bands on an EcoR I digested DNA Southern blot and 5 bands on a Hind III digested DNA Southern blot (figure 4.5).. No EcoR I or Hind III sites were demonstrated within the sequence of cDNA clone 1, as shown by analysis with DNAStar software. Therefore, these bands represented at least 7 individual exons within cDNA clone 1, as the exon proceeding TA 6.35 was already known to be on the same EcoR I fragment as TA 6.35.

When cDNA clone 1 was hybridised to DNA from patients with known homozygous deletion of 13q14.3, no signal was detected.

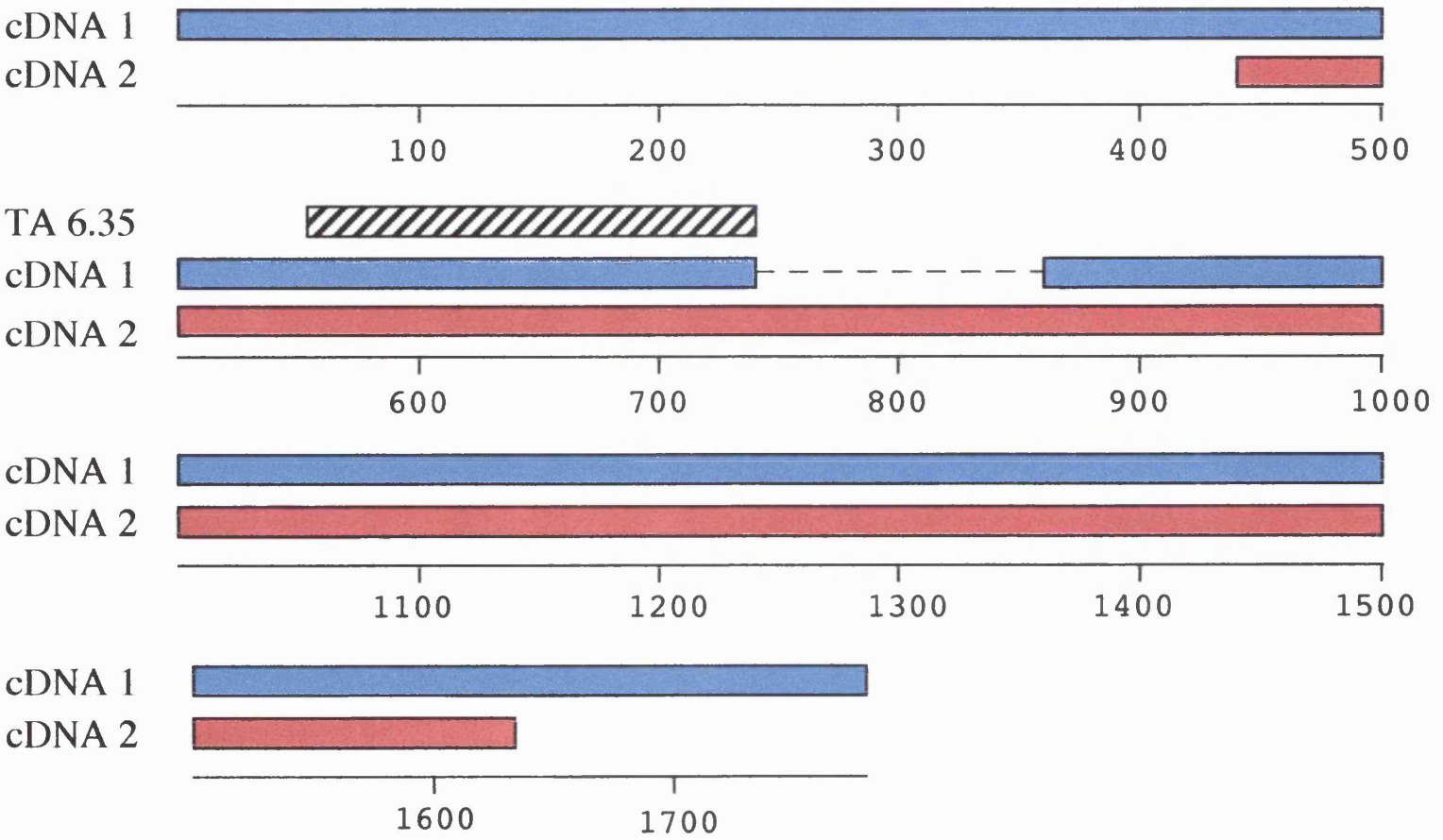
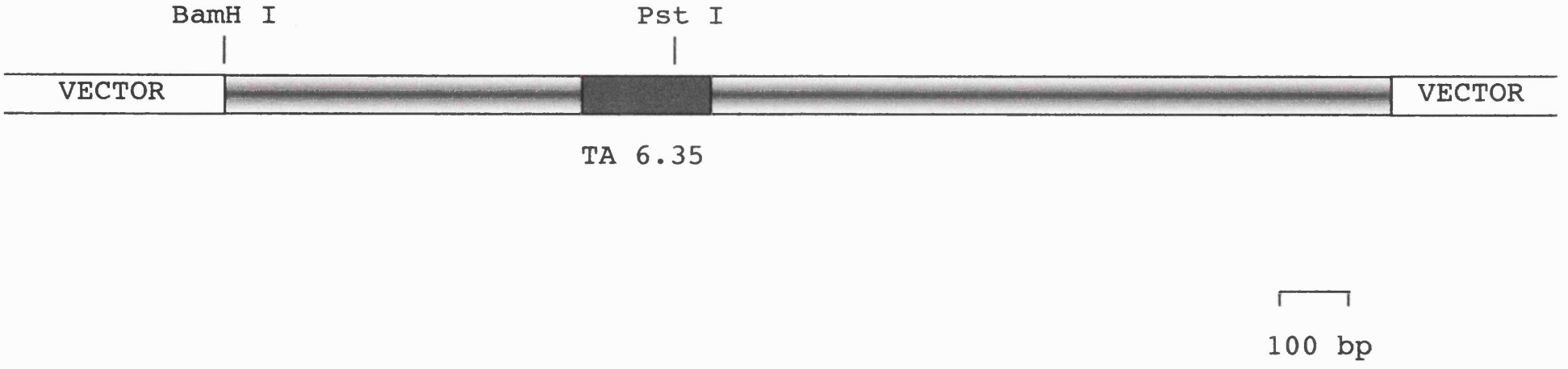


Figure 4.3: Comparison between cDNA clones 1 and 2 and the original putative exon, TA 6.35. Both cDNA clones were isolated from the same peripheral leucocyte cDNA library.

Figure 4.4: Showing cDNA clone 1 with the Pst I and BamH I sites used for excising the 700bp probe. The Pst I site was located at position 685. This probe included 137 base pairs of the original putative exon, TA 6.35



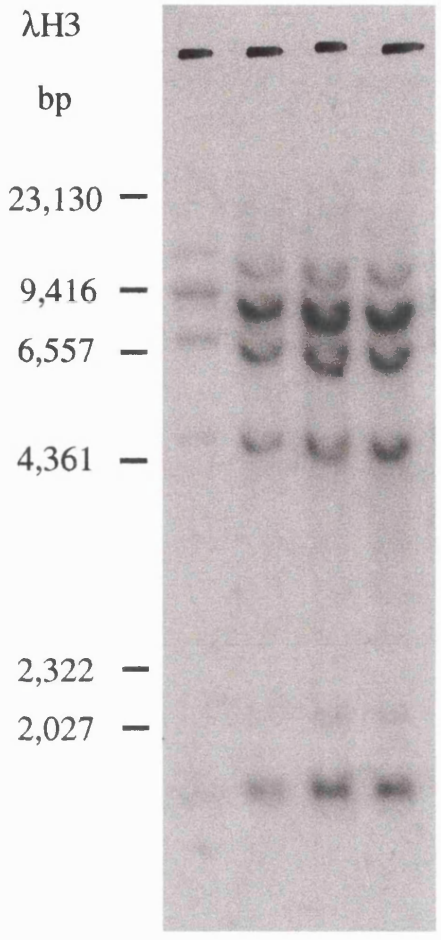


Figure 4.5A

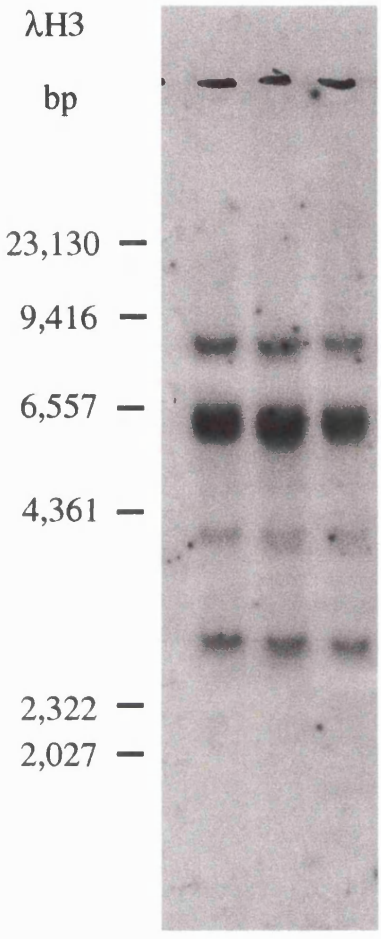


Figure 4.5B

Figure 4.5A: Southern blot of 10µg of normal genomic DNA digested with EcoR I and hybridised to the human probe, cDNA clone 1 in Cambridge buffer. The filter was washed at a final stringency of 0.5 x SSC/0.1% SDS, 65°C and exposed to Fuji Medical autoradiography film for 7 days. 6 bands are visible at 1.5 kb, 4.8 kb, 6 kb, 7 kb, 8.5 kb and 12 kb.The 7 kb and 8.5 kb appear as a doublet.

Figure 4.5B: Southern blot of 10µg of normal genomic DNA digested with Hind III and hybridised to the human probe cDNA clone 1 in Cambridge buffer. The filter was washed at a final stringency of 0.5 x SSC/0.1% SDS, 65°C and exposed to Fuji Medical autoradiography film for 7 days. 5 bands are visible at 2.5 kb, 3.8 kb, 5.5 kb, 6 kb (appear as a doublet) and 8 kb.

The 700bp BamH I/Pst I probe (4.5.1) was labelled and hybridised to a genomic blot. With an EcoR I digested DNA blot, the probe hybridised to 5 bands including the 6kb band identified by the original TA 6.35. These bands corresponded to those identified by cDNA clone 1 except for the 4.8kb band (figure 4.5). This suggested that a minimum of 4 exons lay upstream of TA 6.35. The additional band of 4.8kb seen when an EcoR I genomic blot was hybridised to cDNA clone 1 was presumed to contain the sequence 3' of the 700bp probe (figure 4.4).

4.5.3. Hybridisation of cDNA clone 1 probes to YAC Southern blots

When cDNA clone 1 was hybridised to PFGE blots of YACs from 13q14, positive signals could be seen with YACs 22FA4, 39FC12, 32EH6 and 11AF3. There was no signal with YAC 37AC6 (see figure 3.1). This was evidence that at least one exon of the transcript was contained within the centromeric YAC 11AF3. Therefore, cDNA clone 1 spanned from the genomic region centromeric of D13S319 to a region 400kb telomeric of this, close to D13S25 (figure 3.1). When the 700 bp BamH I / Pst I probe was hybridised to the Southern blot of YAC 11AF3 and 37AC6, positive signals were seen for YAC 11AF3 but no signals were seen for YAC 37AC6. This suggested that the 5' exons of the clone were the most centromeric in the genomic organisation of the transcript.

These data demonstrated cDNA clone 1 to have exons lying in the minimally deleted areas described by both Kalachikov *et al.* (Kalachikov *et al.*, 1997) and Bullrich *et al.* (Bullrich *et al.*, 1996)(figure 1.9).

4.5.4. Hybridisation of cDNA clone 1 probes to PAC Southern blots

When the 700 bp BamH I / Pst I fragment was labelled and hybridised to a Southern blot of DNA from the 13q14 PAC clones digested with EcoR I and Hind III, a pattern of bands could be seen. The same two bands were seen with PACs 173a12 and 208p23, at 8.5kb and 12kb with an EcoR I digest and 5.5kb and 3.5kb with a Hind III digest. There was no signal for PAC 269i22 or PAC 29j15. Two bands were seen with PAC 246m14, at 1.5kb and 3.8kb with a Hind III digest and 1.5kb and 7kb with an EcoR I digest. The expected 6kb band for TA 6.35 was seen on PAC 55o6 (figure 4.6). Note that no 1.5kb band can be seen with the Hind III digested genomic Southern blot (figure 4.5) when hybridised to the same probe.

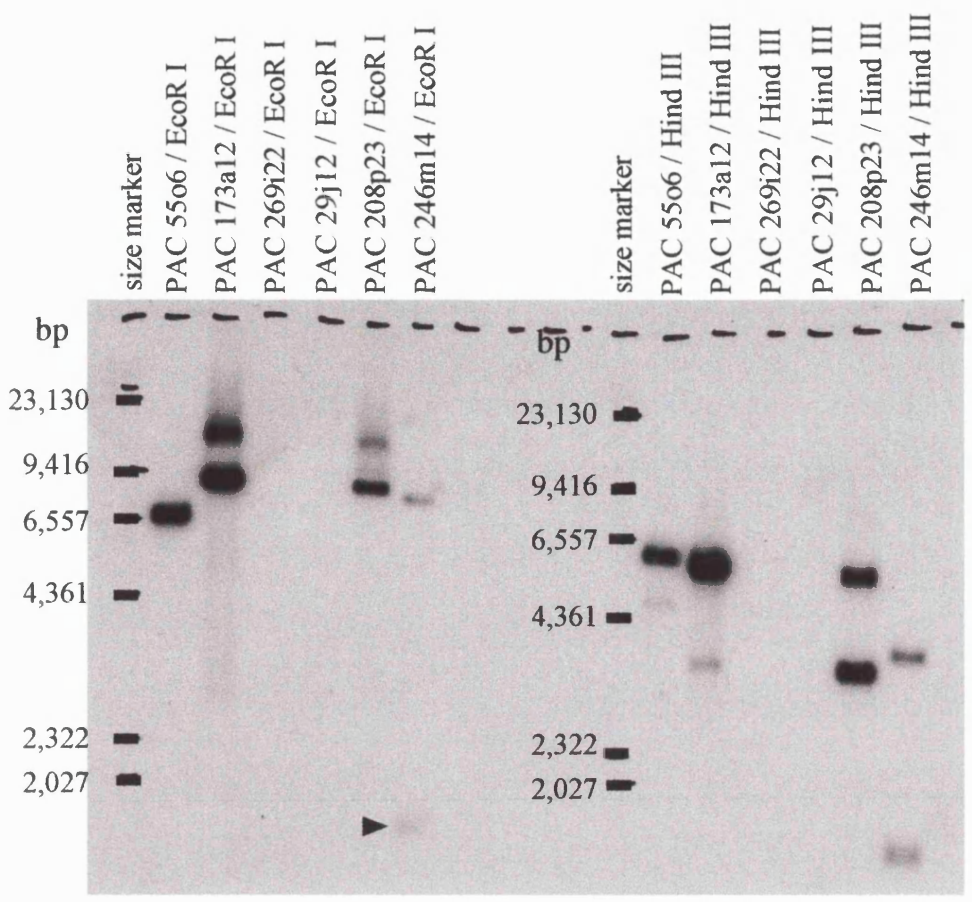


Figure 4.6: A Southern blot of 13q14.3 PAC clones. 1µg of DNA from each PAC clone was digested with EcoR I or Hind III as shown. The gel was blotted and hybridised to the human 700bp BamH I/Pst I probe in Church buffer and was washed at a final stringency of 0.2 x SSC/0.1% SDS. The filter was exposed to Fuji Medical autoradiography film at -20°C overnight. See text for a full description of the bands identifiied.The arrow marks a faint 1.5kb band identified in PAC 246m14 digested with EcoR I.

Further investigation showed the 2.5kb band seen on the genomic Southern to be correct (exon 2, discussed in chapter 6). It is not clear why PAC 246m14 consistently gave a 1.5kb signal instead of a 2.5kb signal when hybridised to cDNA clone 1, but it is likely that the PAC contains a polymorphism. The hybridisation for the EcoR I digested PAC and genomic Southern blots were consistent.

The lack of any signal on PACs 269i22 and 29j15 was in keeping with the negative hybridisation seen with YAC 37AC6 already discussed.

Figure 4.7 illustrates the genomic organisation of cDNA clone 1 according to these data from the mapping experiments described. Each exon of cDNA clone 1 located on the physical map represents at least one potential exon.

4.5.5 Hybridisation of cDNA clone 1 probes to cosmid Southern blots

A cosmid library of YAC 22FA4 constructed in the vector Supercos (Stratagene) from a partial Sau 3AI digest of the YAC DNA was available (previously made by Dr Panayiotidis). The DNA from all the cosmids had been immobilised on nylon filters by Southern blotting. These filters were hybridised to TA 6.35 and 4 cosmids positive for this putative exon were identified. The clones were grown from the original glycerol stocks and DNA was made using the plasmid maxiprep method. DNA from each cosmid was then digested with EcoR I and Hind III, electrophoresed and blotted onto filters. The filters were hybridised to cDNA clone 1. Two bands were identified in all the cosmids with an EcoR I digest; the 6kb band that was expected for TA 6.35, and a second 4.8kb band not seen when the 700bp BamH I / Pst I probe was hybridised to the PACs as previously described (figure 4.8). This band was attributed to represent the most 3' exon/s of cDNA clone 1. The inserts of the cosmids clones were known to be approximately 40kb. Therefore, the last 3' exon/s of cDNA clone 1 were presumed to be within 40kb of TA 6.35 (figure 4.7).

4.5.6. Design of primers along the length of cDNA clone 1

In order to assign the exon-exon boundaries within this clone, forward and reverse oligonucleotide primers were designed spanning the full sequence of cDNA clone 1. The sequence of these primers can be seen in table 4.1 and the positions of these primers in relation to cDNA clone 1 are shown in figure 4.9. The

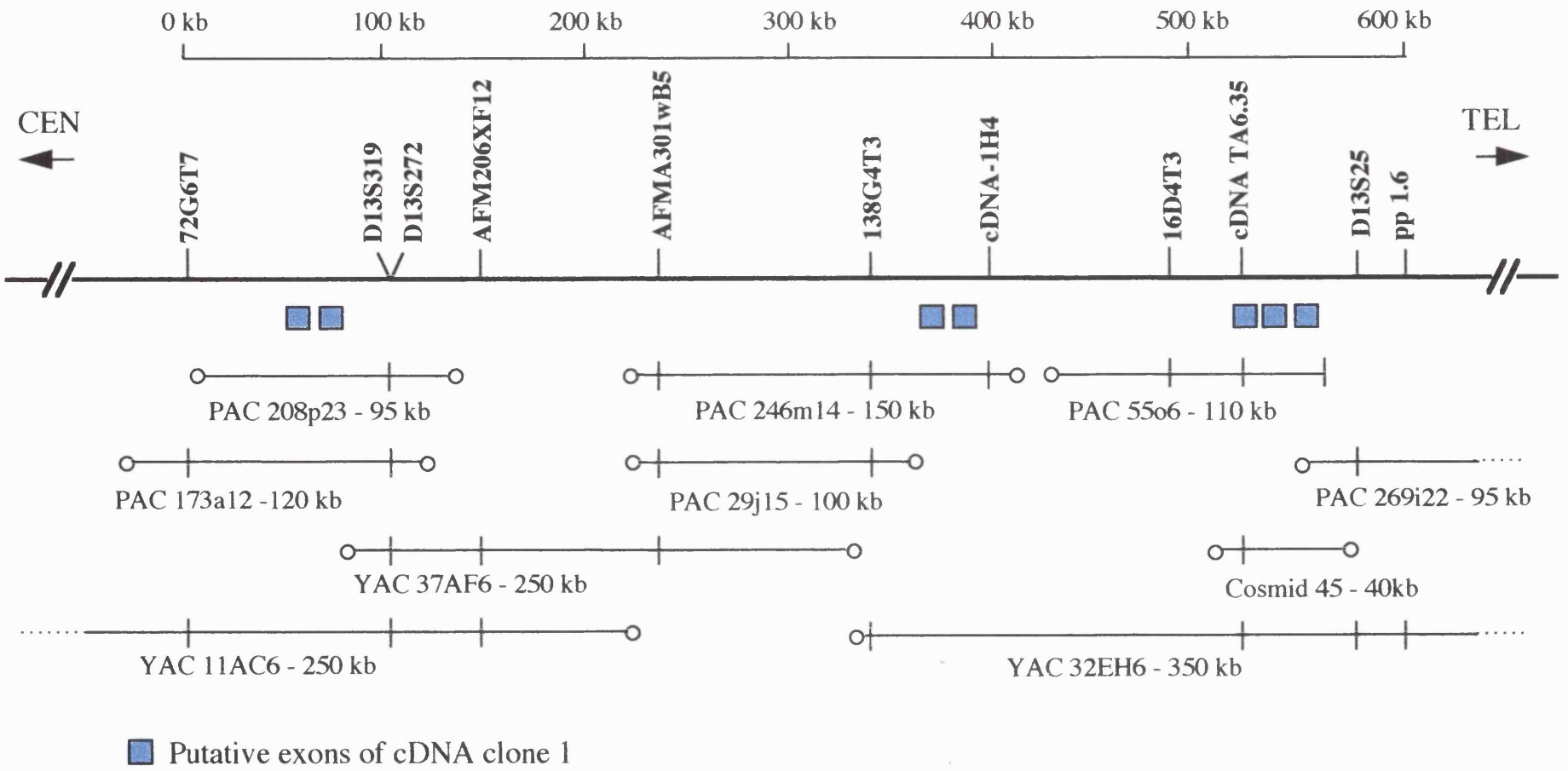


Figure 4.7: Physical positions of the putative exons of cDNA clone 1 at 13q14.3, as shown by hybridisation and PCR experiments. Cosmid 45 is included to highlight some of the additional distances described in the text.

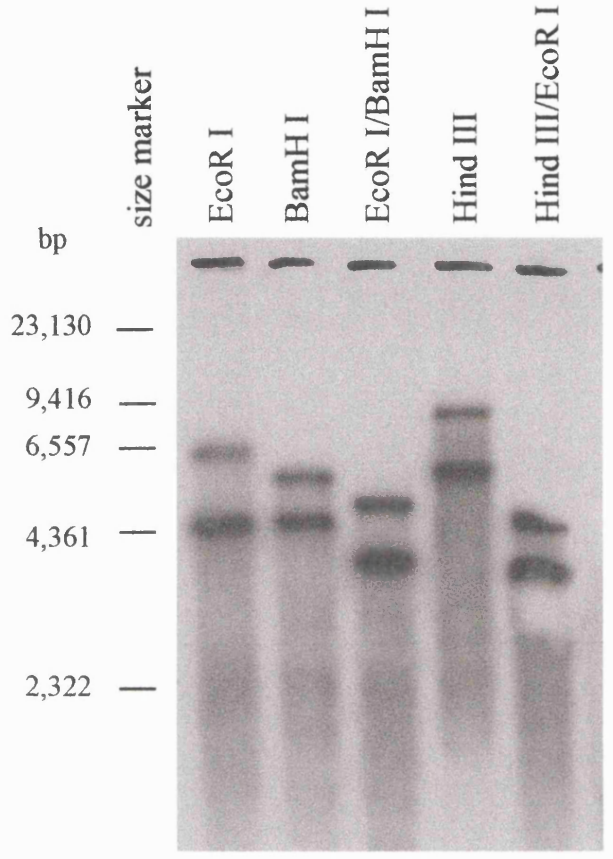


Figure 4.8: A Southern blot of a cosmid positive for TA 6.35. 3 µg of cosmid DNA was digested with the restriction enzymes as shown for each lane. The gel was blotted and hybridised to the human probe cDNA clone 1 in Cambridge buffer at 65°C. The final wash was at a stringency of 3 x SSC/ 0.1% SDS. The blot was exposed to Fuji Medical autoradiography film at -20°C overnight. Two bands can be seen for each digestion.

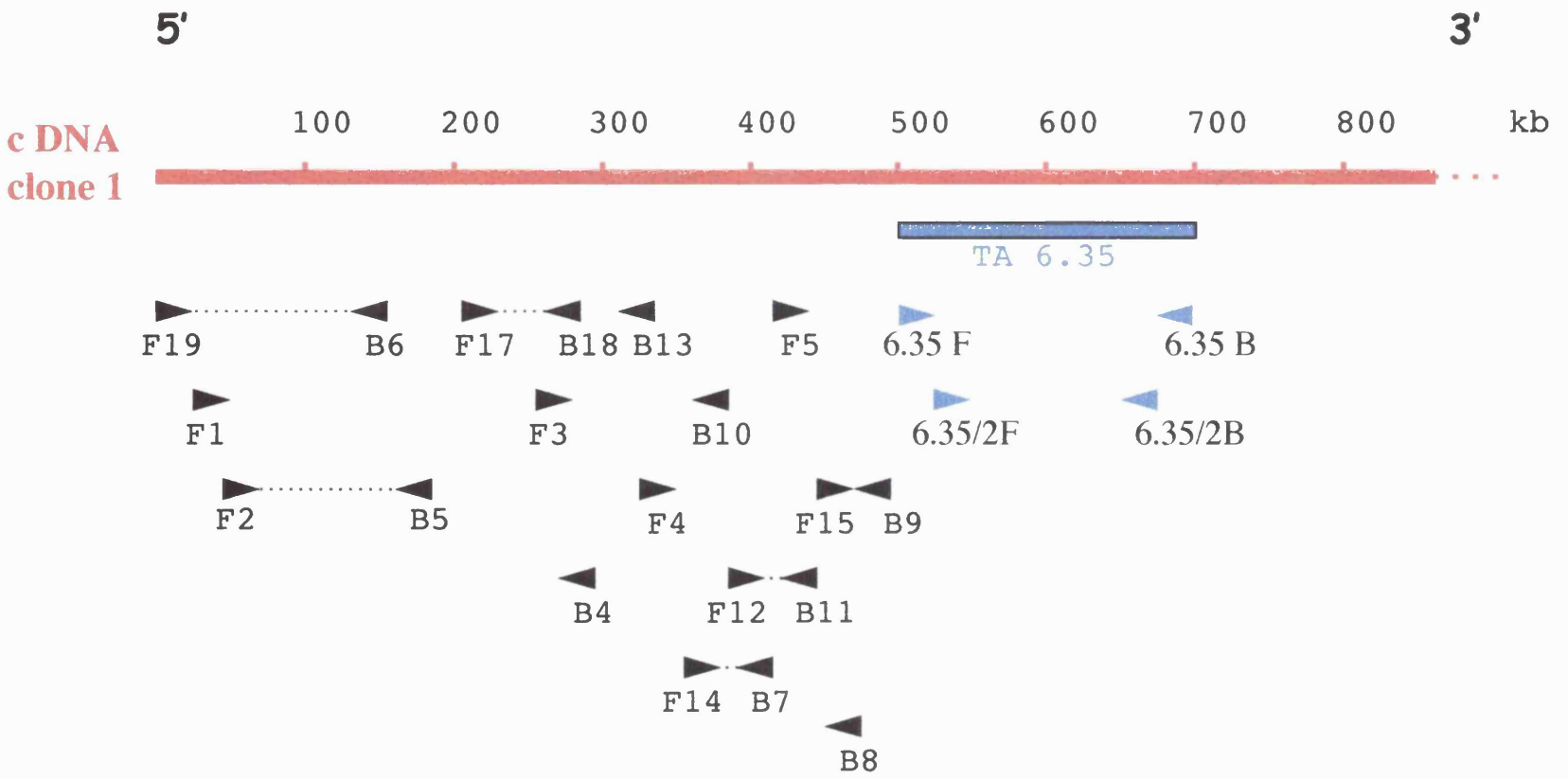


Figure 4.9: Positions of the primers described in table 4.2, spanning the 5' part of cDNA clone 1. The direction of the arrow head depicts whether the primer is a sense or antisense sequence. The dotted black lines show the pairs of primers that amplified the expected product from genomic DNA. cDNA clone 1 is shown in red and the position of the exon TA 6.35 and it's primers are shown in blue.

PRIMER	SEQUENCE
F1	GTCGTCTTATAGCTACCAGCCCAC
F2	GCCCACAGGCATTTAGTCTACGTTG
F3	CGCACTTAAGCCTCGGAACAAC
F4	CTGCCCCAAAGTCTTAGAACTGG
B4	TAAAGTTGTTCCGTGGCTTAAGTGCG
F5	CAGGGCAGACAGCAGGAGTGG
B5	GCACGTGTATCCATATACTAGG
B6	CTCATTGATACATTCCGTTGTTTG
B7	GGGTTCCCTTTTCCTCTGGCT
B8	CCTGTTCCTTTCCTCCAATTCAGTC
B9	GATTCTGTTTTCTGTGTTCTGTCCC
B10	CCAGTTCTAAGACTTTGGGGCAG
B11	GGAGGCCAATCTATATGAAAACTG
F12	AGCCAGAGGAAAAGGGAACCC
B13	TCTTAATCTGAACCTCAGGTGTC
F14	CTTAGAACTGGTATTCTCCAC
F15	GACTGAATTGGAGGAAAGGAACTGG
B16	ACTTTGGGGCAGATTTCTTAATG
F17	CGGCTTGTAGAGACTTTTGCCG
B18	AGGCTTAAGTGCGATAACCC
F19	GCGGTGAACTTAGAGGTTGTG
TA 6.35 F	GGTCTGATGGCCGGGAGGACA
TA 6.35 2F	TAGGGATTCGTCCTCACCATTCCTCTCCTGGAGCTG
TA 6.35 B	CTGGTCGTGCAACAGTGTAC
TA 6.35 2B	TAGCTCGAGAGCTTCCAGTTTAGGACAGGTCCA

Table 4.1: Showing the sequence of the primers designed from cDNA clone 1

following primer pairs amplified products of the expected size from genomic DNA; F2 and B5, F17 and B18, F14 and B7 and F15 and B9. These data were in keeping with the hybridisation data indicating that there were at least 4 exons upstream of TA 6.35. Using these primer combinations, the PCRs were repeated with 1μg of DNA from each of the PAC clones from 13q14.3. Pairs F2 - B5 and F17 - B18 gave the expected sized products with DNA from PACs 173a12 and 208p23. Primers F14 - B7 and F15 - B9 gave the expected PCR products with DNA from PAC 246m14.

These data were in keeping with the hybridisation data already described and shown in figure 4.7.

4.6. Expression of cDNA clone 1

cDNA clone 1 had been shown to be a candidate tumour suppressor gene transcript from the region at 13q14.3 which was deleted in B-cell CLL patients with known 13q deletions. The original putative exon, TA 6.35, had identified a 3.5kb lymphoid specific message on Northern blot analysis. To determine whether the transcript from clone 1 was expressed by normal and malignant B lymphocytes, further Northern blot analysis and RT PCR analysis was performed.

4.6.1. Northern blot analysis

A probe was made from the purified PCR product obtained when cDNA clone 1 was amplified with primers F14 and B9 (figure 4.9). When this probe was tested on a genomic Southern blot digested with EcoR I, two clear bands were seen at 1.5kb and 7kb. This probe was then hybridised to a Northern blot of mRNA from immune tissues (Clontech). The manufacturers instructions were followed and the blot was hybridised in Express Hyb^{TM} solution (Clontech) for 1 hour at 65°C. The final washing stringency was 0.1 x SSC / 0.1% SDS at 65°C. The blot was exposed to autoradiography film (Fuji Medical). However, no signal was seen from the filter, even after 3 weeks exposure.

4.6.2. Amplification from cDNA using cDNA clone 1 primers

The following combinations of primers spanning across postulated exon-exon boundaries of cDNA clone 1 were used in RT PCR reactions with cDNA from normal mononuclear cells and from thymus mRNA (Clontech); F2-B7, F14-B9, F15-6.35 B and F12-6.35 B (figure 4.9). These reactions were positively controlled using cDNA clone 1 as a template. A standard 35 cycle PCR protocol was used.

Only one combination of primers amplified a product; primers F2 (position 56-80) and B7 (position 390-369) amplified 2 bands, with an annealing temperature of 60°C, from most cDNAs tested (figure 4.10). The largest band was approximately 300bp as expected from the known sequence. The second band was approximately 50bp smaller. When this reaction was repeated on cDNAs from B-cell CLL samples the same result was obtained in most patients. However, in two patients only one fragment was detected by PCR (figure 4.10). In patient 53 only the smaller band was seen, whilst in patient 51 only the larger band was detected by electrophoresis (see chapter 7 for data on CLL patients). Primers F2 and B7 were designed containing 5' restriction sites for EcoR I and Hind III respectively. The PCR was then repeated in these 2 patients with the new primers and the products were digested and cloned into a Bluescript (KS+) plasmid vector. The cloned inserts were manually sequenced. The sequence obtained from the PCR product in patient 51 was identical to the sequence of cDNA clone 1. The sequence for patient 53 was missing base pairs 288-340 of the original cDNA clone 1 sequence. The sequence otherwise matched the known sequence of cDNA clone 1.

This result yielded two pieces of information about the nature of cDNA clone 1. Firstly, despite the negative Northern blot data, expression of the 5' part of this transcript could be shown in both normal lymphocytes and malignant B-cell CLL lymphocytes by RT PCR. Secondly, a splice variant appeared to exist with both versions being present in lymphocyte cDNA. The sequence data from patient 53 suggested that there was an exon-exon boundary at position 288 of cDNA clone 1 and a second exon-exon boundary at position 340. The difference between the two bands seen by electrophoresis of the PCR product in most samples appeared to be the result of alternative splicing of a small 53bp exon.

No amplification between the 5' exons of cDNA clone 1 and the TA 6.35 sequence was successful.

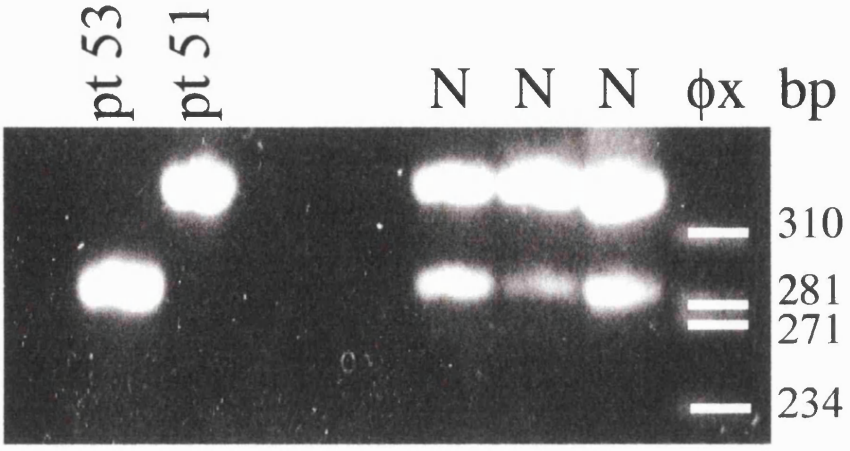


Figure 4.10A: The products of a nested PCR reaction on cDNA from B lymphocytes, using F2 and B7 primers.The PCR products were electrophoresed on a 2% agarose 0.5 x TBE ethidium bromide gel at 100 volts. Two bands can be seen in the normal controls whilst only one band is seen in both of the patient samples shown.

```
56
  GCCCACAGGC ATTTAGTCTA CGTTGGAGGT AAACAAATAC GGGTCCTGCT
106
  TAGGAGAAAA GAAAAACGTC TTACAGCCAG TGTCTAAACT CCAAACAACG
156
  GAATGTATCA ATGAGACCTA GTATATGGAT ACACGTGCAT TTAAAACCGC
206
  CCTGCCGGCT TGTAGAGACT TTTGCCGTTC TCCAGCGCTT TACAGGGGTT
256
  ATCGCACTTA AGCCTCGGAA CAACTTTACC AGATGAGGAC ACCTGAGGTT
306
  CAGATTAAGA AATCTGCCCC AAAGTCTTAG AACTGGTATT CTCCACTGGT
356
  TGAATGCAAA TGGAAGCCAG AGGAAAAGGG AACCC
```

Figure 4.10B: Showing the sequence of the larger PCR product from patient 53. The forward and reverse primers are boxed. The sequence in bold type is the 53 base pair sequence that was 'missing' when the smaller band from patient 51 was sequenced. The numbers refer to the position of the sequence relative to cDNA clone 1.

4.7. Rescreening the cDNA library with the BamH I / Pst I 700 bp probe

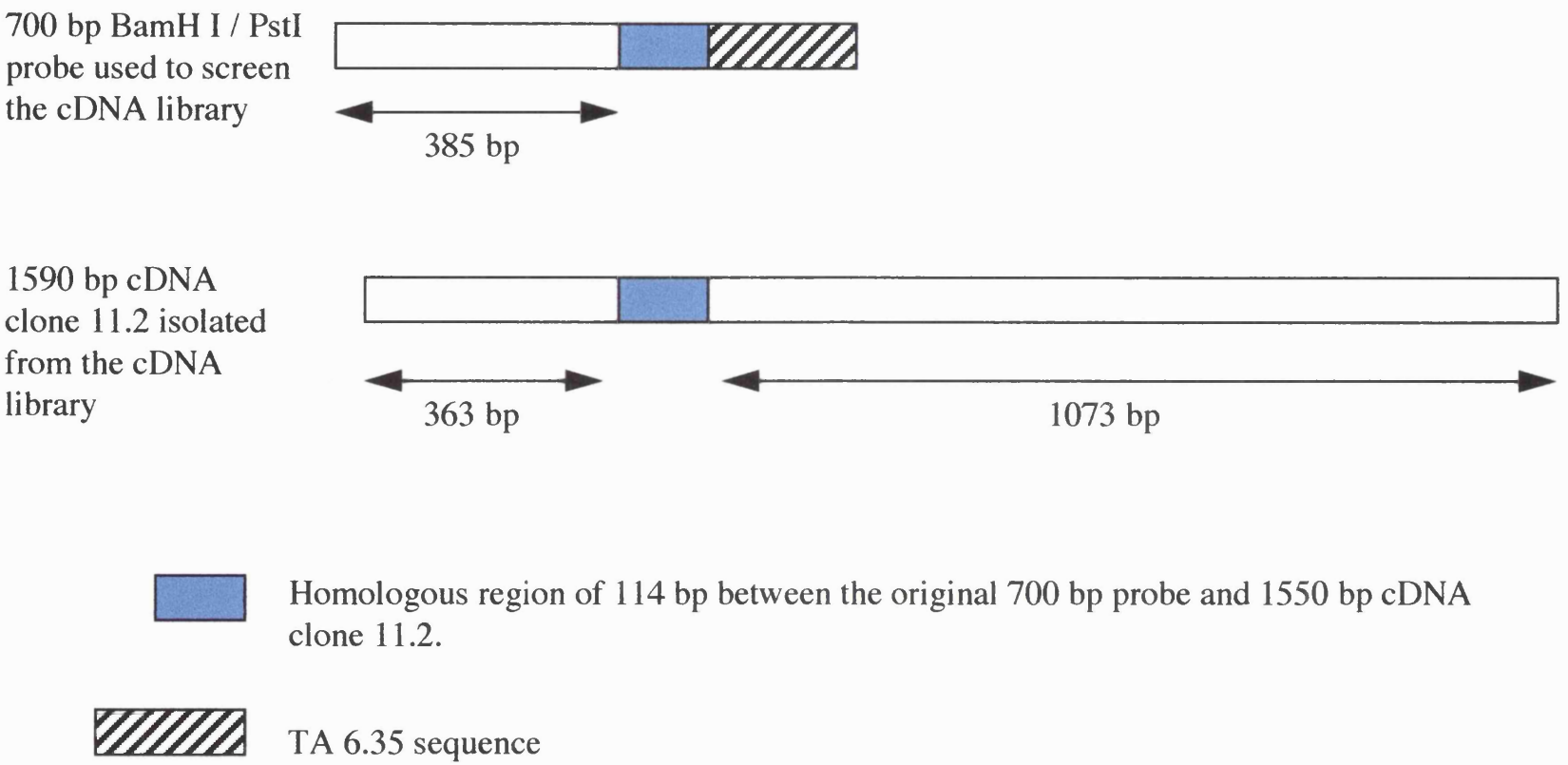
cDNA clone 1 was 1.7kb in length, consisting of multiple exons spliced together and terminating in a poly A tail. The original putative exon, TA 6.35, identified a 3.5kb mRNA on Northern blot as already shown. Assuming that cDNA clone 1 was part of this 3.5kb message, it was postulated that the rest of the message (1.8kb) lay 5' of this sequence, as cDNA clone 1 was poly adenylated. The 700bp BamH I / Pst I probe designed from the 5' part of cDNA clone 1 (figure 4.4) was an ideal probe to identify novel cDNA clones containing the upstream sequence of this message. The peripheral leucocyte cDNA library filters (Clontech) previously described were rescreened with this probe. Three new positive clones were identified by hybridisation to this probe.

The three clones were named cDNA 11.2, cDNA 14.2 and cDNA 19.3. When analysed by electrophoresis, the inserts of both 14.2 and 19.3 looked identical at approximately 900bp. The insert of clone 11.2 was much larger at approximately 1600 bp. The inserts of all three were purified, subcloned into a Bluescript (KS+) vector and the DNA was prepared by the plasmid maxiprep method (section 2.1.3). The ends of the 2 smaller clones were sequenced manually and the two clones were shown to be identical. Their most 5' sequence was identical to the 5' sequence of clone 1 and the 900bp long sequence terminated in a poly A tail. Therefore, sequence from these two shorter clones failed to yield any new sequence extending 5' to clone 1.

4.8. The characterisation of cDNA 11.2

DNA from cDNA clone 11.2 was sonicated and subcloned into a Sma I M13mp18 vector. The clones containing DNA from the insert were then identified by hybridisation to the original purified cDNA insert. Positive clones were manually sequenced. This process was continued until the full 1550bp sequence was obtained as a contig of overlapping sequences which had been sequenced in both directions. Only 114bp of the sequence was homologous to the original 700bp probe used to isolate this clone. This was the 114bp directly upstream of TA 6.35. This 114bp sequence contained the primers F15 and B9 which had been shown to amplify an expected sized product with genomic DNA (figure 4.9). It was postulated that this 114bp sequence, known to lie on PAC

Figure 4.11: cDNA clone 11.2 and its homology to the original 700bp BamH I / Pst I probe, highlighted in blue.



```
   1
   TGCCGCCTTG CAGTTTGATC TCGGACTGCT GTGCTAGCAA TCAGCGAGAC
  51
   TCCGTGGGCG TAGCACCCTC CGAGCCAGGT GCGGGATATA ATCTGGTGGT
 101
   GCACCGTTTT TTAAGCTGGT CCGAAAAGCG CAGTATTCGG GTGGGGAGTG
 151
   ACCCGATTTT CCAGCTGCTG CTAGGCGTTC TCCTTCAAAC ACACCTGGAT
 201
   TGCTGTGAGG ACACTGCCCC ACAGGCGGAA CACAATGTGT TTCTCAGAGC
 251
   TGCGGACTGT GGGCTGTAGG CTCCCACAGC CTGTGTGTGT GAGCCCGTTG
 301
   GCCTCACAAG CTCTTCTGAA GTTCTCTGGT TTCCTCTGAC TTCTCTTCTG
 351
   GATGGGAACT TAGGAGTGGG GTTTTATCTG GACTGAATTG GAGGAAAGGA
 401
   ACTGGCACCA GAGTGGAGGC TGGAAGGATG GTGGAACTGA CACTGGGACA
 451
   GAACACAGAA AACAGAATCT AATGCTATTG CTCTGGCTAT ATCAAATAAA
 501
   AGTGTCAAGA GTGAGCATCC TTGCCTTGTG CTGAATCACA AAGGAATACC
 551
   TTTCAGTTTT TCTCCATTGA TTATGATAGC AGTGGGCTTT TCACAGTGGG
 601
   CTTTACTGTG TTGAGAGTGG AATAAGGACA ATGATGCAAA AAAGAGCCAT
 651
   TCTCGGAGGG AGTCAGTCAG GGGCAAGCTC AGCGTGTCAG GAGACACCAG
 701
   GGAGGCCCAG GTGGAACCTC TCATGCTTGC ACTGCGCAAG GGGAGCCAAT
 751
   CCCGGGTGGA GTCCTGGGGC CAGGACAGAA GTGAACCCTA ACTTGTGAGA
 801
   GCTGATATAC AAAGGATGAC GGCTGGACCC CACATAAAAA TAGAACTCTG
 851
   ACCCACAGCC TGCAGCAACC TGTCCTAAAA AACCAATCCT CTTATCTACA
 901
   ATGAAAAGCC CAGGTAGCCA GCCTGCTAGA AGTCAGACTT GCAGGAAACC
 951
   AGATTGTTAT CTTTAGTCAT GATCCAAGAA GCTAAACAAT AACTTCTATA
1001
   ACAATCAGCC CCGAATGGCC AGGATGTGAT TAATAATTGG CAGCTTCCCT
1051
   AATTTTTATC CTGACTTCCA ACTTGGACCA ATCAGAGAAA GCTAACTATG
```

```
1051
     AATTTTTATC CTGACTTCCA ACTTGGACCA ATCAGAGAAA GCTAACTATG
1101
     CCCCCTAACC AATCCCATAG AATGCCAGGC TTCTGGTTAG CTCACCTACA
1151
     GCTTCCTCAG GCACAGCCCC AGGTATGGCC TGATTGGAGC CTGTCAGCAT
1201
     AAGGAAGTGG GAATGGACTC TCCCTTTTTC CACTCTAATC TTTCTCACTC
1251
     TCCTGCCCTC TTTTGAGTCT CTGCCAAAAT GTAAATGATG GCAGCTGACT
1301
     CCCTTGTTAT GGCAAGCTCT GAATAAACAG CCTTTGCTTT TCTCATCTGG
1351
     TTGGTCTTAA TTTATGTCCA TACTTGCTTG AAGCCCAGGT AGAATGGCTC
1401
     ATGTGTAGCC ATTGTCAGAC AGGGCCCTTT GTTTTTAATG GAAGTTGACT
1451
     TGGTCTCTTT GTGAGGGCAG GCACTGGACG GCAAGGATCA TAACTTACTC
1501
     ATCTCAGTAA TCTAAGTGTC CAGCACTTGC CTTGCTCATA ATAAAGATTG
```

Figure 4.12. The sequence of cDNA clone 11.2. The sequence with homology to the original probe used to identify this clone is boxed.

246m14, represented one of the exons upstream of the TA 6.35 in cDNA clone 1. The other 1436bp of sequence was novel (see figures 4.11 and 4.12).

Much of the sequence of cDNA clone 11:2 was highly repetitive. When the insert of clone 11.2 was used as a probe on a genomic Southern blot a smeary background signal was detected, even when the probe was competed with placental DNA. Translation of the sequence of clone 11:2 using the DNAStar software programme showed stop codons throughout the sequence in all possible frames.

A data base search using the public database BLAST N was done with the entire sequence of cDNA clone 11:2. The first 160bp was homologous to a known human L1 element (LINE). The rest showed no significant homology to any known sequence. This clone did not seem to contain any sequence that belonged to the unidentified 5' part of cDNA clone 1. It was thought that cDNA clone 11.2 may be a recombinant clone, formed artificially in the process of making the cDNA library, containing one exon of our original sequence with at least one repetitive L1 element and other uncharacterisable sequence. No further work was done on this clone at this time as it did not appear to further our understanding of cDNA clone 1.

4.9. Conclusion

In conclusion, a 199bp putative exon from within the region of deletion in B-cell CLL at 13q14.3 was identified by λ Get exon trapping techniques. This putative exon identified a 3.5kb lymphoid specific message on Northern blot analysis. When hybridised to a peripheral leucocyte cDNA library, this sequence identified 3 positive cDNA clones. The longest of these, cDNA clone 1, contained a 1.7kb insert. The original sequence, TA 6.35, was located 500bp from the 5' end of the clone and the sequence terminated in a poly A tail. When the entire clone was labelled and hybridised to DNA from patients known to have homozygous deletion of 13q14.3, no signal could be detected. Therefore, the clone was shown to lie completely within the deleted area at 13q14.3 in these patients.

Using a combination of hybridisation and PCR analysis, cDNA clone 1 was demonstrated to consist of at least 7 exons, including TA 6.35, lying on 6 separate EcoR I genomic fragments. The clone was physically mapped and was shown to span a 600kb genomic region at 13q14.3, extending from an area 5' of D13S319 to a telomeric area close to D13S25 i.e. exons from cDNA clone 1

spanned both the minimal areas of deletion described by Bullrich *et al.*(Bullrich *et al.*, 1996) and Kalachikov *et al.*(Kalachikov *et al.*, 1997)

Despite the clear 3.5kb mRNA identified by the original TA 6.35, no Northern blot signal could be detected with a probe for the 5' sequence of this clone, amplified with the PCR primers F14 - B9. Expression of the 5' sequence had been demonstrated using RT PCR, in both normal lymphocytes and B-cell CLL lymphocytes. However, using the same technique, no link could be established between the sequence 5' of TA 6.35 and TA 6.35 itself.

cDNA clone 1 represented a candidate tumour suppressor gene cDNA from the deleted area at 13q14.3 that was expressed in normal and malignant lymphocytes. The transcript spanned a large genomic distance with exons located in the conflicting minimal areas of deletion described by other groups working in this field. Many questions about this transcript remained unanswered at this time. The aim at this time was to isolate the full length 3.5kb transcript identified by TA 6.35 on Northern blot hybridisation. Therefore, further cDNA library screening experiments were planned. These will be discussed in the next two chapters.

Chapter 5. The isolation of a second novel candidate tumour suppressor gene cDNA from within the minimal region of deletion at 13q14.3

5.1. Introduction

This chapter describes how, in an attempt to isolate the 3.5kb message identified on Northern blot analysis with the TA 6.35 probe described in chapter 4, a new gene was characterised.

Initial screening of a peripheral blood leucocyte cDNA library with the putative exon TA 6.35 had yielded a 1.7kb cDNA, cDNA clone 1. In order to identify the presumed additional 1.8kb of this message, both the 700bp BamH I / Pst I probe described in chapter 4 and the original TA 6.35 probe were further used to screen cDNA libraries from different tissues as described in this chapter. Although the 700bp probe contained 5' sequences from clone 1 including TA 6.35, the TA 6.35 sequence was the only probe that had been effective when hybridised to a Northern blot. Therefore, both probes were used in combination.

5.2. Screening of cDNA libraries to identify the full length message identified by TA 6.35.

5.2.1. JY (T lymphoblastoid cell line) plasmid cDNA library

The message identified by TA 6.35 on Northern blot analysis was strongest in thymic tissue and peripheral blood (figure 4.1), both of which are rich in T cells. A T lymphoblastoid cell line cDNA library was available to us, originally a gift from Dr David Simmonds (ICRF, Oxford). This library was constructed in the plasmid CDM8, which carries the ampicillin resistance gene, and propagated in *Escherichia coli* MC1061/P3.

The library was titrated and 780,000 colonies were then plated on LB agar with ampicillin. Hybond-nylon filter lifts were taken and hybridised to the 700bp BamH I / Pst I probe. After 3 days exposure to autoradiography film (Fuji Medical) at -20°C all candidate positive clones were picked and taken through to secondary screening. No clones were confirmed positive. A further 720,000 colonies were plated and the nylon filter lifts were hybridised to the TA 6.35

probe. Again, no positives were seen after 3 days exposure of the hybridised filters to autoradiography film.

It should be noted that a lymphoblastoid cell line consists of immature, abnormal T cells that are arrested prematurely in their development, and is not representative of the T cells found in normal peripheral blood and thymic tissue. One explanation for the discrepancy between the negative screening of this library and the Northern blot data was that the gene represented by this 3.5kb message was not expressed by T cells at the lymphoblastoid stage of development.

5.2.2. Foetal spleen cDNA library

The Northern blot message was also detected in adult spleen. No adult spleen cDNA library was available, but a commercial library made from foetal spleen cDNA was obtained (Stratagene). These tissues are at different developmental stages and may, therefore, express quite different genes. This library was a bacteriophage library constructed in λ Zap and propagated in *Escherichia coli* XL-1 Blue MRF strain (tetracycline resistant).

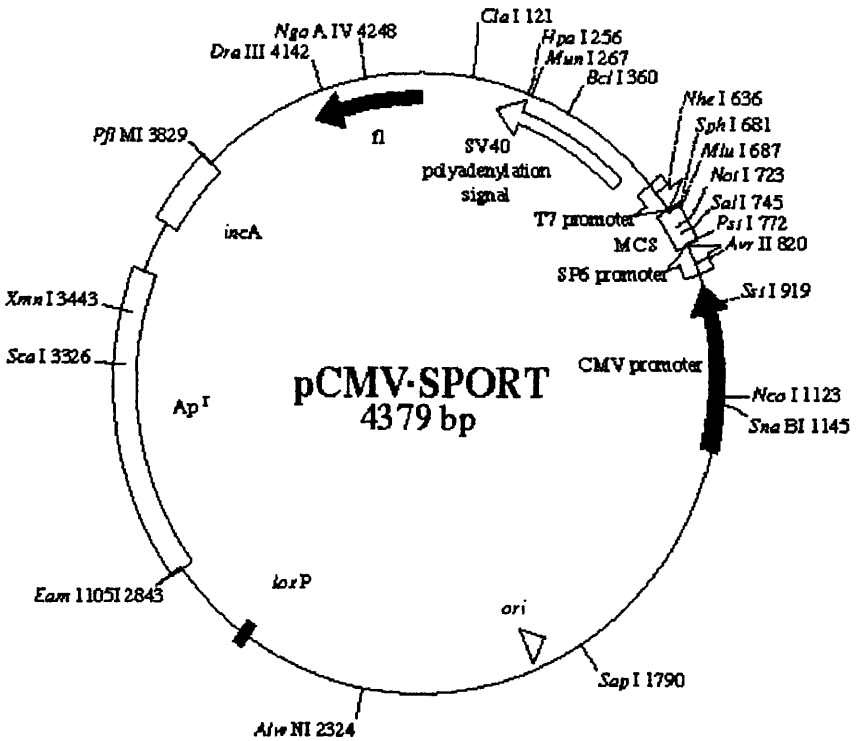
The library was titrated and 1,200,000 colonies were then plated on LB agar supplemented with tetracycline (7.5μg/ml). Filter lifts were taken as for bacteriophage libraries and were hybridised to the TA 6.35 probe. None of the 8 possible positive signals identified during the primary screening were confirmed as positive at the secondary screening stage.

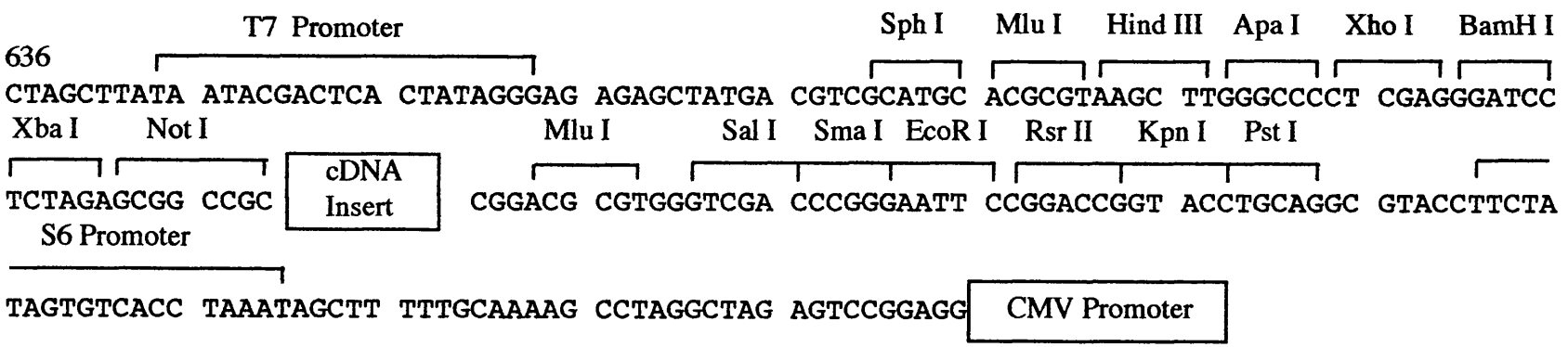
5.2.3. Peripheral leucocyte cDNA library

Another peripheral leucocyte cDNA library (Gibco BRL) became available at this stage of the project. The library was made using peripheral blood leucocytes from 4 normal males and was constructed in the plasmid vector pCMV.Sport, which carries the gene for ampicillin resistance. The inserts were cloned into a Not I - Sal I cloning site (figure 5.1). The plasmid was propagated in ElectroMAX™ DH12S™ cells.

The library was titrated and 2 million colonies were then plated on LB agar supplemented with ampicillin. Filter lifts of the library were taken and screened with the 700bp BamH I / Pst I probe. Six positive colonies were seen following exposure to Fuji Medical autoradiography film at -20°C overnight. These were then picked and taken through to secondary screening.

Figure 5.1
Schematic of Plasmid pCMV.Sport and the relevant vector sequence showing the cloning site





5.3. Results of secondary screening of positives from the peripheral leucocyte cDNA library

Of the 6 positives seen after screening 2 million colonies from this peripheral leucocyte cDNA library, 5 remained positive on secondary screening. An example is shown in figure 5.2. Individual colonies were then picked and grown in 1.5ml cultures of LB media with ampicillin. The DNA was prepared using a Qiagen miniprep kit (section 2.1.2). An aliquot of the DNA was digested with EcoR I and BamH I to release the insert (figure 5.1). The inserts were then characterised by electrophoresis. The gel was blotted and the filter was hybridised to check the genuine positivity of the clones. Clones 2, 3, 4 and 5 all contained inserts that were positive by hybridisation to this probe. Clone 2 had the largest insert at approximately 950bp. The inserts for the other clones were all identical in size, at approximately 900bp. Clone 1 did not contain any insert but was spuriously positive when hybridised to the 700bp BamH I / Pst I probe (figure 5.3).

In order to manually sequence the inserts from clones 2-5, the inserts were all subcloned into an M13 vector. DNA for each clone was prepared using the plasmid maxiprep method. An aliquot of the plasmid DNA was then digested with the restriction enzymes EcoR I and BamH I to release the inserts (figure 5.1). The inserts were then subcloned into M13mp18 and M13mp19 vectors prepared with EcoR I and BamH I. The clones were manually sequenced using the M13 -40 primer and subsequent primers derived from the sequence as it was obtained. Some of the primers used are those described in chapter 4 (figure 4.9);

F3 primer

5'-CGCACTTAAGCCTCGGAACAAC-3'

F17 primer

5'-CGGCTTGTAGAGACTTTTGCCG-3'

F19 primer

5'-GCGGTGAACTTAGAGGTTGTG-3'

2:2 mp18 R primer

5'-CTGTTATAGAACAGGATAAGGCAGG-3'

Clones 3, 4 and 5 were identical to each other and were shown to all be contained within clone 2 (the sequence of these clones started at position 45 of

Figure 5.2 A

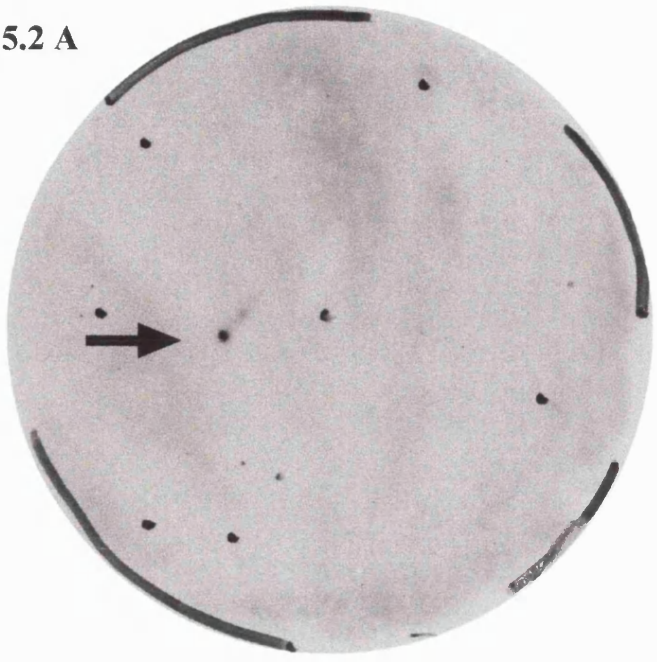


Figure 5.2 B

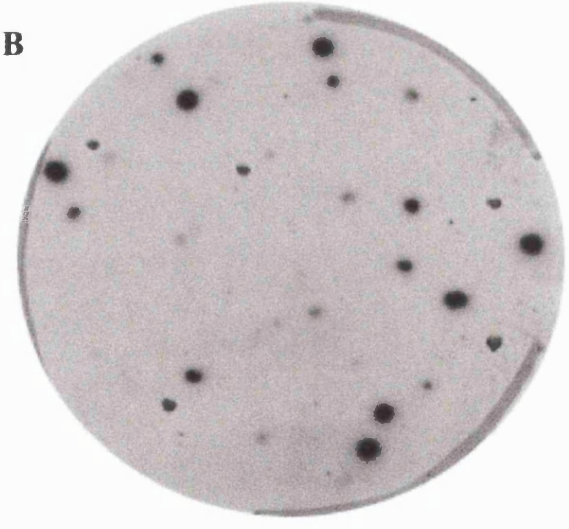


Figure 5.2 A: Primary cDNA library filter showing one positive colony, marked by the arrow, when hybridised to the 700bp BamH I / Pst I probe in Cambridge buffer overnight. The other marks are orientation marks. Final wash was at a stringency of 0.1 x SSC/ 0.1% SDS at 65°C. The filter was exposed to Fuji Medical autoradiography film overnight.

Figure 5.2 B: Positive colonies identified following secondary screening of the positive shown in 4.2 A. The filter was hybridised to the 700bp BamH I / Pst I probe and was washed at a final stringency of 0.1 x SSC/ 0.1% SDS at 65°C.The filter was exposed to Fuji Medical autoradiography film overnight.

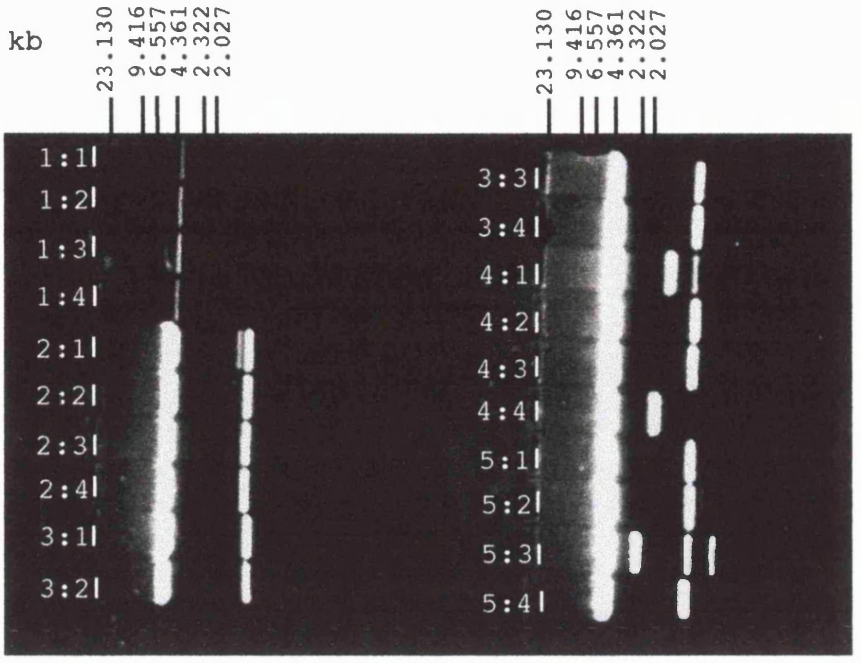


Figure 5.3 A

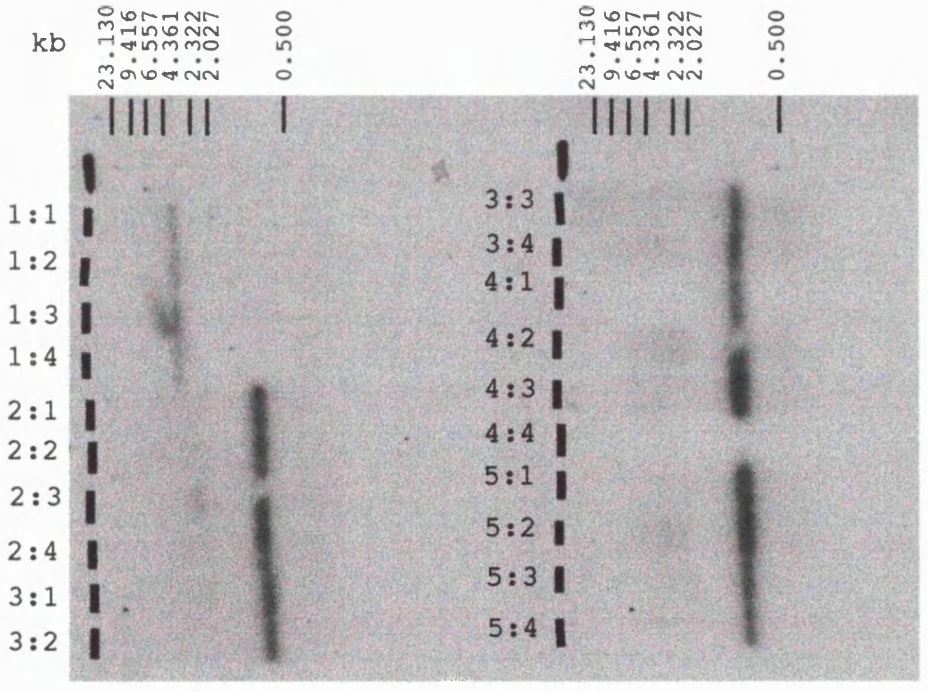


Figure 5.3 B

Figure 5.3 A: Polaroid of DNA from colonies positive for the 700bp BamH I/ Pst I probe, digested with BamH I and EcoR I and electrophoresed on a 0.8% 0.5 x TBE agarose gel stained with ethidium bromide. The 2.9kb band seen is the plasmid vector. Inserts can be seen in clones 2-5.

Figure 5.3 B: Southern blot of the gel shown in 5.3 A. The filter was hybridised to the 700bp BamH I / Pst I probe in Church buffer and then washed to a final stringency of 0.2 X SSC/0.1%SDS at 65°C. The filter was exposed to Fuji Medical autoradiography film overnight. Positive inserts can be seen in clones 2, 3, 4 and 5.

clone 2). All further work was, therefore, carried out on clone 2 (clone number 2:2 in figure 5.3).

5.4. The sequence of cDNA clone 2:2

The full sequence of cDNA clone 2:2 is shown in figure 5.4. The clone was 952bp in length terminating in a poly A tail. The homology to the original probe is shown in figure 5.5. Clone 2:2 extended 70bp further 5' from the original probe and, therefore, 70bp 5' of cDNA clone 1. The first 10bp of cDNA clone 1 were not present in cDNA clone 2:2. Homology with the probe ended at position 347 of cDNA clone 2:2 and the subsequent sequence had no homology to the original probe or to cDNA clone 1.

The sequence of cDNA clone 2:2 was sent to the public database BLAST (http://www.ncbi.nlm.nih.gov). The sequence had 100% homology to an EST, accession number w05790. The sequence of the EST had an additional 53bp sequence inserted at position 345. This additional sequence was identical to the small 53bp exon, described in chapter 4, which formed alternative splice versions when lymphocyte cDNA was amplified by RT PCR using primers F2 and B7 (figure 4.10).

5.5. Mapping of cDNA clone 2:2

cDNA clone 2:2 was mapped to the region at 13q14.3 using PCR. The first 347bp of the sequence were amplified using primers from cDNA clone 1, F1 and B4 (figure 4.9). Two further primers were designed to amplify the remaining sequence; 2:2 mp18 F and 2:2 mp18 B.

2:2 mp18 F

5'- TGGGAGATTGTTGCAAGGAA -3'

2:2 mp18 B

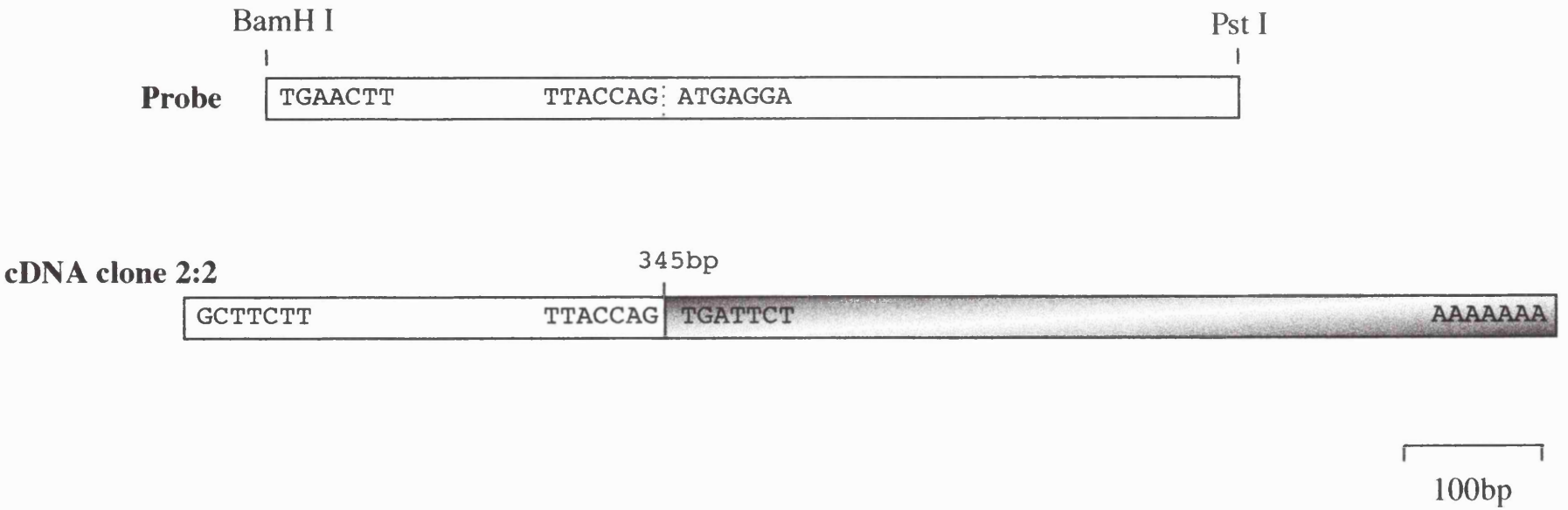
5'- ATTGAGTGCTTACTGTGTGCCCAG -3'

A standard 35 cycle PCR protocol was used with an annealing temperature of 54°C. Positive signals were seen for both amplifications on PAC clones 173a12 and 208p23. YAC 11AF3 was also positive for both amplifications. This positioned cDNA clone 2:2 centromeric of D13S319 at 13q14.3 (see map of 13q14.3, figure 3.1).

```
1
    GCTTCTTCGG TTGCAGTCCT CTTGCTTCTT GCGCGTGCGT GTAGCGCTTT
51
    TGCAAAGCCG CGGAGGTGAA GTGAACTTAG AGGTTGTGGG GCCGAGGGGT
101
    CGTCTTATAG CTACCAGCCC ACAGGCATTT AGTCTACGTT GGAGGTAAAC
151
    AAATACGGGT CCTGCTTAGG AGAAAAGAAA AACGTCTTAC AGCCAGTGTC
201
    TAAACTCCAA ACAACGGAAT GTATCAATGA GACCTTGTAT ATGGATACAC
251
    GTGCATTTAA AACCGCCCTG CCGGCTTGTA GAGCTTTTGC CGTTCTCCAG
301
    CGCTTTACAG GGGTTATCGC ACTTAAGCCT CGGAACAACT TTACCAGTGA
351
    TTCTACCAGA AAGGAATGAA GAACAGAACC TTCAGGAATT GAGTCACAAT
401
    GCAGACAAAT ATCAAATGGG AGATTGTTGC AAGGAAGAGA TTGATGATAG
451
    TATTTTCTAC TAGCCATTGG GAAGATAAAA GGAGACAGAA GATTGAAGCC
501
    TTTGCCAGCC ATTCTTTCCC TTTTTGCTTC CAAACTCCTC AACTGGGAAC
551
    CTTCATATGT GCAGTATTTA TATTGGATCA TACTGGTGAT TATAAAAGTT
601
    CCTAGGAGGC TAGAAGAGCC AACCAACAGA GAAGGGAAAG CAGTCTGTTC
651
    TGAACATAGG GACATAAGTT CATTCATGCC AAGTATCTTT CCAGCATGTT
701
    TCTCCCATTT AGAATATCTA GCATGTAAGG CCTTTCAATA TTAATATAAG
751
    CCCAATATCA GCTCTTTCTC TTTGTATTTC ATCTCTTTCT ACTCTCCTAT
801
    TTGTATTTTG TGTTCCTATC AAAGTGTCGT ATCTGGGAGA TGACCTGCCT
851
    TATCCTGTTC TATAACAGTT TTGTTTGCTG CTGTGTCTTT AGAACAGTGC
901
    CTGGCCACAC AGTAAGCACT CAATAAATCT TTGATGAATG AAAAAAAAAA
951
    AA
```

Figure 5.4. Sequence of cDNA clone 2:2. The border of the homology with the original cDNA clone 1 is indicated by an arrow.

Figure 5.5. Comparison between the 700bp cDNA probe used to screen the peripheral leucocyte cDNA library and the cDNA clone 2:2 isolated from the screening.



5.6. Characterisation of cDNA clone 2:2

To characterise cDNA clone 2:2 further, a probe for the insert was made by digesting the original plasmid DNA preparation with the restriction enzymes Sal I and Not I (figure 5.1). This fragment was then purified from a gel slice following separation from the vector on a 1% agarose gel. The purified DNA was then labelled and hybridised to Southern blots of genomic DNA digested with both the restriction enzymes EcoR I and Hind III. The results of these hybridisations can be seen in figure 5.6. With an EcoR I digest the cDNA clone identified an 8.5kb fragment and a 20kb fragment. With a Hind III digest, the probe identified a 5.5kb fragment and an 8.5kb fragment. When the 700bp BamH I / Pst I probe was hybridised to genomic Southern blots one of the bands seen with an EcoR I digest was 8.5kb and one of the bands seen with a Hind III digest was 5.5kb. These bands were likely to be those identified by the first 345bp of clone 2:2 which was the sequence common to both clones. This interpretation was confirmed by hybridising the same Southern blots to a PCR probe made by amplifying this 347bp sequence, with primers F1 and B4, from the original cDNA clone 1 plasmid DNA (figure 4.9).

The insert from cDNA clone 2:2 was then hybridised to Southern blots from DNA of patients known to have homozygous deletion of D13S319 and D13S25. No signal could be detected. The hybridisation was controlled with DNA from non CLL samples. This experiment showed that cDNA clone 2:2 was located within the deleted region at 13q14.3 in these patients.

5.7. Expression of cDNA clone 2:2

5.7.1. Northern blot data

The 700bp BamH I / Pst I probe (containing 347bp of cDNA clone 2:2 and including the TA 6.35 sequence) was hybridised to a Northern blot of mRNA from a variety of immune tissues (Clontech Immunoblot). The manufacturers instructions were followed, the blot being hybridised to the probe in the provided Express Hyb™ buffer at 65°C for 1 hour. The blot was then washed at a final stringency of 0.2% SDS / 0.2 x SSC at 65°C. After exposure to autoradiography film (Fuji Medical) for 72 hours at -20°C a band could be seen at 1.1kb in all lanes.

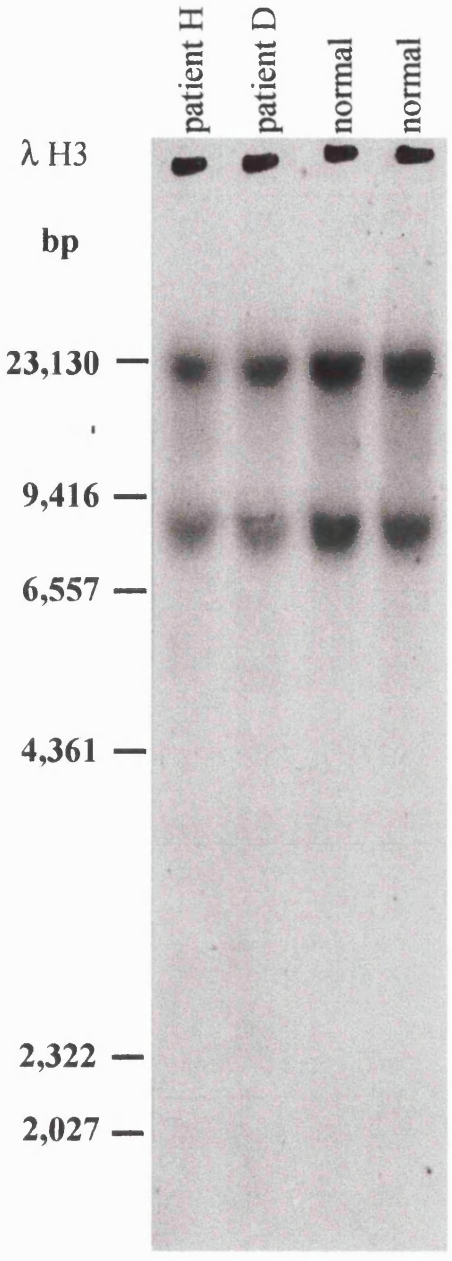


Figure 5.6. A

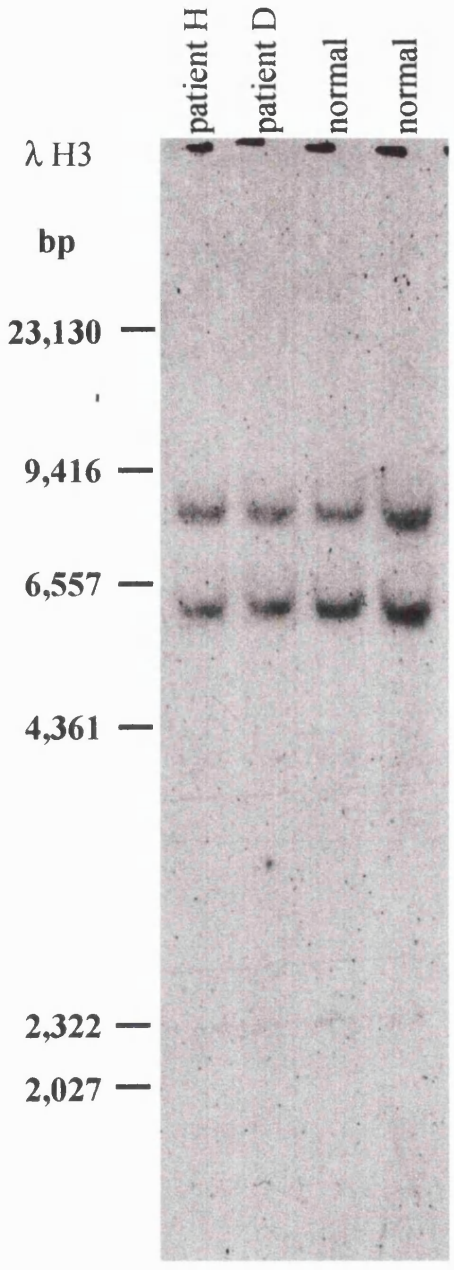


Figure 5.6. B

Figure 5.6.A. Southern blot analysis of the cDNA clone 2:2. 10µg of human genomic DNA from both normal controls and patients with B-cell CLL was digested with EcoR I, blotted and hybridised to the human probe cDNA clone 2:2. Final washing conditions were 0.2 x SSC/ 0.1% SDS, 65°C. The filter was exposed to Fuji Medical autoradiography film for 10 days at -20°C. 2 bands can be seen at 8.5kb and 20kb.

Figure 5.6.B. Southern blot analysis of the cDNA clone 2:2. 10µg of human genomic DNA from both normal controls and patients with B-cell CLL was digested with Hind III, blotted and hybridised to the human probe cDNA clone 2:2. Final washing conditions were 0.5 x SSC/ 0.1% SDS, 65°C. The filter was exposed to Fuji Medical autoradiography film for 6 days at -20°C. 2 bands can be seen at 5.5kb and 8.5kb.

This was more pronounced after 2 weeks exposure (figure 5.7). A faint band was seen at 3.5kb, as was identified by the original TA 6.35 probe.

5.7.2. Amplification of cDNA clone 2:2 from normal cDNA by RT PCR

Primers were designed to amplify the full length clone by RT PCR; 2:2 F and 2:2 R.

2:2 F primer

5'-GCAGTCCTCTTGCTTCTTGCGCGTGGT-3'

2:2 R primer

5'-AGGCTTAAGTGCGATAACCCCTGTA-3'

A standard 35 cycle PCR programme was used with an annealing temperature of 65°C. No signal was obtained with cDNA made from total RNA extracted from normal bone marrow mononuclear cells or cDNA made from thymus mRNA (Clontech). Different PCR conditions were used; the annealing temperature was systematically decreased to 58°C and the Taq was changed from Promega Taq to a Long Range Taq (Clontech). However, the PCR remained negative

5.8. Publication of the novel candidate tumour suppressor gene, Leu 1

A paper was published at this stage in the project describing the isolation of two novel candidate tumour suppressor genes from within the minimally deleted area at 13q14.3 in B-cell CLL (Liu *et al.*, 1997). These genes were named Leu 1 and Leu 2 and were located centromeric of D13S319. The sequence and characteristics of Leu 1 were almost identical to those of cDNA clone 2:2. The expression of Leu 1 was demonstrated by Northern blot analysis, identifying a band of approximately 1.1kb in many tissues including testis, thymus, spleen and peripheral blood leucocytes. This corresponded to our Northern blot data described above (figure 5.7). The published sequence of Leu 1 extended 40 base pairs further 5' of cDNA clone 2:2. An open reading frame encoding for a 72 amino acid protein was described by the authors. When the sequence was compared to that of cDNA clone 2:2 there was an extra T nucleotide at position 483 of Leu 1 (position 443 of clone 2:2). This T was part of the stop codon in the open reading frame of Leu 1. The absence of this base in cDNA clone 2:2 gave

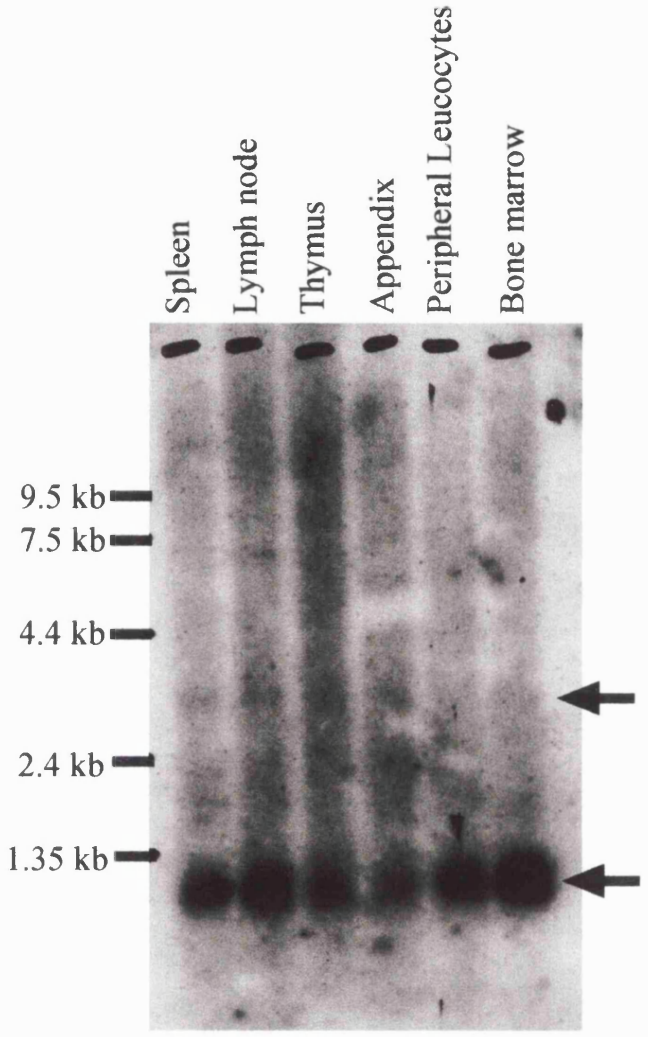


Figure 5.7. Northern blot analysis of the 700bp Pst I/BamH I probe A Clontech Human Immunoblot was hybridised for 1 hour at 65°C to the human cDNA clone 2:2 using the manufacturers Express Hyb buffer. After a high stringency wash at 0.2 x SSC/0.2% SDS at 65°C, the filter was exposed to Fuji Medical autoradiography film for 2 weeks. A band of approximately 1.1 kilobase can be seen in all lanes. A fainter band at 3.5kb can also be seen in the first 4 lanes.

cDNA clone 2:2 an open reading frame of 234bp. The corresponding open reading frame for cDNA clone 2:2 is shown in figure 5.8 and the differences between the reading frames of the two clones are schematically represented in figure 5.9.

5.9. The demonstration of splice variant forms of cDNA clone 2:2

Amplification by PCR of the entire cDNA clone 2:2 from cDNA had been unsuccessful. However, when primers were designed flanking the open reading frame described above, products were amplified from cDNA made from normal lymphocytes and thymus. The same result was achieved when the PCR was repeated on cDNA made from malignant lymphocytes obtained from patients with B-cell CLL. The PCR programme used was a standard 35 cycle protocol with an annealing temperature of 52°C. The primers used were C12T7(5)- F and C12T3(6)- R. They correspond to two of the primers published by Liu *et al.* (Liu *et al.*, 1997), except that restriction sites for EcoR I on the sense primer and Hind III on the antisense primer were added to each oligonucleotide respectively to enable cloning of the PCR products.

C12T7(5)- F

5'- GTCTACGTTGGAGGTAAACA -3'

C12T3(6)- R

5'- ATGTCCCTATGTTCAGAACA -3'

When the PCR products were resolved by electrophoresis, two products could be clearly seen in all lanes. The two PCR products were amplified from both normal thymus cDNA and cDNA from patients' malignant B lymphocytes (figure 5.10.A). The PCR products from the normal cDNA were cloned into a Bluescript plasmid (KS+) vector and the inserts were sequenced using M13 -40 and T3 primers with an ABI 377 automated sequencer. The sequences obtained can be seen in figure 5.10.B. The sequence of the smaller band was the sequence predicted from cDNA clone 2:2. The sequence of the larger band had an additional 53 base pairs at position 347 which was identical to the 53bp 'exon' described in chapter 4 that formed an alternative splice version, revealed by RT PCR using primers F2 and B7 (figure 4.9). The EST w05790 discussed above was 100% homologous to the longer PCR product amplified by these primers i.e. it contained this extra 53bp sequence. It appeared that this 53bp 'exon' formed an alternative splice version of both cDNA clone 1 and cDNA clone 2:2.

```
   1
   GCTTCTTCGG TTGCAGTCCT CTTGCTTCTT GCGCGTGCGT GTAGCGCTTT
  51
   TGCAAAGCCG CGGAGGTGAA GTGAACTTAG AGGTTGTGGG GCCGAGGGGT
 101
   CGTCTTATAG CTACCAGCCC ACAGGCATTT AGTCTACGTT GGAGGTAAAC
 151
   AAATACGGGT CCTGCTTAGG AGAAAAGAAA AACGTCTTAC AGCCAGTGTC
 201
   TAAACTCCAA ACAACGGAAT GTATCAATGA GACCTTGTAT ATGGATACAC
                                 M  R   P  C  I   W  I  H
 251
   GTGCATTTAA AACCGCCCTG CCGGCTTGTA GAGCTTTTGC CGTTCTCCAG
   V  H  L  K   P  P  C   R  L  V   E  L  L  P   F  S  S
 301
   CGCTTTACAG GGGTTATCGC ACTTAAGCCT CGGAACAACT TTACCAGTGA
    A  L  Q   G  L  S  H   L  S  L   G  T  T   L  P  V  I
 351
   TTCTACCAGA AAGGAATGAA GAACAGAACC TTCAGGAATT GAGTCACAAT
     L  P  E   R  N  E   E  Q  N  L   Q  E  L   S  H  N
 401
   GCAGACAAAT ATCAAATGGG AGATTGTTGC AAGGAAGAGA TTGATGATAG
   A  D  K  Y   Q  M  G   D  C  C   K  E  E  I   D  D  S
 451
   TATTTTCTAC TAGCCATTGG GAAGATAAAA GGAGACAGAA GATTGAAGCC
    I  F  Y   .
 501
   TTTGCCAGCC ATTCTTTCCC TTTTTGCTTC CAAACTCCTC AACTGGGAAC
 551
   CTTCATATGT GCAGTATTTA TATTGGATCA TACTGGTGAT TATAAAAGTT
 601
   CCTAGGAGGC TAGAAGAGCC AACCAACAGA GAAGGGAAAG CAGTCTGTTC
 651
   TGAACATAGG GACATAAGTT CATTCATGCC AAGTATCTTT CCAGCATGTT
 701
   TCTCCCATTT AGAATATCTA GCATGTAAGG CCTTTCAATA TTAATATAAG
 751
   CCCAATATCA GCTCTTTCTC TTTGTATTTC ATCTCTTTCT ACTCTCCTAT
 801
   TTGTATTTTG TGTTCCTATC AAAGTGTCGT ATCTGGGAGA TGACCTGCCT
 851
   TATCCTGTTC TATAACAGTT TTGTTTGCTG CTGTGTCTTT AGAACAGTGC
 901
   CTGGCCACAC AGTAAGCACT CAATAAATCT TTGATGAATG AAAAAAAAAA
 951
   AA
```

Figure 5.8: cDNA clone 2:2 showing the postulated open reading frame from the ATG to the TAG. The primers used to amplify the coding region in patients are underlined.

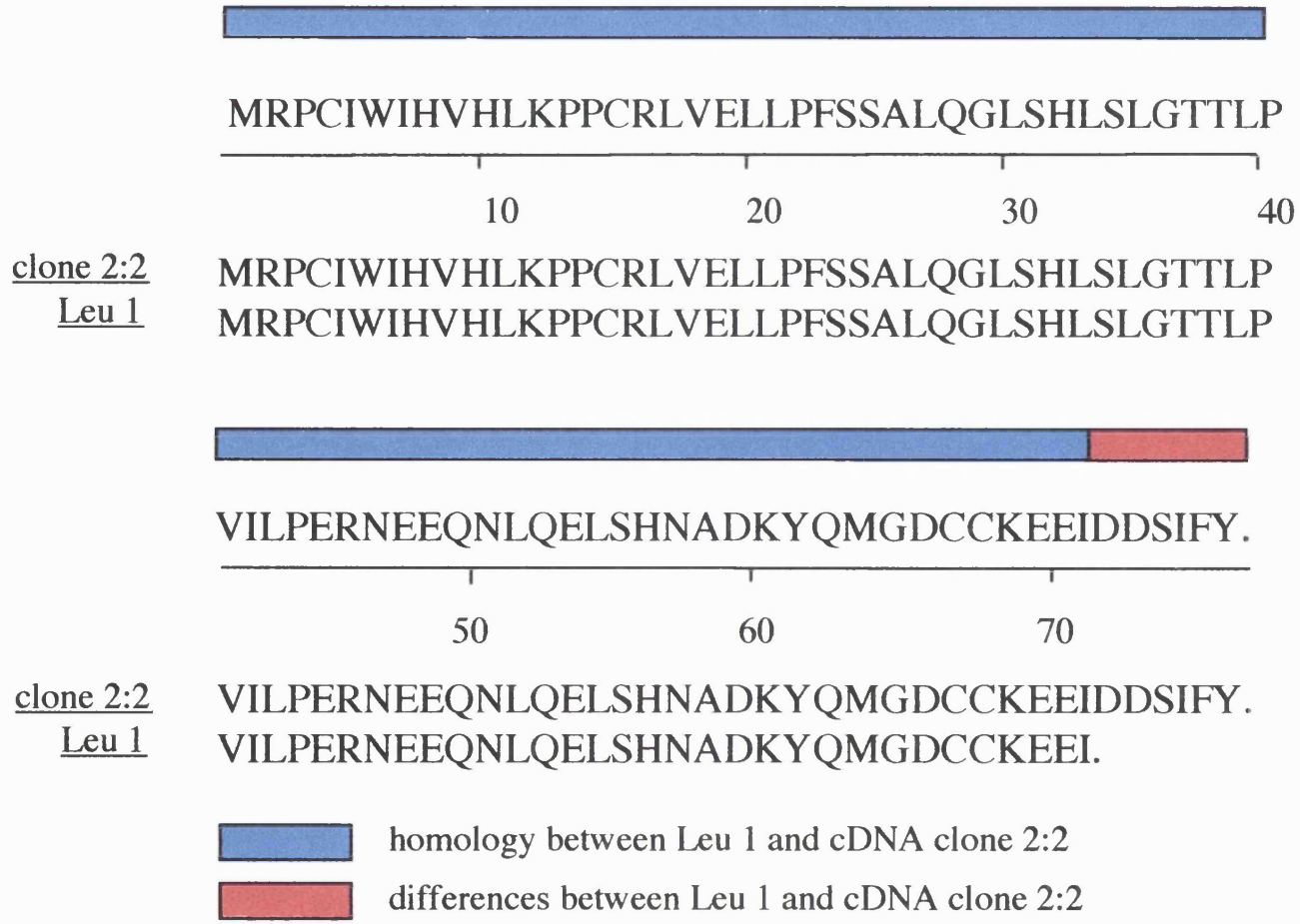


Figure 5.9: Showing the homologies and differences between the proteins potentially coded for by the gene Leu 1 and cDNA clone 2:2.

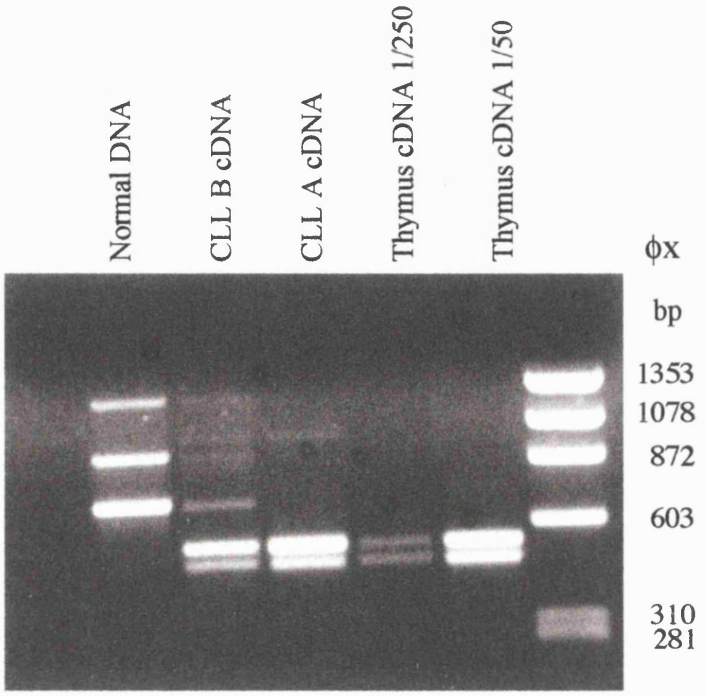


Figure 5.10.A: Showing PCR products from amplification of the open reading frame of cDNA clone 2:2, elctrophoresed on a 1% agarose 0.5 x TBE gel stained with ethidium bromide. All cDNAs, including those from CLL patients A and B, show 2 bands when amplified with the primers C12T7(5)-F and C12T3(6)-R.

```
  1
  GTCTACGTTG GAGGTAAACA AATACGGGTC CTGCTTAGGA GAAAAGAAAA
 51
  ACGTCTTACA GCCAGTGTCT AAACTCCAAA CAACGGAATG TATCAATGAG
101
  ACCTTGTATA TGGATACACG TGCATTTAAA ACCGCCCTGC CGGCTTGTAG
151
  AGCTTTTGCC GTTCTCCAGC GCTTTACAGG GGTTATCGCA CTTAAGCCTC
201
  GGAACAACTT TACCAGATGA GGACACCTGA GGTTCAGATT AAGAAATCTG
251
  CCCCAAAGTC TTAGAACTGT GATTCTACCA GAAAGGAATG AAGAACAGAA
301
  CCTTCAGGAA TTGAGTCACA ATGCAGACAA ATATCAAATG GGAGATTGTT
351
  GCAAGGAAGA GATTGATGAT AGTATTTTCT ACTAGCCATT GGGAAGATAA
401
  AAGGAGACAG AAGATTGAAG CCTTTGCCAG CCATTCTTTC CCTTTTTGCT
451
  TCCAAACTCC TCAACT
```

Figure 5.10B: Sequence of the larger band, shown in 5.10A. The 53 base pair sequence shown in bold is the sequence spliced out in the smaller band also shown. The stop codon within the additional sequence is underlined.

When the additional sequence was included in the sequence of cDNA clone 2:2, the open reading frame was disrupted as there was a stop codon within this added sequence. This is underlined in figure 5.10.B. This 53bp sequence is not the intron between the two exons of Leu 1 as they have been shown to lie over 19kb apart at the genomic level (Liu *et al.*, 1997).

5.10. Mutational analysis of cDNA clone 2:2 in our B-cell CLL patients

Liu *et al.* (Liu *et al.*, 1997) had carried out mutational analysis of Leu 1 in 180 of their B-cell CLL patients with heterozygous loss of 13q14 and found no abnormalities. It was, however, relevant to analyse a small cohort of our patients with heterozygous deletion of 13q14.3 as the sequence of our clone had some minor differences to Leu 1.

According to Knudson's 'two hit' theory of tumour suppressor genes (Knudson, 1971), both copies of the gene must be inactivated within a cell for malignant potential to occur. Six patients known to have heterozygous deletion of 13q14.3 were chosen for analysis. The patients chosen were numbers 4, 14, 48, 54, 62 and patient K. The data showing the full genomic status of these patients are described in chapter 7. Subsequent to this work, patients 4 and 62 were shown not to have heterozygous deletion of 13q14.3. Patients 14, 48 and 54 were shown to have heterozygous deletion of all markers tested at 13q14.3. Patient K had heterozygous deletion of D13S25 and D13S319. The data for patient K are not shown as they were from work done by Dr Jabbar from our group.

cDNA was made from the total RNA isolated from malignant, purified B lymphocytes in the selected patients. The PCR reaction described in section 5.9. to amplify across the open reading frame of cDNA clone 2:2 was repeated using patient cDNA. Both splice versions of the clone were seen in all patients when the RT PCR products were analysed by electrophoresis. This was an indication that the 3' splice site of exon 1 and the 5' splice site of exon 2 were intact. The existence of the two splice versions also indicated that both splice sites of the extra 53bp exon were present within the RNA of these patients. The entire PCR product for each patient was purified, digested and cloned into a Bluescript plasmid (KS+) vector. Despite all samples having been monocyte and T cell depleted (see chapter 7), at least 3 clones were sequenced for each patient to minimise the chance of sequencing cDNA amplified from normal contaminating cells, and to account for error that may arise from PCR amplification. When all the

clones were analysed for each patient using DNAStar software, no base changes were consistently seen for any patient. The sequence of cDNA clone 2:2 seen in all the patients was the same as our original cDNA clone and differed from the published sequence of Leu 1 at base pair 443 of cDNA clone 2:2 as already described. It was concluded that the postulated open reading frame of cDNA clone 2:2 was not mutated in any of these patients. It should be noted that we had not analysed the full non coding 5' and 3' regions in these patients. This data was in keeping with the data from Liu *et al.* (Liu *et al.*, 1997) and, more recently, Rondeau *et al.* (Rondeau *et al.*, 1999).

5.11. Conclusion

This chapter has shown how a novel gene was isolated by cDNA library screening with a probe initially designed from a longer, more complex cDNA. Both cDNAs had different 3' regions with poly adenylation sites and poly A tails. Only the first 5' exon was shared between the 2 transcripts. The subsequent 53 base pairs of cDNA clone 1 (exon 1a) also existed in the cDNA clone 2:2 as an alternative splice version. This sequence was not contained within the clone isolated from cDNA library screening, but was only seen when normal lymphocyte cDNA was amplified by RT PCR using primers designed to amplify across the two exons. The EST homologous to clone 2:2 identified through the BLAST N database also contained this additional sequence.

cDNA clone 2:2 was 960bp in length and the 700bp probe containing exon 1 of this clone identified a message on Northern blot analysis of 1.1kb (corresponding with the Northern blot data published by Liu *et al.*(Liu *et al.*, 1997)). The original clone isolated by cDNA library screening consisted of 2 exons which mapped to 13q14.3, centromeric of D13S319. Both of these exons were shown to lie within the deleted area in our patients with homozygous deletion of the markers D13S319 and D13S25 as described. The cDNA existed as two splice forms. The smaller version had a postulated open reading frame encoding for a 78 amino acid protein. The larger version would give rise to a very short open reading frame of 44 amino acids. No mutations could be found within the open reading frame of this gene on the retained allele in any of the four patients with heterozygous deletion of 13q14.3 that were tested. Although no mutations were detected, this gene remains a candidate tumour suppressor gene from 13q14.3 because expression of the retained allele could be reduced or absent due to

mutations in other sections of the gene such as the promoter region, which has yet to be identified (discussed further in chapter 8), or the 3' and 5' untranslated regions.

In view of the data available at this stage of the study, we also analysed clone 14:2 and 19:3 (chapter 4) which were found to be 100% homologous to cDNA clone 2:2. Clones 3, 4 and 5, isolated from the peripheral leucocyte library (Gibco) discussed above, were contained within cDNA clone 2:2, starting at position 40.

The relationship between cDNA clone 1 and cDNA clone 2:2 remained to be clarified. It was possible that cDNA clone 1 was a hybrid clone, formed during the construction of the cDNA library by abnormal fusion of part of cDNA clone 2:2 with a further transcript from 13q14.3, comprising the 3' part of the sequence. However, the RT-PCR reaction using primers F2 and B7 from cDNA clone 1 successfully amplified the expected product, linking these two sequences (chapter 4).

Another explanation for the relationship between cDNA clones 1 and 2 was that the first 5' exon of cDNA clone 2:2 was part of a second transcript, due to alternative splicing. It had already been demonstrated that this exon could be alternatively spliced either to exon 2 of cDNA 2:2 or to the third 53bp exon, exon 1a, which was part of the sequence of cDNA clone 1. Alternative splicing of RNA transcripts is a common, normal phenomenon. There are 3 main forms of alternative splice versions seen in RNA processing (Lewin, 1994);

1. The first form of alternative splicing occurs when the 5' and 3' exons remain the same but the number of exons within the message may change; exon 'skipping' (figure 5.11.A). All variants may be functional and translated into a protein, or only some spliced forms may be translated. In some genes, the exon that is 'spliced in' contains a stop codon causing premature termination of the protein e.g. P elements of *D. melanogaster* show a tissue specific splicing pattern causing a longer protein to be translated in germline cells but not in somatic cells due to the splicing out of a termination codon. In the case of cDNA clone 2:2, the 2 alternative splice versions that have been demonstrated to occur would fit into this category. The shorter splice variant has a postulated open reading frame which is disrupted when the third exon is spliced in.

2. The second form of alternative splicing occurs when the 5' part of the gene remains constant, but the 3' part is completely different (figure 5.11.B). An example of this type of alternative splicing is seen in sex specific exons being

Figure 5.11: The different forms of altermative splicing which may generate a variety of protein products from an individual gene.

Figure A: Type 1
The 5' and 3' exons remain the same but the exons in the middle can vary.

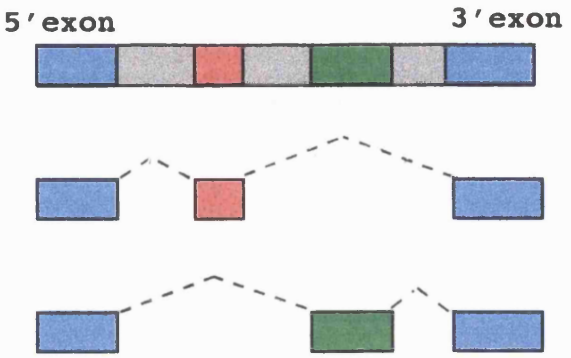


Figure B: Type 2
The 5' region remains constant but the 3' part of the gene is completely different.

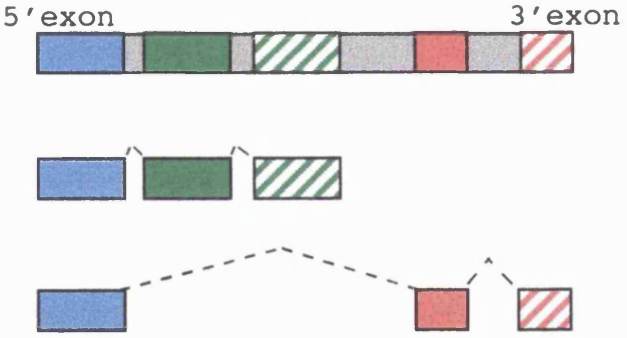
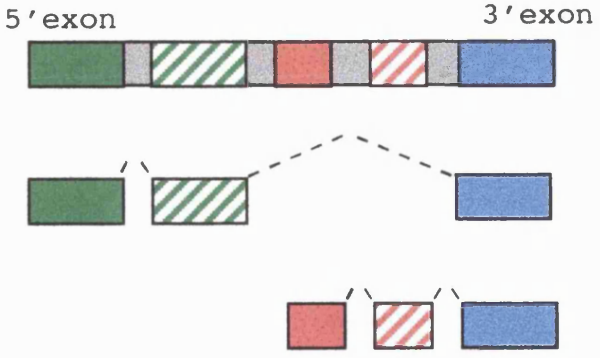


Figure C: Type 3
The 5' region of the gene varies but the 3' region remains constant.



spliced to the 3' region of the *dsx* mRNA in *D. melanogaster*. If cDNA clones 1 and 2:2 do share the same first two 5' exons, then their existence could be explained by this form of alternative splicing.

3. The third form of alternative splicing is similar to that in example 2, except that the 3' part of the gene remains constant and the 5' part changes. The large T/t antigens of SV40 are generated by connecting a varying 5' site to a constant 3' site (figure 5.11.C).

Despite the attractiveness of this theory, some of our expression data remained to be clarified. The message identified by TA 6.35 was 3.5kb. cDNA clone 1 was 1.7kb in length. If the 5' exon of clone 1 was shared with clone 2:2 then the model of splicing adopted would suggest that there were no further exons upstream. That would leave 1.8kb of message unaccounted for.

cDNA clone 2:2 appeared to be a genuine gene from the area of deletion. We had identified it from more than one library and it had been published by others. cDNA clone 1 had only been isolated from one cDNA library even though 3 other libraries had been extensively screened with probes containing the TA 6.35 sequence, including a second peripheral leucocyte library. Therefore, it was decided to go back to the original putative exon, TA 6.35 to try and identify the transcript that it identified on Northern blot by using RACE PCR and further cDNA library screening. This work will be discussed in chapter 6.

Chapter 6. Identification of a novel coding sequence from 13q14.3 by RACE PCR and the consequent clarification of cDNA clone 1.

6.1. Introduction.

Earlier chapters have described the identification of two candidate tumour suppressor gene cDNAs from within the region of deletion at 13q14.3. The first, cDNA clone 1, was identified following screening of a peripheral leucocyte cDNA library with the putative exon TA 6.35. Concerns regarding the genuine existence of cDNA clone 1 had arisen and it was decided to return to the original exon, TA 6.35, to try and characterise the full 3.5kb message that it identified on Northern blot, from the region at 13q14.3. The techniques that were chosen to do this were further cDNA library screening and RACE PCR.

6.2. cDNA library screening with the TA 6.35 probe.

The peripheral leucocyte plasmid cDNA library (Gibco) described in chapter 5 was re-titrated and 500,000 colonies were plated. The filters were hybridised to the TA 6.35 probe in Cambridge buffer and washed at a final stringency of 0.2 x SSC/ 0.1% SDS. No positives were seen.

A human thymus cDNA library (Clontech) constructed in the vector λGT10 and propagated in C600 *Escherichia coli* cells was obtained. The library was titrated and a total of 2.5 million plaques were screened with the TA 6.35 probe as described before. No positive colonies were seen. To control for the hybridisation and for the efficiency of this library, these filters were hybridised to a probe for the TCR (T cell receptor) β gene (Foroni *et al.*, 1991). This gene is expressed in T lymphocytes. The total number of positives was compatible with a fully representative library for thymus tissue.

Dr V. Duke from our laboratory then screened the following cDNA libraries with the TA 6.35 probe, using conditions as described before.

1. A further 2 million plaques from the human thymus cDNA library (Clontech).
2. A further 1 million clones from the peripheral leucocyte cDNA library (GibCo).

3. One million colonies from a K562 (human erythroleukaemia cell line) cDNA library (Clontech). This library was constructed in the vector λTriplEx^{TM} and propagated in XL1-Blue *Escherichia coli* cells.
4. One million colonies from an embryonic mouse stem cell genomic library that was constructed in the phage vector λ2001 and propagated in the *Escherichia coli* cells Q358.
In addition the above cDNA library clones were screened with DNA from both YAC 37AF3 and YAC 32EH6 (see figure 3.1). The YAC DNA was sonicated into fragments of approximately 500 base pairs and was competed with human placental DNA before being used as a probe. None of the screenings yielded any positive clones.

One explanation for the negative cDNA library screening, when TA 6.35 was used as a probe, compared to the strong signal seen on a lymphoid Northern blot was that the mRNA identified on Northern blot had a strong secondary structure. This would potentially lower the efficiency of reverse transcription by the superscript enzyme in the cDNA reaction used to prepare cDNA libraries. As a consequence the relatively abundant RNA message would be barely detectable at the cDNA level.

A different approach for isolating full length cDNA messages is RACE PCR (discussed in chapter 1). This method also relies on cDNA initially being synthesised from RNA. However, the message is identified by PCR techniques rather than hybridisation, and consequently cDNAs present in very low copy numbers can be identified (Frohman *et al.*, 1988). The efficiency of RACE PCR is further increased if a nested PCR reaction is used.

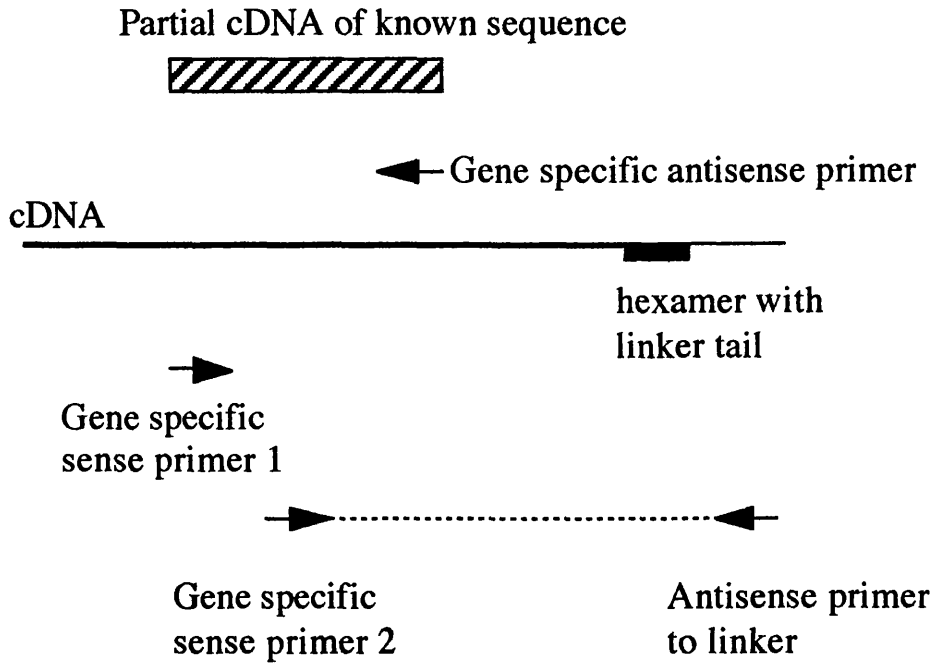
6.3. RACE PCR

Rapid amplification of cDNA ends (RACE) by PCR is a technique that has been developed in order to complete cDNA sequences both in the 5' and 3' directions. It uses PCR to amplify from regions of known sequence into regions of unknown sequence (Frohman *et al.*, 1988). In order to do this, the cDNA substrate has to be prepared specifically for RACE PCR by attaching known sequences to the 5' and/ or 3' ends, following the generation of a cDNA using either random priming or gene-specific priming methods. In this project, two RACE PCR methods were used. The first was a conventional technique where

separate cDNAs were made for the 3' and the 5' reactions (see 6.3.1-2). The second was an 'all in one' method using a commercial kit known as Marathon cDNA Amplification (Clontech), as described in 6.3.3.

6.3.1. 3' RACE PCR methods

Figure 6.1. Principles of 3' RACE PCR



cDNA for 3' RACE PCR is made from total or poly A+ selected RNA using random hexamers. This is a similar method to that described in 2.14. except that the random hexamer primers used have linker tails of known sequence attached to the 5' end containing a Sal I site to enable cloning of PCR products. The reverse transcriptase used for these reactions was Superscript II (Gibco BRL).

1μg of total RNA was diluted to 10μl with DEPC water. 1μl of RNasin and 500ng of adaptor primer were added and the reaction was incubated at 65°C for 5 minutes before being cooled on ice for 5 minutes. The following reagents were then added:

2µl 10x Superscript II buffer (GibcoBRL)
2µl 25 mM MgCl2
1µl 10mM dNTP
2µl 0.01M DTT(Gibco BRL)
1µl Superscript II (Gibco BRL)

The reaction was incubated at 42°C for 1 hour before being terminated at 65°C for 15 minutes. 2µl of cDNA was then used for subsequent reactions

Sequence of adaptor primer attached to random hexamers

5'- GGCCACGCGTCGACTAGTACGAC$(T)_{17}$ -3'
Sal I

In order to amplify the cDNA containing the gene of interest, the first round of PCR was performed using a gene specific primer only i.e. no antisense primer was added.

50µl reaction	2µl of cDNA	
	250ng of the gene specific sense primer 1	
	5µl of 10 x PCR buffer (1.5mM $MgCl_2$)	
	5µl of 2mM dNTP	
	0.5µl Taq polymerase (Promega)	
Protocol: 1 cycle	94°C	5 minutes
30 cycles	94°C	1 minutes
	65°C	1.5 minutes
	72°C	2 minutes
1 cycle	72°C	10 minutes

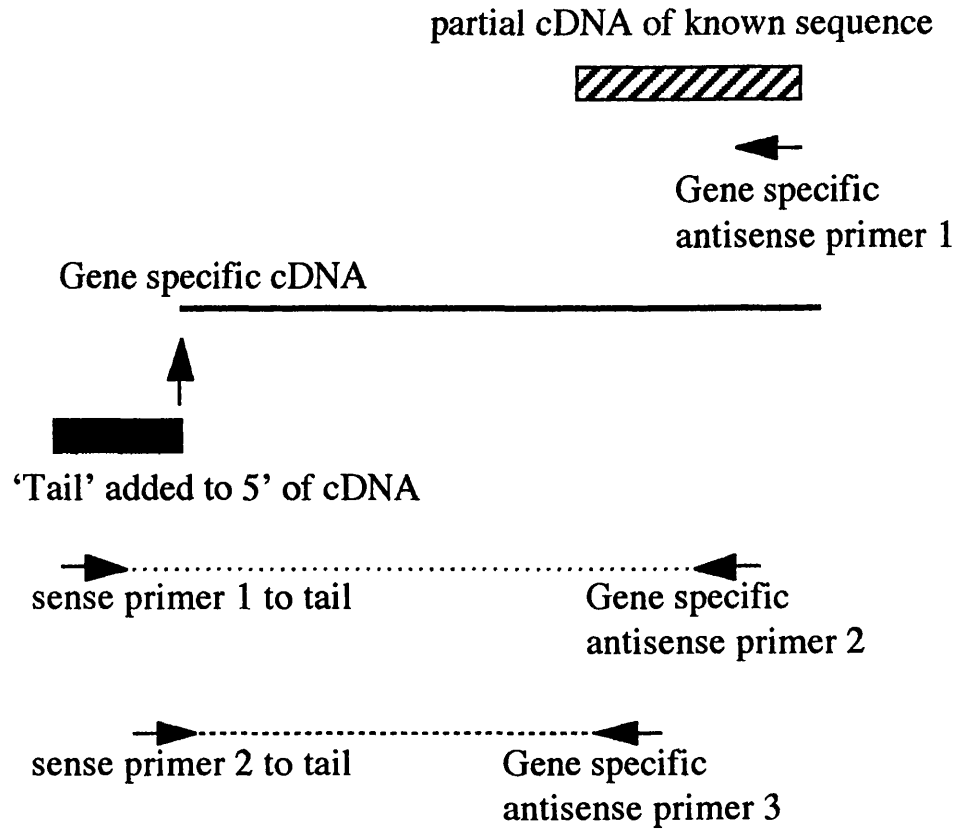
A nested second round PCR was then set up to amplify between the gene of interest and the 3' linker. The gene specific sense primer 2 had a BamH I site tagged to the 5' end of the sequence to aid the cloning of PCR products.

50µl reaction	2µl of the first round PCR
	250ng of the gene specific sense primer 2
	250ng of the primer antisense to the polylinker
	5µl of 10 x PCR buffer (1.5mM $MgCl_2$)
	5µl of 2mM dNTP
	0.5µl Taq polymerase (Promega)

The same PCR program as for round one was used. The PCR products were cleaned using a Qiagen PCR clean kit and the DNA was eluted in 50µl of sterile, distilled water. A 100µl restriction digest was then set up with the PCR product, using the enzymes BamH I and Sal I. Following 3 hours incubation at 37°C, 15µl of the digest was electrophoresed on a 1.5% agarose 0.5 x TBE ethidium bromide gel. The cDNAs amplified were of different lengths and a smear of PCR products was seen on the gel. Products between the sizes of 300-600bp were then excised from the agarose gel and purified using a Jetsorb (Genomed) kit. The final purified, size selected product was then eluted in 30µl of water and used for ligation into a M13mp18 BamH I/ Sal I vector. These ligations were then transfected into competent cells. The following day lifts were taken using nylon filters and they were then hybridised to [γ-^{32}P]dATP labelled gene specific antisense primer 1 from the 3' terminus of the cDNA sequence (fig. 6.1). After overnight hybridisation, the filters were washed as described 2.7. They were then exposed to Fuji Medical autoradiography film overnight at -70°C and positive colonies were picked the next day for sequencing. From the novel 3' sequence generated, new primers were designed to repeat the process until the end of the cDNA was reached.

6.3.2. 5' RACE PCR methods

Figure 6.2. Principles of 5' RACE PCR



For 5' RACE PCR, gene specific cDNA was made by using a gene specific primer as a template for the reverse transcriptase enzyme. A known 5' tail was then added to the cDNA so that PCR could be used to amplify between the tail and the gene of interest.

The cDNA was made using 1µg of total RNA added to 500ng of gene specific primer and 1µl RNasin, made up to a volume of 11.5µl with DEPC treated water. This was then incubated for 10 minutes at 65°C before chilling on ice for 5 minutes. The following reagents were added;

2.5µl 10 x PCR buffer
2.5µl 25mM $MgCl_2$
1µl 10mM dNTP s
2.5µl 0.01M DTT
1µl Superscript II Reverse Transcriptase (Gibco BRL).

The reaction was then incubated at 42°C for 50 minutes before being inactivated by heat at 70°C for 15 minutes. To remove the initial RNA template, 1µl of RNAse mix (Gibco BRL) was added and the reaction was incubated at 37°C for 30 minutes. The cDNA was then cleaned prior to tailing by diluting it to 1ml with sterile water and loading onto an Amicon 30 column. This was spun for 10 minutes at 5K in a SS34 rotor. The cDNA was washed with 1ml of sterile water and spun again for 20 minutes. This step was repeated twice, the final spin being for 45 minutes. The cDNA was then eluted in 40 - 60µl of sterile water following the manufacturers instructions.

Tailing: The following reagents were placed in a 1.5ml Eppendorf;

34µl of cDNA
35µl 2x tailing buffer
1µl 10mM dGTP
1µl TdT (10-20 U)

2x tailing buffer
0.4M K cacodylate
50mM Tris/HCl (pH 6.9)
4mM DTT
1mM $CoCl_2$
500µg/ml Bovine Serum Albumin

This was then incubated at 37°C for 20 minutes. The TdT was inactivated by incubation at 65°C for 15 minutes. The reaction mix was then diluted to 100µl with TE, pH 8.

5µl of cDNA was then used for the PCR in a 50µl reaction. 250ng of gene specific antisense primer 2 and 250ng of a poly C primer were used. The PCR program was the same as that used for the 3' RACE except that the annealing temperature was decreased to 60°C. A semi nested reaction was then performed. 5µl of the primary PCR product were used for the semi nested reaction. A 50µl reaction was again set up using 250ng of a poly G primer with an EcoR I site added. The same program was used as in round one. The reaction was then cleaned, size selected and cloned following the same protocol as for the 3' RACE.

6.3.3. Marathon RACE PCR methods

The method for the technique described here was obtained from the commercially available Marathon cDNA amplification kit produced by Clontech (catalog K1802-1). All components of this reaction (except the RNA) described were provided within this kit.

1µg of poly A+ RNA (Thymus RNA from Clontech) was used to make the cDNA. 1µl of synthesis primer and sterile water was added to a total volume of 5µl. The tube was incubated at 70°C for 2 minutes and then cooled on ice for a further 2 minutes. The following reagents were then added to the tube;

2µl 5X first-strand buffer
1µl dNTP mix (10 mM)
1µl sterile water
1µl MMLV reverse transcriptase

The tube was incubated at 42°C for 1 hour in a water incubator. The reaction was then stopped by placing the tube on ice. The original RNA strand was degraded and a second DNA strand was synthesised. The following were added to 10µl of the first strand reaction;

48.4µl sterile water
16µl 5X second strand buffer
1.6µl dNTP mix (10mM)
4µl second strand cocktail

The second strand cocktail includes RNase H, *Escherichia coli* DNA polymerase I and *Escherichia coli* DNA ligase. The contents of the tube were incubated at 16°C for 1.5 hours. 2µl of T4 DNA polymerase were added and the incubation was then continued for a further 45 minutes. The reaction was terminated with 4µl of EDTA/glycogen (Clontech) and the aqueous phase was extracted following mixing with an equal volume (100µl) of PCIA. Traces of phenol were removed by a second extraction with 100µl CIA. The double stranded cDNA was then precipitated with one half volume of 4M ammonium acetate and 2.5 volumes of 95 % ethanol at room temperature. The pellet was collected immediately by centrifugation at 14,000 rpm for 20 minutes. The supernatant was removed and the pellet was overlaid with 80% ethanol and spun for a further 10 minutes. The supernatant was then removed and the pellet was air dried for 10 minutes. The double stranded cDNA pellet was resuspended in 10µl of sterile water and stored at -20°C.

Adaptor ligation

A Marathon cDNA adaptor was then ligated to both ends of the cDNA to allow 5' and 3' RACE PCR amplifications. The following components were added to 5µl of the double stranded cDNA in a 0.5µl micro-centrifuge tube;

2µl Marathon cDNA adaptor (10µM stock concentration)
2µl 5X DNA ligation buffer
1µl T4 DNA ligase (1 unit/µl)

Marathon Adaptor

5'- -3'

CTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGT

Not I

The reaction was then incubated overnight at 16°C. The ligase was inactivated by heating the tube to 70°C for 10 minutes. The adaptor ligated double stranded cDNA was then diluted in Tricine-EDTA buffer (Clontech). Two dilutions were used for further reactions, 1 : 50 and 1 : 250. These dilutions were placed at 94°C for 2 minutes to denature the cDNA before being stored at -20°C. A schematic representation of the final cDNA with adaptor linkers can be seen in figure 6.3.

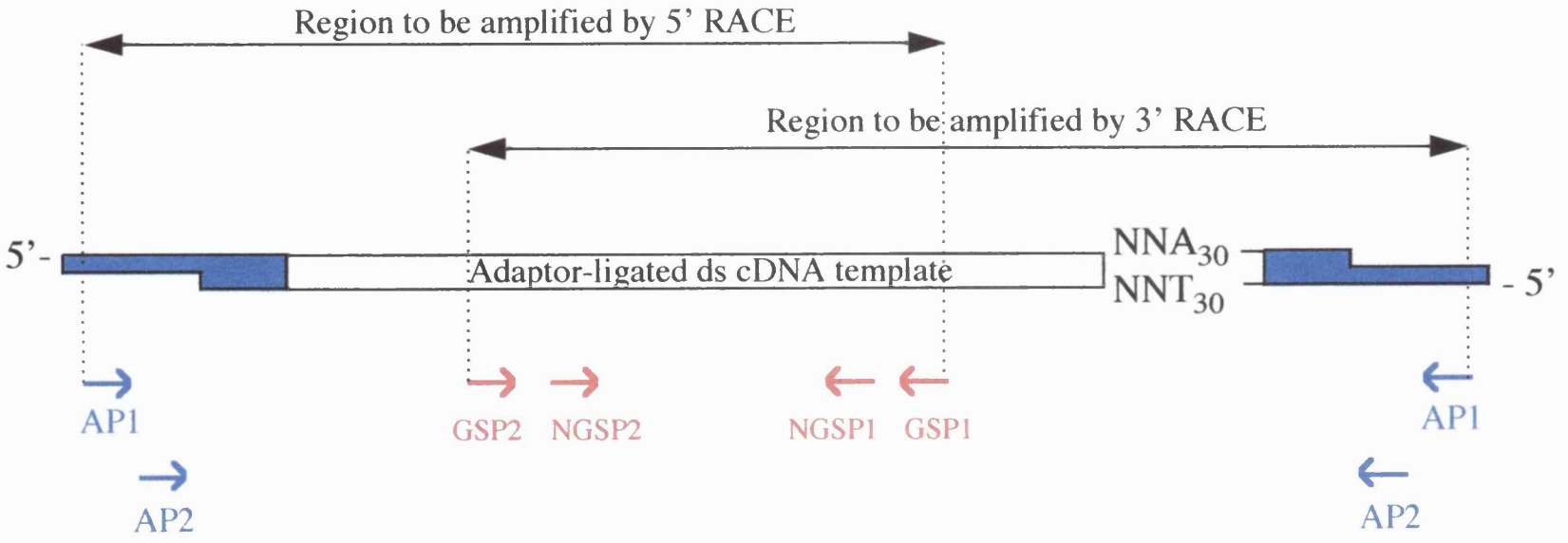
The RACE PCR reactions were then set up in the same way for both 5' and 3' reactions. GC rich primers with a T_m of over 70°C were designed from the known sequence of the cDNA to be amplified. No restriction sites were added to the primers as this reduces the efficiency of the PCR. 50µl reactions were performed using 5µl of both dilutions of the double stranded, adaptor ligated cDNA. The following reagents were then added;

36µl sterile water
5µl Klen Taq PCR buffer (catalog K1905-1)
1µl dNTP (10mM)
1µl Advantage Klen Taq Polymerase Mix (50X) (catalog K1905-1)
1µl AP1 primer (10 µM)
1µl gene specific primer (10 µM)

AP 1 primer

5'-CCATCCTAATACGACTCACTATAGGGC-3'

Figure 6.3 : The template and primers used in the Clontech Marathon RACE PCR reactions.
(adapted from the Clontech Marathon cDNA user guide)



The contents of the tube were mixed and overlaid with a drop of mineral oil. A Touchdown PCR protocol was then used for all Marathon RACE PCRs. Annealing and extension occur simultaneously in this protocol.

Touchdown PCR protocol;

	94°C	for 1 minute
5 cycles:	94°C	30 seconds
	72°C	4 minutes
5 cycles:	94°C	30 seconds
	70°C	4 minutes
25 cycles:	94°C	30 seconds
	68°C	4 minutes

Appropriate positive controls were set up using test primers that amplify the house keeping gene TFR (supplied). A negative control was also set up where no cDNA was added to the PCR reaction. 5µl of each reaction was then analysed on a 1% agarose 0.5 x TBE gel stained with ethidium bromide. If no product was seen then the PCR reaction was nested. An internal primer from the adaptor sequence (AP2) was provided in the Marathon cDNA kit.

Adaptor primer 2

5'-ACTCACTATAGGGCTCGAGCGGC-3'

Gene specific nested primers were designed in the same way as the first round primers. Again, a 50µl reaction was set up for the PCR. 5µl of the primary PCR was diluted in 245µl of Tricine/EDTA buffer and 5µl of this was used for the nested reaction. The Touchdown PCR protocol was the same, although 20 cycles were used instead of 25 in the last stage.

The RACE products were then characterised by cloning and sequencing. The total reaction was cleaned using a Qiagen PCR clean up kit and was eluted in 40µl TE. A 50µl digestion using enzymes appropriate for the restriction sites adjoining the primers was then set up. 15µl of the reaction was analysed on a 1% agarose 0.5 x TBE gel stained with ethidium bromide. The band(s) of interest were then excised and the DNA was extracted using a Jetsorb kit (Genomed). The purified, digested PCR products were then ligated into a Bluescript plasmid (KS+)

vector. Competent cells were transfected with the cloned RACE PCR products and the positive clones were identified for sequencing as for the conventional protocol.

6.4. RACE PCR results.

6.4.1. Conventional RACE PCR.

Primers from the TA 6.35 sequence were designed (see below). Initially, total RNA was isolated from peripheral leucocytes, separated from the blood of a normal donor, using the GITC method (section 2.13). Conventional RACE PCR methods were then used. The cDNA for 3' RACE was made using random hexamers, and the cDNA for the 5' RACE PCR was made using the gene specific primer - 6.35 B. The gene specific primers used to amplify the 3' reaction were 6.35 F and 6.35 2F. The gene specific primers used to amplify the 5' reaction were 6.35 2B and 6.35 5B . The positions of these primers are shown in figure 4.9.

6.35 F

5'- GGTCTGATGGCCGGGAGGACA -3'

6.35 2F

5'- TAGGGATCCGTCCTCACCATTCCTCTCCTGGACCTG -3'

6.35 B

5'- CTGGTCGTGCAACAGTGTAC -3'

6.35 2B

5'- TAGCTCGAGAGCTTCCAGTTTAGGACAGGTCCA -3'

6.35 5B

5'- TAGCTCGAGTGTCCTTCCCGGCCATCAGACC -3'

Both the 3' and 5' RACE PCR reactions were unsuccessful with total RNA purified from normal peripheral lymphocytes. As the signal on Northern blot analysis with TA 6.35 was strongest with thymus RNA, poly A+ selected human thymus mRNA (Clontech) was used for subsequent reactions. The cDNA was made using the same methods as before and the amplifications were repeated with the same primers. Although no satisfactory amplification was obtained in the 5' reactions, a good product was amplified from the 3' RACE reaction. This was size selected and cloned. When sequenced, the 3' RACE PCR product showed 100%

homology to cDNA clone 1, described in chapter 4 (sequence shown in figure 4.2). A further primer, 6.35 3F, was designed from the sequence (36bp downstream from TA 6.35) and the RACE PCR reactions were repeated. The further sequence obtained was identical to cDNA clone 1 until position 987 of clone 1, when homology was lost. This sequence was compared to the genomic data previously obtained from sequencing of genomic clones including the TA 6.35 sequence (see chapter 4). The RACE PCR clone identified was identical to the genomic DNA sequence beyond position 987. Therefore, contaminating DNA was being amplified from the thymic cDNA despite it being made from poly A selected RNA.

6.35 3F

5'- TGAAGGAAGCTGGTGAGGCAGATGTCC -3'

6.4.2. Marathon RACE PCR.

In order to extend the message from TA 6.35 in the 5' direction, 5'-Marathon RACE PCR was set up. Initially, the same primers were used as for the conventional RACE PCR except that the nested primer, 6.35 2B, was re-designed without the BamH I restriction site to improve its specificity. All the primers had melting temperatures above 70°C and the Touchdown PCR protocol was used. The results from the first round of the 5' RACE PCR were evaluated by electrophoresis. Only a smear could be seen. Unlike the conventional RACE PCR techniques, this method should give a clear band or bands representing the amplification of the complete 5' part of the gene. The reaction was nested using the internal primers and re-analysed by electrophoresis. Multiple bands could be seen between 500bp and 2kb. To ascertain which bands were correct, the gel was blotted and the nylon filter was hybridised to an internal oligonucleotide (6.35 F). A doublet at approximately 850bp was seen as a positive hybridisation signal after overnight exposure of the filter to autoradiography film. The total PCR product was electrophoresed as before and the two bands were excised. The DNA was then cleaned, blunt cloned into an M13mp18 vector and manually sequenced. Sequence analysis of first 270bp 5' of the TA 6.35 sequence showed no homology to any cDNA or genomic clones previously analysed (RACE PCR clone A in figure 6.5).

However, sequence homology analysis using the database BLAST N showed 100% homology to an EST, number AA431979, isolated from a testicular cDNA library. The homology with the EST ended with the first nucleotide of TA 6.35. The sequence of the EST contained multiple stop codons in all 6 possible reading frames. A PCR primer was designed from the 5' part of this sequence - RACE 1F (below), and a PCR reaction was set up using this primer and 6.35 B. A standard 35 cycle PCR protocol was used with an annealing temperature of 65°C. No products were seen with the thymus cDNA template or with genomic DNA under the conditions of PCR amplification used.

RACE 1 primer

5'- ATGATAGCAGTGGGATTTCAC -3'

As the link between this novel sequence and TA 6.35 was unproven, further RACE PCR experiments were set up. New primers were designed with a CG content of over 50% and, consequently, high melting temperatures. These were named 6.35 RACE B and 6.35 RACE 2B. Although these primers were designed without restriction sites to increase their specificity, 6.35 RACE 2B was also ordered with a Hind III site tagged to the 5' end to aid cloning of products. The sense nested primer, AP2, was redesigned with a restriction site for EcoR I tagged to the 5' end, also to aid cloning.

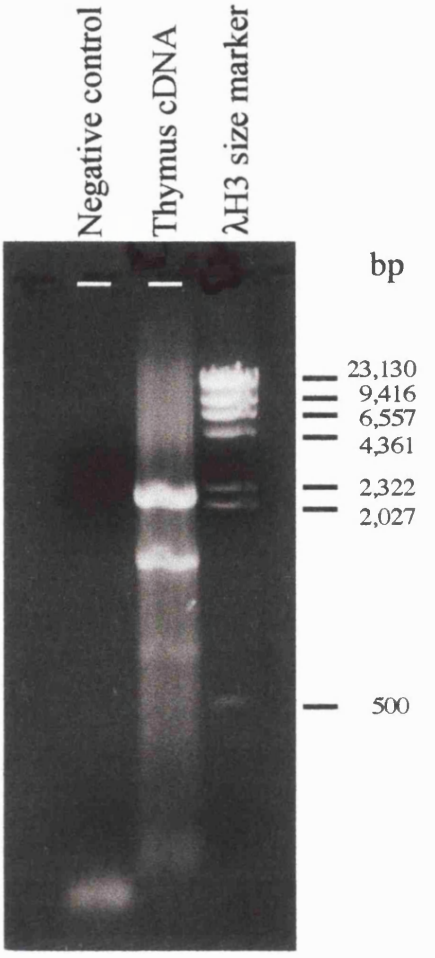
6.35 RACE B

5'- CACATGTGGTGAACACCAGGCGCCCTCCC -3'

6.35 RACE 2B

5'- GTGGGTTCCTTCTTCCTGTCCTCCCGGCC -3'

As before, no clear product was seen with the first round PCR. The reaction was nested using the internal primers 6.35 RACE 2B and AP2. Two bands were seen at 1.4kb and 2.2kb (figure 6.4A). These were purified and re amplified in a 20 cycle PCR using the same nested primers with added restriction sites. The amplified product was checked by electrophoresis and both bands were seen to have been re-amplified. The entire PCR product was then cleaned, using a Qiagen PCR clean up kit, and digested with the restriction enzymes EcoR I and Hind III. The digested DNA was then cloned into a Bluescript (KS+) vector.



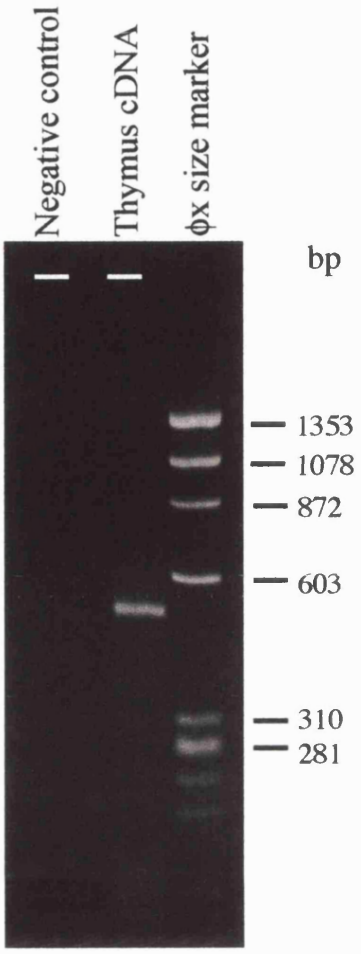


Figure 6.4A

Figure 6.4B

Figure 6.4A: Nested 5'RACE PCR product electrophoresed on a 1% agarose 0.5 x TBE gel stained with ethidium bromide. Two products can be seen at approximately 1.4kb and 2.2kb in the thymus cDNA lane.

Figure 6.4B: Nested 5' RACE PCR product electrophoresed on a 1% agarose 0.5 x TBE gel stained with ethidium bromide. One product at approximately 500bp can be seen in the thymus cDNA lane.

DNA from the cloned products was characterised by electrophoresis following digestion with EcoR I and Hind III to release the inserts. The clones all contained the smaller 1.4kb fragment. The DNA was sequenced using M13 -20 and T3 primers with an ABI 377 automated sequencer.

A contig extending 5' from TA 6.35 was compiled and the sequence was again 100% homologous to the EST AA431979. The fact that this sequence had been isolated by two different experiments was compelling evidence that it was the correct 5' transcript. However, a repeat attempt to amplify between the primer RACE 1F and TA 6.35 was again negative with thymic cDNA as a template.

A third 5' RACE PCR experiment was set up. A change was made to the protocol. It was known that the human oncogene AML 1 was not amplifiable at the cDNA level due to the complex secondary structure of its cDNA (unpublished work by Mr S. Hart from Dr Foroni's lab). Raising the denaturing temperature of the PCR to 98°C was shown to facilitate amplification of the gene by RT PCR due to a reduction in its secondary structure (unpublished data). Because of the difficulties that we had encountered trying to amplify the full cDNA associated with the TA 6.35 sequence, it was decided to adopt this approach. The 5' RACE PCR was repeated with the denaturing temperature being increased to 98°C in each cycle of the Touchdown PCR programme. As before, no product was seen with the first round PCR, but a very clean, clear band of 550bp was seen with the nested PCR (figure 6.4B). This was purified and re-amplified with the nested primers containing the restriction sites as before. The product was cloned and sequenced with an ABI 377 automated sequencer. The sequence obtained was 440bp in length (RACE PCR clone B in figure 6.5) and showed homology to 138bp of the sequence previously obtained by RACE PCR (i.e. homologous to EST AA431979) and then 92bp which was 100% homologous to part of the sequence of the original cDNA clone 1 (exon 2 in figure 6.5). The 5' 170bp of the sequence was shown to be an L1 element by homology search using the BLAST N database. The homologies of these sequences are represented schematically in figure 6.5.

The sequence that had homology to the EST AA431979 had now been isolated by 5'-RACE PCR on three occasions. The full sequence that had the homology to the EST was 249bp in length. Only the 5' 138bp had been identified in this most recent sequence (see figure 6.5). This 249bp sequence was compared to all the sequences that we had identified during this project so far and the same

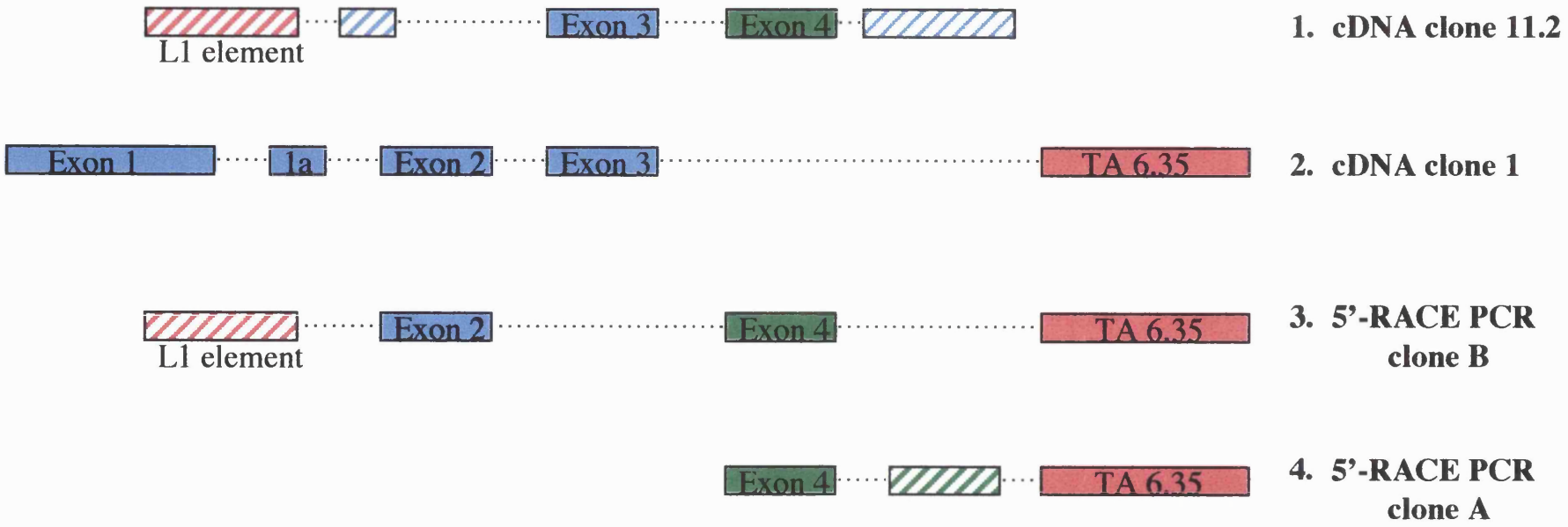


Figure 6.5 : Comparison between the cDNA clones isolated by both 5'-RACE PCR and cDNA library screening. The sequence represented by the green hatched box was later shown to be part of cDNA clone 1 and was named exon 5 (see text). The blue hatched boxes are sequences of cDNA clone 11.2 that were not identified.

138bp were found to have 100% homology to part of the cDNA clone 11.2 (described in chapter 4). This cDNA only contained homology to exon 3 of clone 1, as shown in figure 6.5. This was the first evidence of the genuine existence of cDNA clone 1 from a source other than the original peripheral leucocyte library from which the clone was isolated. These data also suggested that this novel sequence consisted of 2 exons, one of 138bp that had been isolated in cDNA clone 11.2 as well as by RACE PCR, and the other that was 111bp and had been isolated by RACE PCR only. These clones were named exons 4 and 5 (fig. 6.5).

When the cDNA library clones and RACE PCR clones were further analysed, the L1 element at the start of the 5' RACE sequence was shown have been previously identified in cDNA clone 11:2 (fig.6.5). It was unclear whether this L1 element was genuinely part of cDNA clone 1 or whether it was artifactually amplified from cDNA. Different amounts of the L1 sequence existed in the two cDNA clones and the splice junctions varied. In addition, the L1 element was spliced to a different exon in each clone i.e. 5' of exon 3 in cDNA 11.2 and 5' of exon 2 in the RACE clone.

6.5. Further experiments to verify the existence of cDNA clone 1.

The sequence called exon 2 from cDNA clone 1 had now been isolated from 2 sources of cDNA and was common to the original cDNA clone 1 and the 5' RACE PCR sequence. RACE PCR experiments from this sequence, both in the 5' and the 3' directions, were set up to confirm or refute the other contigs that had been generated so far. The cDNA already made for Marathon RACE PCR was used again for these reactions. Primers were designed for the Touchdown RACE PCR from exon 2 and are shown in table 6.3, at the end of this chapter.

As before, the nested primers were designed with and without Hind III restriction sites added to the 5' end. The Touchdown PCR protocol was set up for both the 3' and the 5' reactions. A denaturation temperature of 98°C was used again. Both reactions required nesting. Multiple bands were seen for both the 5' and the 3' PCR reactions. Instead of trying to clone individual bands as previously, the entire PCR product was cloned in this experiment. The product was re-amplified with the primers containing restriction sites before being cloned into an EcoR I / Hind III Bluescript (KS+) vector. After transfection into competent cells, colonies were randomly picked i.e. the hybridisation step was omitted, and characterised by electrophoresis following digestion with Hind III

and EcoR I. All clones with inserts (89% of 3'-RACE clones and 87% of 5'-RACE clones) were then sequenced with an ABI 377 automated sequencer using the M13 -20 and T3 primers. Many DNA contigs were generated from the sequences of these inserts, some of which contained novel sequence and others which contained known sequences previously identified. The figure 6.6 shows the contigs in schematic form that were isolated from these experiments. From now on TA 6.35 will always be referred to as exon 6.

6.5.1. 3' RACE PCR results

The sequences known as exon 3 and exon 4 were repeatedly isolated downstream of exon 2 and appeared to exist as alternative splice forms of the same cDNA. Some clones consisted of exon 2 spliced to exon 3 and then exon 6 (3' RACE clone 4 in figure 6.6). This combination of exons was first seen in the original cDNA clone 1, isolated from the peripheral leucocyte library. Other clones showed exon 2 spliced to exon 4 and then exon 6 (3' RACE clone 5 in figure 6.6). Two 3'-RACE clones were of particular interest in that both exons 3 and 4 were present forming a link between exon 2 and exon 6. The first (3' RACE clone 3 in figure 6.6) read from exon 2 into exon 3 then exon 4 and into exon 6. The second contained a 1.8kb insert (3' RACE clone 1 in figure 6.6) and read as follows;

exon 2 : exon 3 : exon 4 : exon 5: exon 6 : and was then contiguous with cDNA clone 1 until the poly A tail.

Three unknown contigs were also isolated from these experiments. Two of these were novel sequences spliced to exon 2 (3' RACE clones 6 and 7 in figure 6.6). The third was part of a clone where exons 3 and 4 were spliced downstream of exon 2 contiguously. The sequence 3' of exon 4 did not show homology to any previously identified sequence (clone 2 in figure 6.6). Antisense primers for all three novel sequences were designed and used in standard 35 cycle PCR reactions with sense primers from exon 2 (exon 2F) or exon 4 (exon 4F). No product could be amplified from cDNA or DNA with any of these primers. All the sequences isolated were sent to the BLAST database. No significant homologies to known sequences or peptides were seen with any of these RACE PCR sequences. Interestingly, none of these uncharacterised novel sequences were followed by known downstream exons that would have suggested that they were genuinely part of cDNA clone 1.

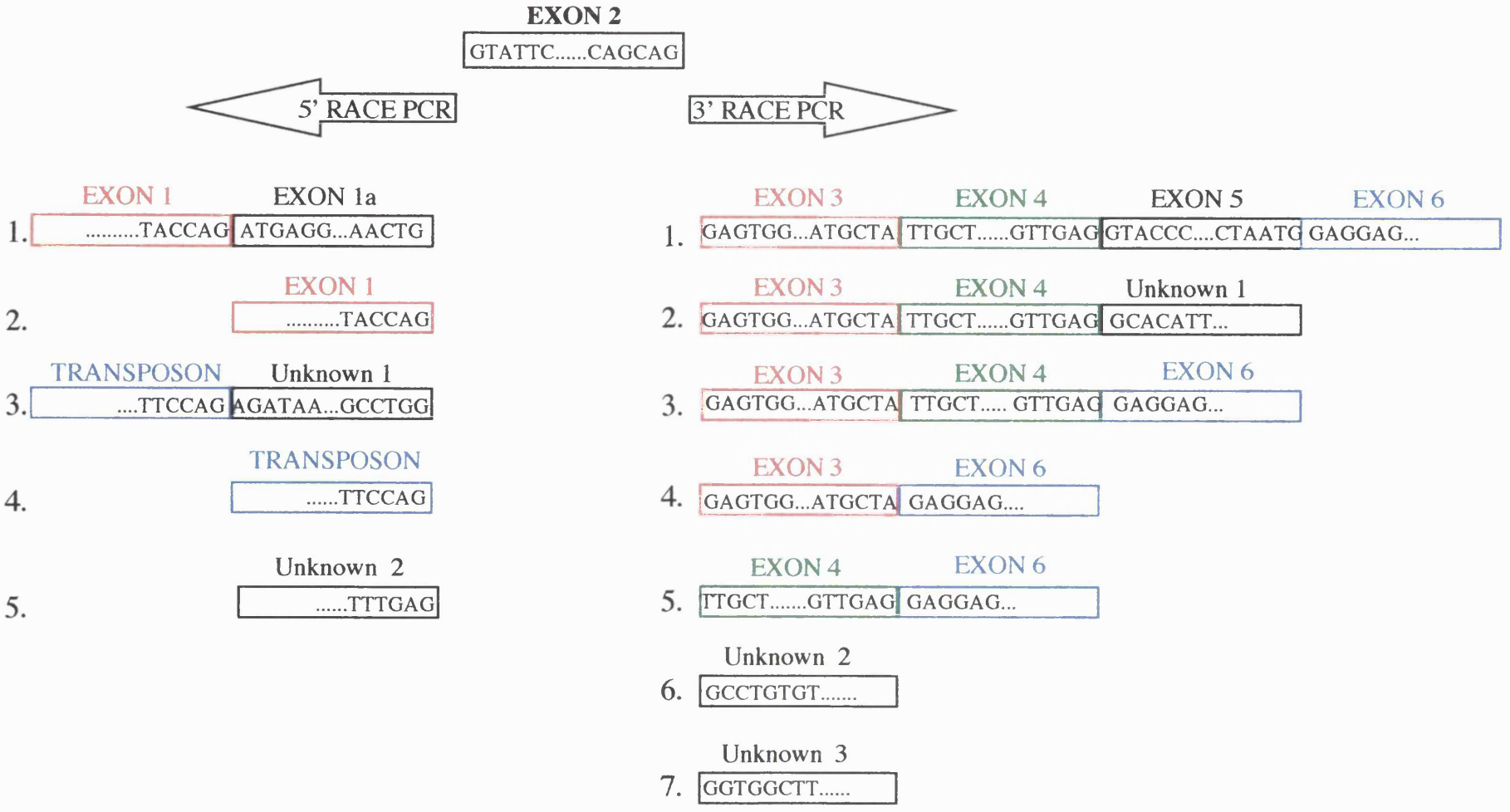


Figure 6.6: A schematic of the clones isolated from Marathon RACE PCR experiments starting from the template of exon 2.

6.5.2. 5' RACE PCR results

Exons 1 and 1a of cDNA clone 1 had also been shown to be part of a second candidate tumour suppressor cDNA, clone 2:2. This gene was published as Leu 1 (Liu *et al.*, 1997) and had raised doubts as to the validity of cDNA clone 1. Both of these exons were repeatedly identified from these 5' RACE PCR clones (5' RACE clones 1 and 2 in figure 6.6). Evidence of this link had been demonstrated before when an RT PCR reaction using primers from exons 1 and 2 of cDNA clone 1 had amplified 2 bands. These were sequenced and shown to be alternative splice versions, with and without an additional 53bp sequence known as exon 1a (chapter 4). Combined with the RACE PCR data, these provided evidence for the existence of an alternative cDNA, starting from or including exon 1 of cDNA clone 2:2.

The L1 element (transposon) previously identified by 5' RACE PCR and from cDNA clone 11:2 was isolated again by 5' RACE PCR from exon 2 (fig. 6.6). Two other novel sequences were also identified, with no known homologies when analysed by BLAST N and BLAST X databases. Sense primers from these novel sequences failed to amplify expected size fragments, when used in combination with antisense primers from exon 2, under standard PCR conditions using RT-PCR on cDNA from thymus and PCR amplification on DNA from normal bone marrow.

When these data were collated, a link between exon 1 of Leu 1 and the other exons of cDNA clone 1, including exon 6, had been proven. There appeared to be a pattern emerging of different splice versions of this cDNA. However, several unresolved questions remained. First, the relevance of the recurring L1 element sequences was unclear. Second, there was no evidence other than the positive RT PCR reaction using the primers F2 and B7, to show that this transcript was expressed either in normal or malignant lymphocytes. Third, whether all the exons of cDNA clone 1 between exon 1 and exon 6 had been isolated. Several alternative splice forms had already been identified and it was possible that others existed. Some of the sequences identified by the 5' and 3' RACE PCR experiments had also remained uncharacterised.

6.6. Expression of cDNA clone 1

6.6.1. Amplification of the extended cDNA clone 1 from normal cDNA.

Many of the informative RACE PCR sequences had been isolated following the increase of the denaturing temperature in the PCR reaction to 98°C. We postulated that this cDNA had a secondary structure which had prevented its amplification from cDNA in the past, but which could be destroyed by higher temperatures. Primers were available that had been designed across the first 7 exons of cDNA clone 1 and their positions are shown in figure 6.7. Primers derived form the sequence of clone 1 were used on normal thymus cDNA in RT-PCR reactions using a 35 cycle amplification with a denaturation step at 98°C. Our standard Taq (Promega) was unstable at 98°C and it was, therefore, replaced with Clontech Long Range Taq, previously used for the RACE PCR.

The initial combinations of primers that were used were: i) exon 2F and exon 4R and ii) exon 4F and 6.35 B (figure 6.7 and table 6.3). The annealing temperature for both reactions was 65°C. Reaction i) gave two products of approximately 320bp and 200bp. The expected product from the sequence of clone 1 was 324bp. If exon 3 was spliced out, a version of the clone that had been suggested by some of the 3'-RACE products, then the expected product would be 210bp. The total PCR products were resolved by electrophoresis and the gel was blotted. The Hybond-nylon filter was hybridised to an internal oligonucleotide from the exon 4 sequence, exon 4/2F (fig. 6.7). Both bands were positive as can be seen in figure 6.8A. This result was highly suggestive that the two bands seen on the agarose gel were the sequences expected from the predicted sequence and splice versions of cDNA clone 1. The PCR was then repeated on cDNAs made from total RNA extracted from malignant lymphocytes from patients with B-cell CLL. The same bands were seen when the PCR products were again resolved by electrophoresis.

Reaction ii) failed to give any product. However, when the PCR was repeated using a more 3' primer from the exon 4 sequence (exon 4/2F shown in fig. 6.7 and table 6.3) two bands were seen in most samples at approximately 420bp and 300bp following electrophoresis on a 2.5% agarose gel. These were correct for the predicted products amplified between exon 4 and exon 6 with exon 5 forming an alternative splice version of the cDNA. The gel was then blotted and hybridised to an internal oligo from exon 6, 6.35F. All bands were positive by

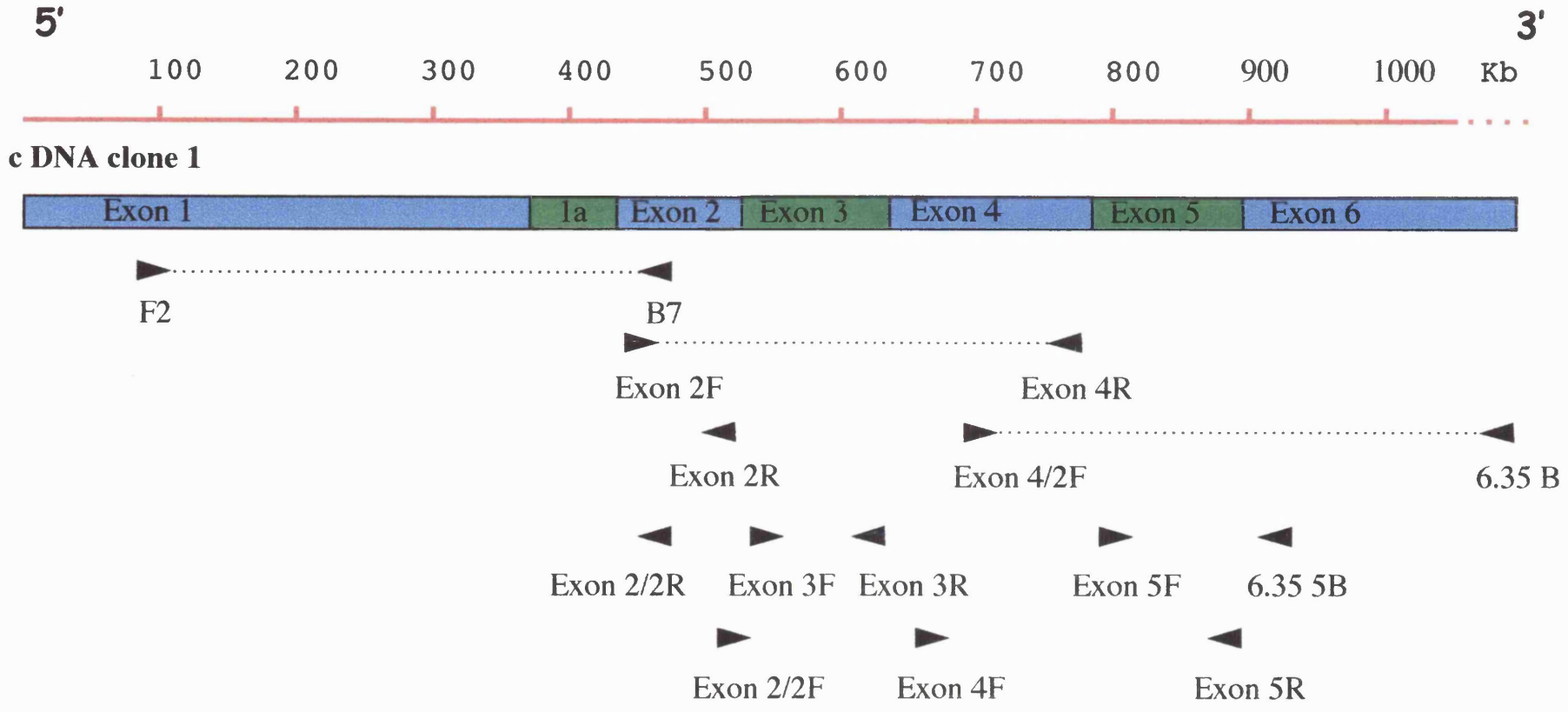


Figure 6.7 : The first 7 exons of cDNA clone 1with the positions of the primers designed from this sequence. The position of the arrow head denotes whether the primer is a sense or an antisense primer. The dotted lines join the primers that amplified the expected sequence from lymphocyte cDNA.

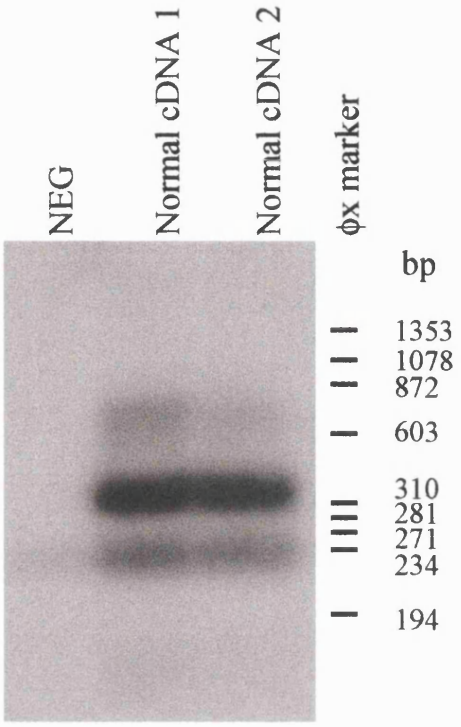


Figure 6.8. A

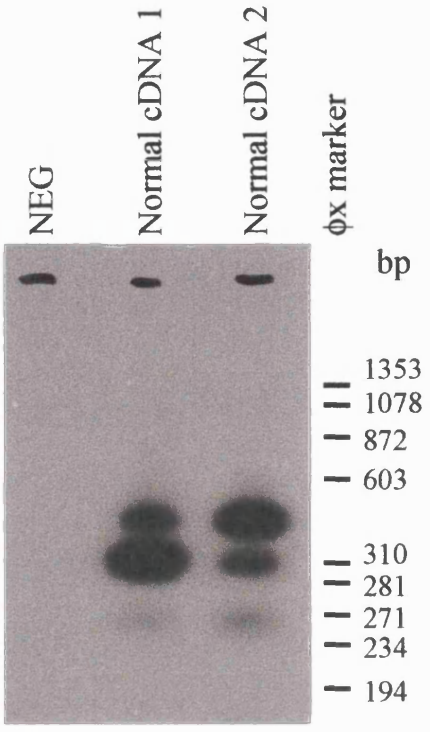


Figure 6.8. B

Figure 6.8.A.: RT PCR products amplified from normal cDNA using primers exon 2F and exon 4R.The products were electrophoresed on a 2.5% agarose gel, blotted and the filter was hybridised to the oligo exon 4/2F.The filter was washed at 50˚C in 6 x SSC/0.1% SDS and exposed to Fuji Medical autoradiography film overnight.Two bands can be seen at approximately 320 bp and 220 bp.

Figure 6.8. B: RT PCR products amplified from normal cDNA using primers exon 4/2F and 6.35R.The products were electrophoresed on a 2.5% agarose gel, blotted and the filter was hybridised to the oligo 6.35 F. The filter was washed at 50˚C in 6 x SSC / 0.1% SDS and exposed to Fuji Medical autoradiography film overnight. Two bands are seen in both samples at approximately 420bp and 300bp.

hybridisation. The reaction was then repeated on cDNAs from CLL patients lymphocytes. The same bands were seen when the products were resolved by electrophoresis.

This was evidence that this region of the sequence of the clone 1 was expressed by normal and malignant lymphocytes. Sequencing of the PCR products was required to confirm that they were amplifications of the predicted sequence.

6.6.2. Sequencing of amplified cDNA products.

Equivalent primers for exon 2F, exon 4/2F, exon 4R and 6.35 B were re-designed with 5' restriction sites to enable cloning (Hind III for the sense primers and EcoR I for the antisense). The PCR reactions described above were repeated on cDNAs from non CLL samples e.g. mononuclear cells purified from the bone marrow of patients with ALL and normal controls. The same bands as before were seen following analysis by gel electrophoresis. The entire PCR product was cleaned and cloned into a Bluescript (KS+) vector. The DNA from the cloned products was digested with Hind III and EcoR I and the insert sizes were assessed by electrophoresis. Clones with both the large and small inserts for each PCR were sequenced with the M13 -20 and T3 primers using an ABI 377 automated sequencer. All sequences matched the predicted splice versions of cDNA clone 1.

These data demonstrated that the first 7 exons of cDNA clone 1 could be amplified in alternative forms from normal and malignant lymphocyte cDNA by RT PCR. The alternative splice forms that had been identified both with RACE PCR and conventional PCR from non CLL cDNAs are depicted in figure 6.9. None of the unknown 3' sequences previously amplified with RACE PCR were re-amplified with conventional RT PCR.

6.6.3. Further Northern blot analysis of cDNA clone 1

The RT PCR reactions described were evidence of expression of the first 7 exons of cDNA clone 1. Whether cDNA clone 1 was part of the original 3.5kb lymphoid specific message identified on Northern blot analysis by the exon 6 probe was unknown. To address this, further probes were prepared to repeat some of the previous Northern blot experiments.

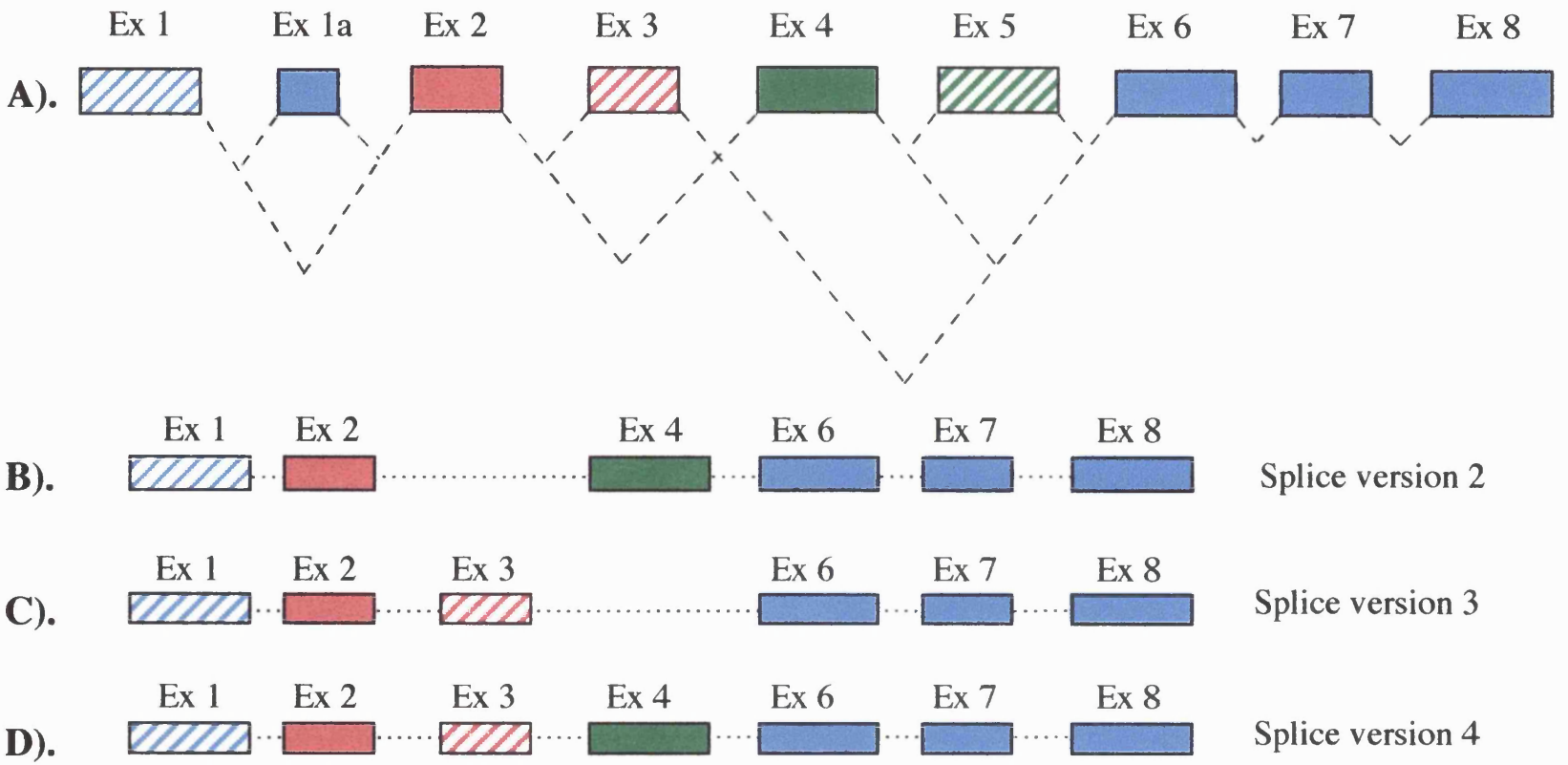


Figure 6.9 : Showing the full length cDNA clone 1 as seen in A. The schematics shown in B,C and D are the splice versions of the full length clone that were identified by 3' and 5'- RACE PCR. All versions including A could be amplified by RT PCR form lymphocyte cDNA. Exons 1 and 1a were also shown to form alternative splice forms with exon 2 of cDNA clone 2:2.

Exon 2 specific primers were designed with restriction sites to facilitate subsequent cloning (exon 2F and exon 2R, shown in figure 6.7). Following PCR amplification, the exon was cloned into a Bluescript plasmid (KS+) vector. The plasmid DNA was then prepared by the maxiprep method and the exon specific probe was purified, following electrophoresis, from an agarose gel slice. The probe was checked by hybridising it to a genomic Southern blot. It identified a 1.5kb EcoR I fragment and a 2.5kb Hind III fragment. The probe was then hybridised to a multi tissue Northern blot (Clontech). The manufacturers instructions were followed as before. The filter was hybridised to the probe for 1 hour at 65°C in Express Hyb^{TM} buffer. The final wash was at 0.2 x SSC/ 0.2% SDS at 65°C and the filter was then exposed to autoradiography film initally for 48 hours and then for 2 weeks at -20°C. No signal was detected.

A larger cDNA probe was then made from 3'-RACE PCR clone 1 (figure 6.6). Exons 2, 3 and 4 were amplified using the primers exon 2F and exon 4R. The product was cloned into a Bluescript (KS+) vector and prepared as for exon 2. The insert was used as a probe on a Southern blot. Three clean bands were seen representing each of the exons. Exon 3 identified a 7kb EcoR I fragment and a 3.8kb Hind III fragment. Exon 4 identified a 10kb EcoR I fragment and a 6kb Hind III fragment. The probe was then hybridised to the same multi tissue Northern blot (Clontech) using the same conditions of hybridisation and washing as before. No signal was detected after 2 weeks exposure to autoradiography film at -20°C.

Probes from exons 2, 3 and 4 had been hybridised to Northern blots in three separate experiments. All the results had been persistently negative. The 700bp probe including exon 1 had given a signal at approximately 1.1 kb corresponding to the message of Leu 1 (Liu *et al.*, 1997). These data suggested that the exons upstream of exon 6 in cDNA clone 1 were not part of the 3.5kb message that was so clearly identifiable with exon 6. However, the existence and expression of cDNA clone 1, including exon 6, had been demonstrated by RT PCR techniques.

6.7. Genomic organisation of cDNA clone 1

The original organisation of the exons of cDNA clone 1 can be seen in chapter 4. In order to position the new exons of cDNA clone 1 at 13q14.3, primers for amplification of exon 4 (exon 4F and exon 4R) and primers for amplification of exon 5 (exon 5F and exon 5R) were designed (see figure 6.7 and

table 6.3). Both exons were amplified from the 3'-RACE PCR plasmid clone 1 (figure 6.6), the products were then purified and hybridised to the Southern blots of both PAC DNA (described in chapter 3) and normal genomic DNA digested with EcoR I and Hind III

Exon 4 was not positive by hybridisation for any of the PAC clones. However, PCR amplification for the exon was positive for YAC 32 EH6. It was postulated that the exon lay within the gap between PAC 55 and PAC 246 (figure 6.10). This probe hybridised to a 10kb EcoR I fragment and a 5.5kb Hind III fragment with normal genomic DNA.

Exon 5 hybridised to PAC 55oe, identifying a 9.5kb EcoR I fragment and an 800bp Hind III fragment. This PAC clone also contains exon 6. Other smaller genomic clones from a cosmid library and a λ Get library containing exon 6 were available. PCR amplification with exon 5F and exon 5R were set up using DNA from 4 cosmids and 4 λ Get clones all known to contain exon 6. The cosmids were all positive for exon 5 but the phage clones were negative. The average size of the cosmid inserts was 40kb whereas the average size of the λ Get clone inserts was 19kb. This positioned exon 5 within 40kb centromeric of exon 6 but not within 19kb. See figure 6.10.

Table 6.1: Showing the size of fragments identified by each exon of cDNA clone 1 when hybridised to a genomic Southern blot of DNA digested with the restriction enzymes Hind III and EcoR I.

EXON	Hind III fragment	EcoR I fragment
1	5.5kb	8.5kb
1a	3.5kb	12kb
2	2.5kb	1.5kb
3	3.8kb	7kb
4	5.5kb	10kb
5	800bp	9.5kb
6	6kb	6kb
7	6kb	6kb
8	8kb	4.8kb

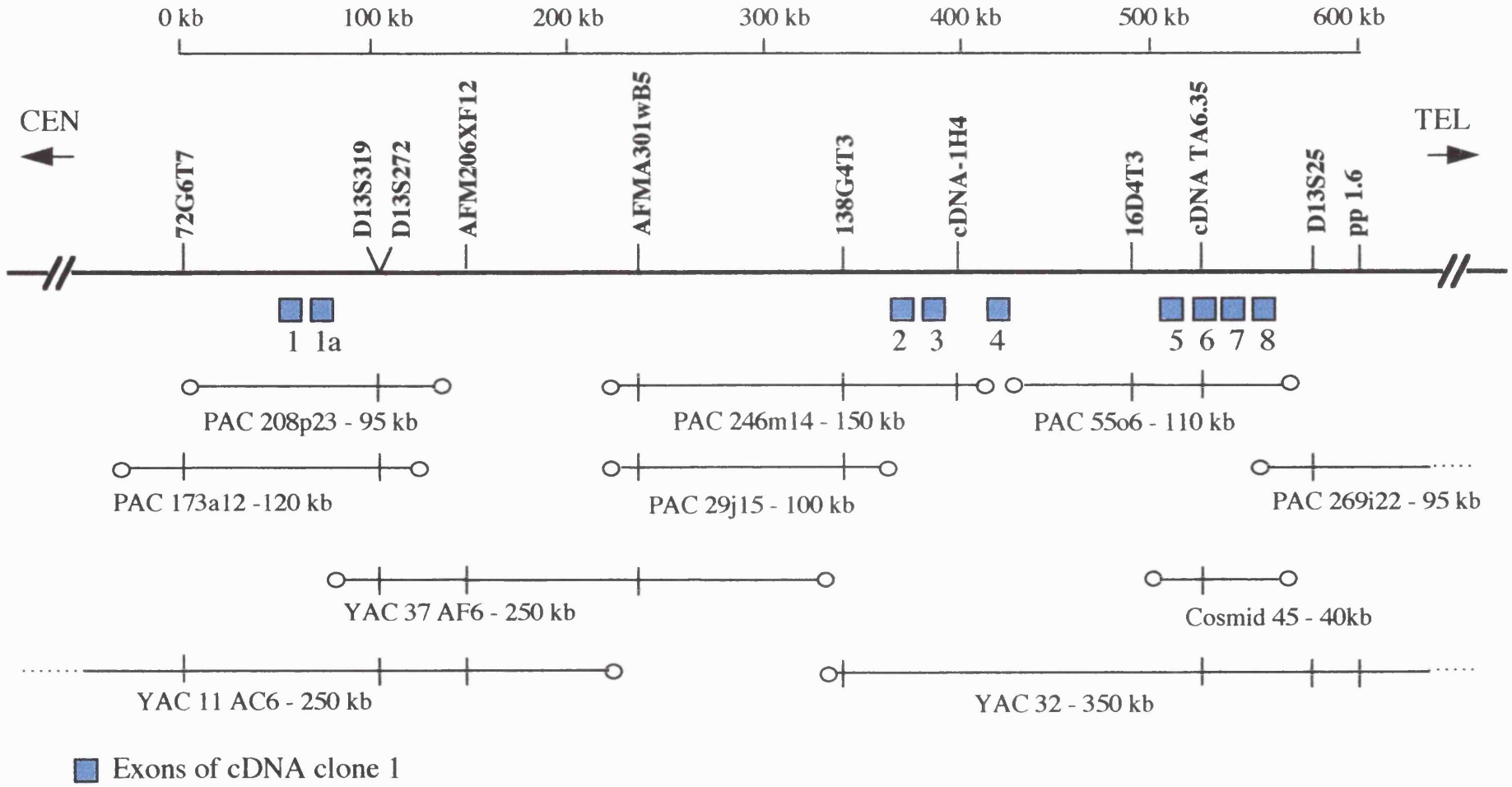


Figure 6.10: Showing positions of the exons of cDNA clone 1 on the physical map of the deleted area at 13q14.3. Note that exon 4 is situated in the gap of the PAC contig, between PACs 246m14 and 55o6.

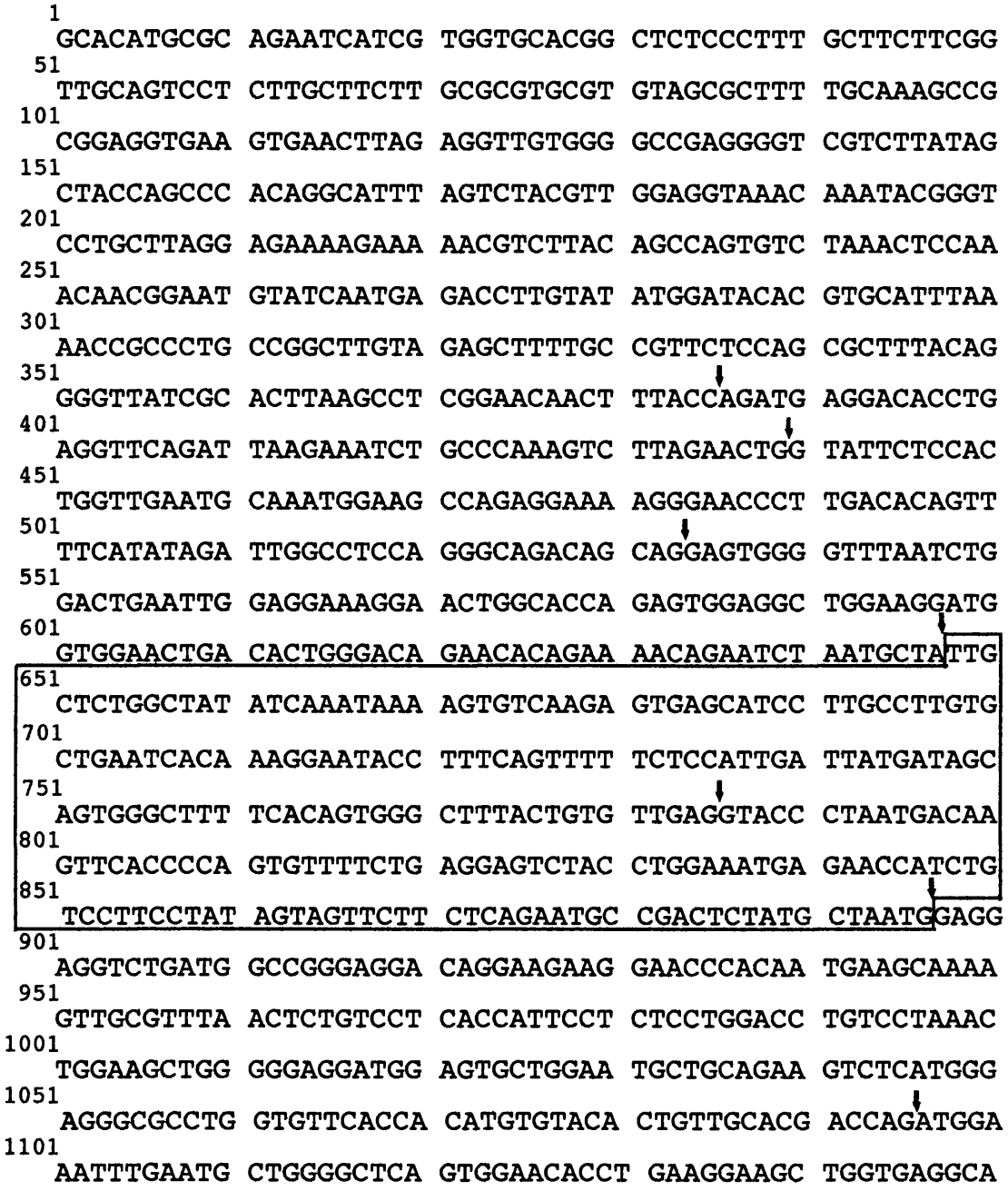
6.7.1. Species conservation of the exons of clone 1

As discussed in chapter 1, the protein coding sequences of many human genes show evolutionary conservation and are found in the DNA of other species e.g. mouse. Probes were now available for each of the exons of cDNA clone 1 and each one had been shown to hybridise to a single fragment on a human genomic Southern blot (table 6.1). Individual exon probes were labelled and hybridised to a mouse genomic Southern blot (prepared by Dr Foroni). The blots were hybridised in Church buffer at 65°C overnight. They were then washed at low stringency, 6 x SSC/ 0.1% SDS at room temperature. If after 20 minutes wash with one change of wash solution the count remained above 10 count per second as measured by a Geiger counter, the blots were washed at the same stringency at 65°C for 10 minutes. Otherwise they were exposed to autoradiography film (Fuji Medical) for a minimum of 48 hours without further washing. The blots were systematically hybridised to exons 2, 3, 4, 5 and 6. None of the probes identified a signal on the mouse genomic Southern blots. It can be concluded that these exons of cDNA clone 1 are not conserved in the mouse genome.

6.7.2. Intron-exon boundaries of the exons of cDNA clone 1.

The full sequence of cDNA clone 1 is shown in figure 6.11. Due to the different alternatively spliced forms identified for this clone, the exon - exon boundaries of the clone were clearly defined (marked with arrows in figure 6.11). The intron - exon boundaries of each exon needed to be identified from genomic clones. Various sources of genomic clones from 13q14.3 were available to us: A λ Get library of YAC 933e9 had been generated and was stored as glycerol stocks of plasmid pools as previously described. A cosmid library of YAC 22FA4 was also available in glycerol stocks and the DNA of these clones was immobilised on a series of Hybond-nylon filters. DNA from all the PACs and YACs shown in map 6.10 were also available as previously described in chapter 3.

The λ Get pools were screened by hybridisation. The λ Get library was plated out on LB agar plates supplemented with ampicillin. There were 80 pools in total containing 3 plasmid (converted phage clones) clones per pool. Filter lifts were taken from these colonies and hybridised to probes for the individual exons. Positive pools were then diluted, re-plated and screened in order to isolate the individual positive clones.


1
GCACATGCGC AGAATCATCG TGGTGCACGG CTCTCCCTTT GCTTCTTCGG
51
TTGCAGTCCT CTTGCTTCTT GCGCGTGCGT GTAGCGCTTT TGCAAAGCCG
101
CGGAGGTGAA GTGAACTTAG AGGTTGTGGG GCCGAGGGGT CGTCTTATAG
151
CTACCAGCCC ACAGGCATTT AGTCTACGTT GGAGGTAAAC AAATACGGGT
201
CCTGCTTAGG AGAAAAGAAA AACGTCTTAC AGCCAGTGTC TAAACTCCAA
251
ACAACGGAAT GTATCAATGA GACCTTGTAT ATGGATACAC GTGCATTTAA
301
AACCGCCCTG CCGGCTTGTA GAGCTTTTGC CGTTCTCCAG CGCTTTACAG
351
GGGTTATCGC ACTTAAGCCT CGGAACAACT TTACCAGATG AGGACACCTG
401
AGGTTCAGAT TAAGAAATCT GCCCAAAGTC TTAGAACTGG TATTCTCCAC
451
TGGTTGAATG CAAATGGAAG CCAGAGGAAA AGGGAACCCT TGACACAGTT
501
TTCATATAGA TTGGCCTCCA GGGCAGACAG CAGGAGTGGG GTTTAATCTG
551
GACTGAATTG GAGGAAAGGA ACTGGCACCA GAGTGGAGGC TGGAAGGATG
601
GTGGAACTGA CACTGGGACA GAACACAGAA AACAGAATCT AATGCTATTG
651
CTCTGGCTAT ATCAAATAAA AGTGTCAAGA GTGAGCATCC TTGCCTTGTG
701
CTGAATCACA AAGGAATACC TTTCAGTTTT TCTCCATTGA TTATGATAGC
751
AGTGGGCTTT TCACAGTGGG CTTTACTGTG TTGAGGTACC CTAATGACAA
801
GTTCACCCCA GTGTTTTCTG AGGAGTCTAC CTGGAAATGA GAACCATCTG
851
TCCTTCCTAT AGTAGTTCTT CTCAGAATGC CGACTCTATG CTAATGGAGG
901
AGGTCTGATG GCCGGGAGGA CAGGAAGAAG GAACCCACAA TGAAGCAAAA
951
GTTGCGTTTA ACTCTGTCCT CACCATTCCT CTCCTGGACC TGTCCTAAAC
1001
TGGAAGCTGG GGGAGGATGG AGTGCTGGAA TGCTGCAGAA GTCTCATGGG
1051
AGGGCGCCTG GTGTTCACCA CATGTGTACA CTGTTGCACG ACCAGATGGA
1101
AATTTGAATG CTGGGGCTCA GTGGAACACCT GAAGGAAGC TGGTGAGGCA

```
1151
   GATGTCCCGC TATGCCTTCC AAGTGTCATG CAACCAGTTC ATGCCTGTCA
1201                          ↓
   GCATCTCTGC CGGTGGACTC TTGAACCTTC AGAAAGGGCA GAAAGGAAGT
1251
   TTACTTCTGC CTCTACAATG GATTTGCCTA CTTTAGACAT GTTTTCAAGA
1300
   TTTATCCACG TCATGGCATA TACCAGAACT TCATTTCTTT TATGATTGAA
1351
   TAATATTACA TGGTTTTAAT ATACTATATT TTGTTTATCT GTTCATCAAT
1401
   TAATAAATCT TTGGGTTGTT TTCACTTTTA GGCTATAATA ATTAACGCTG
1451
   TTAGGAATAT TTACTTGTAA GTTTTTGTTA TATGTTTTTA ATTCTCCTGG
1501
   GTAAGATACT TAGGGGATAA AAATTGCTGA ATTAGATTGT AACTTTATGT
1551
   TTAGAAACTG CCCAACTGTT TTCCAAAGTA ACTATCCATT TTAATATTTC
1601
   ACCAGCATTG TATGACAGTT CTGGTCTTTC CTTATTTAAA CTAATACTTA
1651
   ATTGTTGCCT TTTATTGTAG CTATCCTAGT AGATATGAAG TGTTATCTCA
1701
   TTATGGTTTT GATTTGCATT CTCTAATGAC TAATAATATC AAGCATCTTT
1751
   TCATGTGTAT ACAAGAATAT TTGTATATCT TCTTTAGAGA AATGTCTACT
1801
   CAAATCCTGT GCCTATGTTT TAATTGGATA ATTTATGTTT TTATTGTTGA
1851
   GTTGTAATAG TTTTTAAATA TATTCTGGAT CCTAGACCTT TAACAGCACA
1901
   TGATTTGTAA ATATTTCTCC CATTCTGTGG GTTGGCTTTT CACTTCTTTA
1951
   TATTATCTTT TGAAGCACAA GTTTTTTATT TTGATAAAGT CCAGTTTATA
2001
   TAAAAAA AAAAAAAAAA AAA
```

Figure 6.11 : Showing the full sequence of the complete cDNA clone 1. The arrows mark the exon-exon boundaries. The boxed sequence is the novel sequence identified by RACE PCR and cDNA library screening (known as exons 4 and 5). The two signals for polyadenylation are underlined.

Two positive clones were identified for exon 1. By using primers already available from exon 1 (F2 and B4), sequencing reactions were set up directly on DNA of these clones. Comparison of the λ Get clone sequences with the cDNA sequence allowed the intron - exon boundaries of exon 1 to be clearly defined (table 6.2). Primers were then designed from the intronic sequence for amplification of the exon from DNA by PCR (table 6.3).

When the λ Get clones were screened with exons 2 and 3 no positive clones were successfully isolated. Sequencing directly from PAC 246m14 was tried using internal primers from the exons (exon 2F and 2R and exon 3F and 3R) with an ABI 377 automated sequencer. The DNA was prepared with a Qiagen maxiprep kit from a 250ml overnight culture of the PAC clone in LB supplemented with kanamycin. The sequencing protocol was adjusted as recommended by the manufacturers. The starting DNA was increased to 800ng and the amount of enzyme added was increased from 1μl to 8μl. However, the quality of sequence obtained was unsatisfactory.

A λ Get clone was isolated containing exon 4. DNA from the clone was prepared as before and an aliquot of the DNA was digested with EcoR I and Hind III in separate reactions. These were then electrophoresed and the gel was blotted overnight. The filter was hybridised to the exon 4 PCR probe. A 5.5kb band was seen with the Hind III digest and a 10kb band with the EcoR I digest as expected. A further 3μg of the phage clone DNA was then digested with Hind III and resolved by electrophoresis on a 1% agarose gel. The 5.5kb band was excised, purified and ligated into a Hind III prepared Bluescript (KS+) plasmid vector. Sequencing reactions were set up with the internal primers to exon 4 (exon 4F and exon 4R) and the sequences of both intron - exon boundaries were obtained (table 6.2). Intronic primers were designed to amplify this exon from genomic DNA (table 6.3).

Four cosmids from the YAC 22FA4 library, positive for hybridisation to exon 5, had previously been identified. Using primers designed from the exon 5 sequence (exon 5F and exon 5R), sequencing reactions for automated sequencing were set up on these cosmids. The cosmid DNA was prepared by the plasmid maxiprep method. The sequencing protocol for plasmids was modified in that 600ng of DNA was used per reaction and 4μl of enzyme was used. No satisfactory sequence was obtained. As no λ Get clones were consistently positive by hybridisation for this exon, the boundaries of exon 5 remain to be characterised.

The identification and characterisation of the intron-exon boundaries for exon 6 have already been described in chapter 4.

Exon 7 was contained in its entirety within the same 6 kb EcoR I genomic fragment as exon 6. The cloning and sequencing of this fragment was described in chapter 4. By comparing the sequence of the genomic DNA and cDNA clone 1, the boundaries of this exon were identified.

When cDNA clone 1 was hybridised to the bacteriophage clone that contained the 6kb EcoR I fragment with exons 6 and 7 (from the SHAA library described in chapter 4), a second 4.8kb EcoR I band was also identified. This band was postulated to be that containing exon 8 of cDNA clone 1. When the same genomic clone was digested with the restriction enzyme BamH I, two bands were again seen at 5.5kb and 4.8kb. Exon 6 was shown to hybridise to the 5.5kb band. The 4.8kb band was subcloned into a BamH I prepared Bluescript plasmid vector and the ends of the insert were sequenced using the M13 T7 and T3 primers. The sequence of exon 8 was found to be identical to the sequence obtained from the T3 primer, starting from a BamH I site known to exist beyond the poly A addition site of cDNA clone 1. By comparing the genomic and cDNA sequences, the 5' intron - exon border of exon 8 was identified.

As yet we have been unable to characterise the intron-exon boundaries of exons 1a, 2, 3 and 5.

Table 6.2: Showing the 5' and 3' intron-exon boundaries of exons 1, 4, 6, 7 and 8.

Intron	5' boundary		3' boundary	Intron
cccgccag	GCACATGC	EXON 1	TTTACCAG	gtaggaac
ttgtctag	TTGCTC	EXON 4	GTGTTGAG	gtaagttc
gtttatag	GAGGAG	EXON 6	ACGACCAG	gttggca
tcccacag	ATGGAAA	EXON 7	GGACTCT	gtgagtca
tatttcag	TGAACCT	EXON 8	contiguous	

Primer	Sequence
Exon 2F	CTCCACTGGTTGAATGCAAATGGAAGCC
Exon 2/2F	ATTGGCCTCCAGGGCAGACAGCAG
Exon 2R	CTGCTGTCTGCCCTGGAGGCCAAT
Exon 2/2R	GGCTTCCATTTGCATTCAACCAGTGGAG
Exon 3F	AATCTGGACTGAATTGGAGGAAAGGAAC
Exon 3R	GATTCTGTTTTCTGTGTTCTGTCCCAGTG
Exon 4F	TTGCTCTGGCTATATCAAATAAAAGTG
Exon 4/2F	GAGTGAGCATCCTTGCCTTGTGCTGAATC
Exon 4R	CACTGTGAAAAGCCCACTGCTATC
Exon 5F	GTACCCTAATGACAAGTTCACCCCAGTG
Exon 5R	GCATAGAGTCGGCATTCTGAGAAGAAC
6.35 5B	TAGCTCGAGTGTCCTTCCCGGCCATCAGACC
Exon 1 intron F	GCTGCCTCCACAGCTGTCAATACCGC
Exon 1 intron R	CGTCTCTGAGCCTAGGTTTCCTGGTC
Exon 4 intron F	CTGCAAACATAGGCAATTGTAC
Exon 4 intron R	CTCAACAAAATAGGTACAGAAGG
Exon 7 intron F	CCAAAAAGTAGTTAATCTGAGCACACATC
Exon 7 intron R	GTGGCTCTCTGCTCCTGTAGAT
Exon 8 intron F	TGGCTCCCATGATTATGCATAT
Exon 8 R	TCCCTAAGTGTCTGCCCAGGAG

Table 6.3: Primers for RACE PCR and mutational analysis

6.8. Assessment of open reading frames of varying splice versions of cDNA clone 1.

When the sequence of each exon had been verified, the possible open reading frames for various combinations of splice versions were analysed. The software programme used was DNAStar for Macintosh.

The ATG postulated to be the start of the open reading frame for the gene Leu 1 (Liu *et al.*, 1997) was contained within exon 1 at position 267. Exon 1a had no in-frame sequence. Exon 2 had one in-frame sequence as did exon 3. However, when 2 was spliced to exon 3, a stop codon was introduced to the in frame sequence of exon 2 early in exon 3. Exons 4 and 5 had stop codons in all possible frames. Exon 6 had one in-frame sequence for its 199 base pairs. In addition, an open reading frame existed from position 121 of exon 6, continuing through exon 7 and stopping in exon 8.

Clearly there is no splice version of cDNA clone 1 which has a long open reading frame. The two longest postulated open reading frames, starting with a methionine (ATG), are shown in figure 6.12. The first can be identified when exon 2 is spliced to exon 4. The open reading frame starts at position 267 in exon 1 and continues through exon 2, stopping in exon 4, 234 base pairs later. This transcript would encode for a 78 amino acid peptide but only exists when exons 1a and 3 are spliced out. The other open reading frame starts at position 121 of exon 6 and continues for 297 base pairs through exons 7 and 8. This potentially encodes for 99 amino acids. This open reading frame exists in all possible spliced versions of this clone identified so far.

6.9. Mutational analysis of cDNA clone 1 in patients with B-cell CLL

Five of the patients shown to have heterozygous deletion of the region at 13q14.3 were chosen for mutational analysis of cDNA clone 1. The patients' reference numbers were 14, 33, 54, 48 and 76. The data showing their heterozygous status at 13q14.3 can be seen in chapter 7. cDNA and DNA purified from malignant lymphocytes were available for each patient. Unfortunately, only one of these patients (14) was included in the group assessed for mutations of cDNA clone 2:2, chapter 5. Lack of material dictated the change in the group of patients assessed for this cDNA clone 1.

```
   1
     GCACATGCGC AGAATCATCG TGGTGCACGG CTCTCCCTTT GCTTCTTCGG
  51
     TTGCAGTCCT CTTGCTTCTT GCGCGTGCGT GTAGCGCTTT TGCAAAGCCG
 10
     CGGAGGTGAA GTGAACTTAG AGGTTGTGGG GCCGAGGGGT CGTCTTATAG
 151
     CTACCAGCCC ACAGGCATTT AGTCTACGTT GGAGGTAAAC AAATACGGGT
 201
     CCTGCTTAGG AGAAAAGAAA AACGTCTTAC AGCCAGTGTC TAAACTCCAA
 251
     ACAACGGAAT GTATCAATGA GACCTTGTAT ATGGATACAC GTGCATTTAA
 301                   M  R    P  C  I    W  I  H    V  H  L  K
     AACCGCCCTG CCGGCTTGTA GAGCTTTTGC CGTTCTCCAG CGCTTTACAG
 351   P  P  C    R  L  V    E  L  L  P    F  S  S    A  L  Q
     GGGTTATCGC ACTTAAGCCT CGGAACAACT TTACCAGGTA TTCTCCACTG
 401 G  L  S  H    L  S  L    G  T  T    L  P  G  I    L  H  W
     GTTGAATGCA AATGGAAGCC AGAGGAAAAG GGAACCCTTG ACACAGTTTT
 451  L  N  A    N  G  S  Q    R  K  R    E  P  L    T  Q  F  S
     CATATAGATT GGCCTCCAGG GCAGACAGCA GTTGCTCTGG CTATATCAAA
 501   Y  R  L    A  S  R    A  D  S  S    C  S  G    Y  I  K
     TAAAAGTGTC AAGAGTGAGC ATCCTTGCCT TGTGCTGAAT CACAAAGGAA
 551 *
     TACCTTTCAG TTTTTCTCCA TTGATTATGA TAGCAGTGGG CTTTTCACAG
 601
     TGGGCTTTAC TGTGTTGAGG AGGAGGTCTG ATGGCCGGGA GGACAGGAAG
 651
     AAGGAACCCA CAATGAAGCA AAAGTTGCGT TTAACTCTGT CCTCACCATT
 701
     CCTCTCCTGG ACCTGTCCTA AACTGGAAGC TGGGGGAGGA TGGAGTGCTG
 751
     GAATGCTGCA GAAGTCTCAT GGGAGGGCGC CTGGTGTTCA CCACATGTGT
 801                       M    G  G  R    L  V  F  T    T  C  V
     ACACTGTTGC ACGACCAGAT GGAAATTTGA ATGCTGGGGC TCAGTGGAAC
 851  H  C  C    T  T  R  W    K  F  E    C  W  G    S  V  E  H
     ACCTGAAGGA AGCTGGTGAG GCAGATGTCC CGCTATGCCT TCCAAGTGTC
 901   L  K  E    A  G  E    A  D  V  P    L  C  L    P  S  V
     ATGCAACCAG TTCATGCCTG TCAGCATCTC TGCCGGTGGA CTCTTGAACC
 951 M  Q  P  V    H  A  C    Q  H  L    C  R  W  T    L  E  P
     TTCAGAAAGG GCAGAAAGGA AGTTTACTTC TGCCTCTACA ATGGATTTGC
1001  S  E  R    A  E  R  K    F  T  S    A  S  T    M  D  L  P
     CTACTTTAGA CATGTTTTCA AGATTTATCC ACGTCATGGC ATATACCAGA
1051   T  L  D    M  F  S    R  F  I  H    V  M  A    Y  T  R
     ACTTCATTTC TTTTATGATT GAATAATATT ACATGGTTTT AATATACTAT
     T  S  F  L    L  *
```

Figure 6.12: Showing the first 1100 bp of the cDNA clone 1 version 2, as shown in figure 6.6.The two longest postulated open reading frames are shown starting at positions 267 and 769 respectively.

The intron - exon boundaries were available for exons 1, 4, 6, 7 and 8. These exons were amplified from DNA using intronic primers (table 6.3). Because the intron between exons 6 and 7 was so small (80bp), these two exons were amplified together (see figure 6.13). As exon 8 does not have a 3' boundary, an antisense primer (exon 8R) was designed from the sequence after the poly A addition site (table 6.3).

The other exons were sequenced by amplifying stretches of cDNA clone 1 from the patient's cDNA by RT PCR. All cDNAs were amplified with i) exon 2F - exon 4R, and ii) exon 4/ 2F - 6.35 B. This combination of primers amplified from the beginning of exon 2 to the end of exon 6. As described before, 2 bands were seen in both these RT PCRs representing the two splice versions for each amplification. The larger versions were sequenced in each case so that the sequence of exons 3 and 5 were included in the amplified product. The combinations of primers used to sequence the entire clone within these patients is shown in figure 6.13. The sequences of each of the primers used are shown in table 6.3. All primers included 5' restriction sites to allow cloning of the products (not shown).

When looking for mutations in a gene potentially related to disease, it is important to assess the integrity of both splice sites for each exon as well as the potential coding sequences. Mutations within splice sites frequently cause disease (Vogelstein and Kinzler, 1998), as discussed in chapter 1. The integrity of the splice sites of cDNA clone 1 were assessed with the combination of PCR amplifications shown in figure 6.13. Where the exons were sequenced from DNA, the splice site could be analysed directly. For alternative splicing to occur the splice sites must be intact. Therefore, for those exons that were only sequenced from RT PCR reactions, the integrity of the splice sites was assumed by the presence of expected splice versions at the cDNA level, as assessed by electrophoresis of the RT PCR product.

The products from these PCR reactions were cloned and sequenced. At least 3 clones from each amplification were sequenced in both directions for each patient. This was to allow for PCR error and amplification from contaminating normal cells. Where base changes were discovered in certain clones, more clones for that patient were sequenced. An automated ABI 377 sequencer was used for this part of the project. As none of the products being sequenced were more than 400bp in size, it was possible to obtain double stranded sequencing of each clone with the M13 -20 and T3 primers.

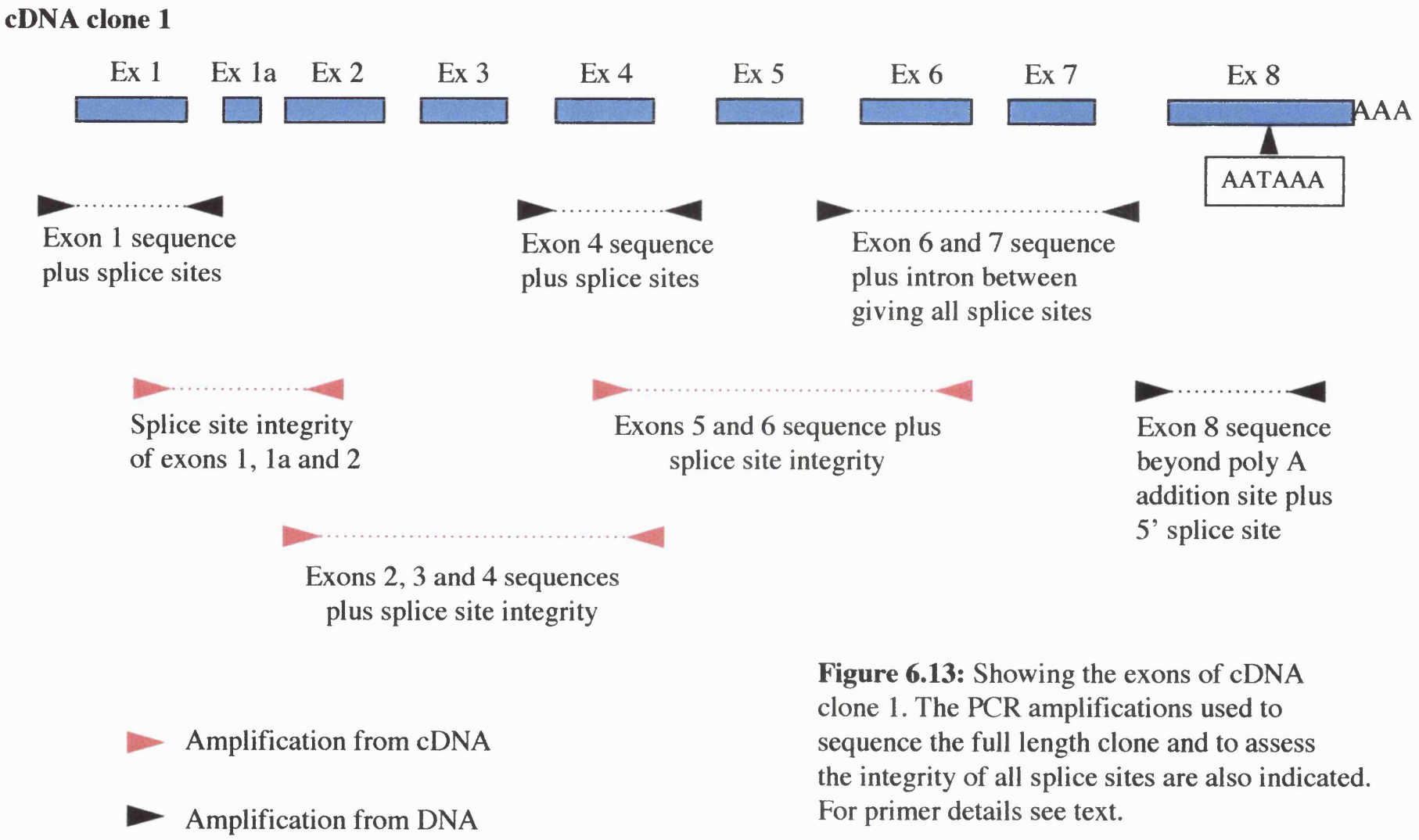


Figure 6.13: Showing the exons of cDNA clone 1. The PCR amplifications used to sequence the full length clone and to assess the integrity of all splice sites are also indicated. For primer details see text.

Of the five patients analysed, patient 14 had a base change at position 276 of exon 1 where an A was substituted for a T. This base was situated 9bp after the ATG triplet postulated to be the start of the open reading frame of the Leu 1 gene (Liu *et al.*, 1997) . This point mutation would be silent i.e. the proline remains unchanged. Patient 33 had a base change at position 81 of exon 4 in approximately 75% of clones sequenced. The base change substituted a C for a T. In the remaining 25% of clones in this patient, there was no base change. This base change was outside the coding regions postulated in section 6.8. No other mutations were found in the remaining exons for these two patients or any of the exons in the other three patients. In addition, all the expected alternative splice versions were demonstrated in each patient using the amplifications shown in figure 6.13. The significance of the two point mutations found is not known.

6.10. Conclusions.

6.10.1. The existence of cDNA clone 1 as shown by RACE PCR

This chapter describes how the existence of cDNA clone 1 was proven using RACE PCR techniques. The original cDNA clone 1 was isolated from a peripheral leucocyte cDNA library following hybridisation to TA 6.35. However, despite repeated screenings of libraries from other lymphoid tissues, no corresponding cDNAs were isolated again. 5'-RACE PCR from the TA 6.35 sequence not only confirmed the link between TA 6.35 and the upstream exons of clone 1, but also identified two novel exons that were also shown to be part of the transcript. Exon 4 had previously been isolated from a peripheral leucocyte cDNA library (cDNA clone 11.2). Exons 4 and 5 were shown to have 100% homology to 157bp of an EST (no. AA431979). Following the completion of this work, a further EST (no. A1337241) was deposited in the BLAST N database which has 100% homology to exons 2, 3 and 4 of cDNA clone 1. Interestingly, this EST also contains the L1 element that had been identified both in cDNA clone 11:2 and using 5' RACE PCR (figure 6.5).

RACE PCR became an increasingly powerful tool in this project as experience with the technique was gained. Using poly A selected mRNA rather than total RNA decreased the numbers of spurious products that were amplified. Marathon RACE PCR was more efficient and reproducible than the conventional techniques initially used. Adding 5' restriction sites to the nested primers greatly

increased the cloning efficiencies of the products generated. A lot of time was wasted with the early Marathon RACE PCR experiments, trying to identify the correct band to clone from the nested reactions. The return on the experiments improved when this procedure was changed and the total PCR product was cloned. With the acquisition of an automated sequencer many more clones could easily be sequenced. Only by continuing to isolate and sequence clones in this way did the pattern of splice versions of cDNA clone 1 become clear. Any inserts that were not flanked by the nested primer sequences were ignored. Finally, increasing the denaturing temperature to 98°C in the RACE PCR reactions increased the yield of products.

The final clone identified consisted of 9 exons which were spliced together to form a polyadenylated mRNA. Physical mapping of these exons showed that they were all positioned at 13q14.3 spanning a genomic region of more than 450 kb. Consequently, 5' exons lay within the minimal area of deletion described by Corcoran *et al.* (Corcoran *et al.*, 1998) and Kalachikov *et al.* (Kalachikov *et al.*, 1997) whilst 3' exons lay within the minimal region of deletion described by Bullrich *et al.* (Bullrich *et al.*, 1996). See figure 6.14.

6.10.2. Expression of cDNA clone 1

Increasing the denaturing temperature to 98°C and changing to Clontech Long Range Taq enabled amplification of sections of cDNA clone 1 by RT PCR. This demonstrated expression of the first 6 exons of cDNA clone 1 in both normal and malignant lymphocytes. Expression can also be assumed in the testis as an EST (no. AA431979) with 100% homology to exons 4 and 5 was isolated from a testicular cDNA library.

However, the lymphoid specific 3.5kb message identified by the original TA 6.35 (exon 6) sequence was not identified by any of the 5' exons i.e. exons 1 to 4. Although the mRNA that these exons belong to was detected by RT PCR, it was not sufficiently abundant to be detectable by Northern blot hybridisation. This is not uncommon, many genes have mRNAs that fall below the level of detection of this technique e.g. Kallmann gene (Legouis *et al.*, 1991). RT PCR is 1000 fold more sensitive than Northern blot analysis and will detect low abundance messages.

It has to be concluded that these exons are not part of the relatively abundant 3.5kb message described in chapter 4. The TA 6.35 (exon 6) sequence

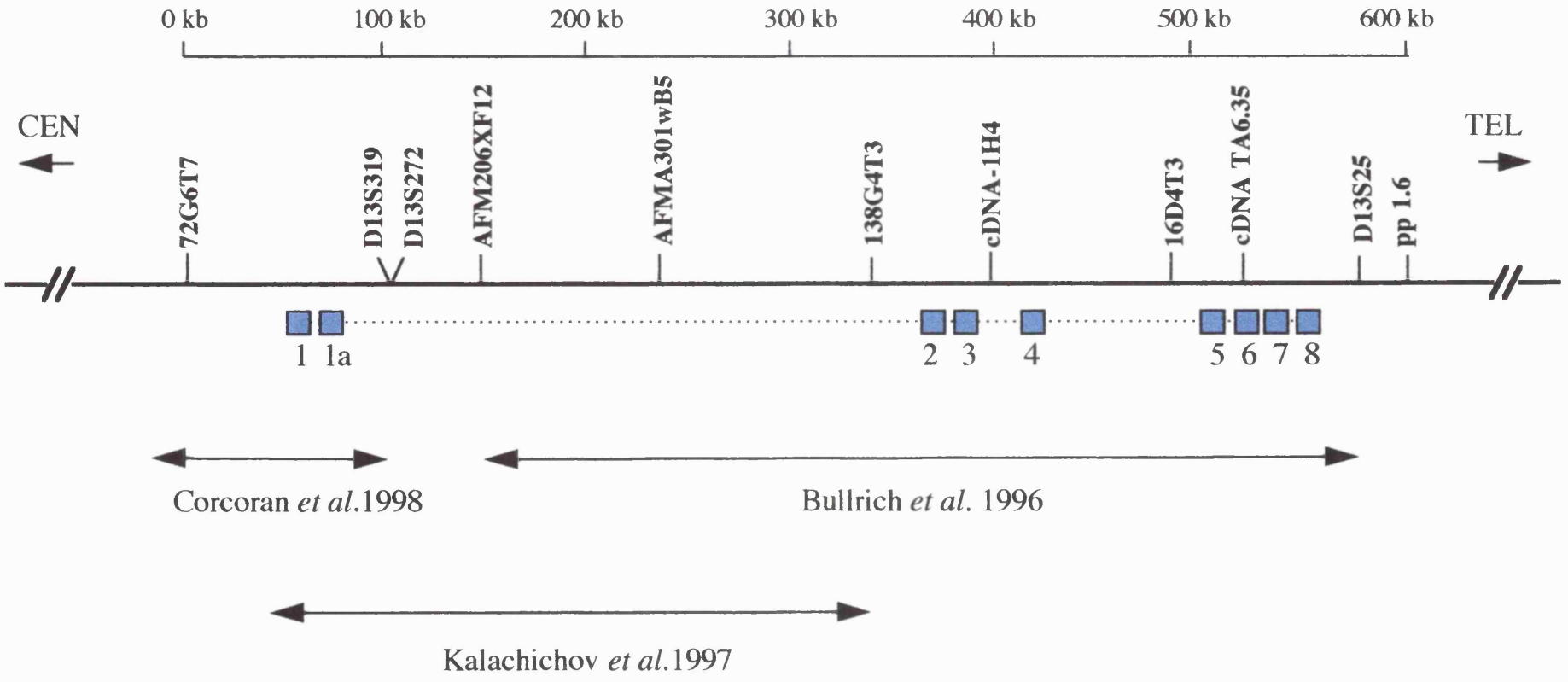


Figure 6.14: Showing positions of the exons of cDNA clone 1 on the physical map of the deleted area at 13q14.3 and how some exons fall into each of the published minimal areas of deletion shown.

must hybridise to another RNA. This RNA could be a functional mRNA existing as yet another splice variant form of cDNA clones 1 and 2:2, but not including the first 6 5' exons of cDNA clone 1. Alternatively, the message seen on Northern blot analysis may be an immature or inappropriate RNA that is non functional.

Discrepancies between abundant non functional messages identified by exons of genes on Northern blot analysis and rarer functional transcripts of the same gene have been previously described. In fact, it has been postulated that some of the commoner RNA transcripts may be non functional or may have unexpected functions (Ivell, 1992). Some examples of this are given below;

1. pro-enkephalin gene - this gene is encoded on a 1.4kb transcript in the brain (both human, and rat and mouse). In the rat and mouse testis, the message identified on Northern blot is 1.7kb. The 1.4kb transcript can be detected within these tissues but only by RT PCR methods. Analysis of the 1.7kb transcript shows that it results from differential splicing of the mRNA resulting in loss of exon 1. The RNA polymerase II starts transcription within the first intron. Although the protein transcribed from the pro-enkephalin gene could be found within the testis, it was shown to be from the 1.4kb transcript expressed at much lower levels. The 1.7kb transcript, although more abundant, was non functional (Kilpatrick *et al.*, 1987).

2. The Vasopressin gene - using exon 3 from this neuropeptide gene as a probe on Northern blot, the same product can be seen in neural tissues and testicular tissues. However, when exon 1 was used as a probe, no message could be detected on a Northern blot of testicular mRNA. Analysis by RT PCR and cDNA sequencing showed that the testicular transcripts were products of differential splicing. Two new exons further upstream of exon 1 were spliced to exon 2. This transcript was shown to be non functional as translational control sequences were contained within exon 1 (Foo *et al.*, 1991).

3. POMC gene - the functional transcript of this gene exists as a 1,100 bp mRNA in pituitary tissue and as a 1,300bp transcript in hypothalamic tissue. The difference in size is due to a longer poly A tail in the hypothalamus transcript. However, the same gene hybridises to an 800bp transcript in testicular tissue. Despite the absence of both full length forms in the testis as judged by Northern blot, the peptide was still shown to exist within this tissue. The 800bp transcript lacked the 5' upstream non coding elements that were responsible for translation initiation, rendering these transcripts sterile. The normal length transcripts were shown to exist by using RT PCR techniques, accounting for the presence of the

peptide. These data illustrate how the abundant message was shown to be non functional, the relevant message being present at undetectable levels by Northern blot analysis (Chen *et al.*, 1986; Kilpatrick *et al.*, 1987).

These examples show how exons may belong to both functional and non functional transcripts of a gene. They also illustrate how the non functional transcript may be far more abundant than the functional one. However, they do not prove that either cDNA clone 1 is functional or that the 3.5kb message is non functional.

The general lack of a detectable signal when screening cDNA libraries with the TA 6.35 probe remains to be investigated. One explanation is that the abundant 3.5kb message identified by TA 6.35 is not polyadenylated (unlike the less abundant cDNA clone 1). It would, therefore, not be reverse transcribed in the cDNA reaction used for the construction of a cDNA library. All libraries used in this project were constructed with cDNAs made from reverse transcription using oligo(dT) primers i.e. only polyadenylated RNAs will be reverse transcribed. Although the Northern blot was also made using mRNA, the mRNA for this is selected by binding to a column coated with oligo(dTs). The mRNA is then eluted with a high salt wash. DNA and abundant non polyadenylated RNAs can still remain in this purified mRNA despite the selection. For this reason it is possible that hybridisation signals can be identified on such Northern blots from non polyadenylated RNAs.

6.10.3. Multiple splice forms of clone 1

RT PCR amplification of cDNA clone 1 from normal and malignant lymphocyte cDNA confirmed the existence of multiple splice versions of this transcript. These had been suggested by the varying forms identified by RACE PCR. Firstly, and arguably most significantly, clone 1 was proven to exist as an alternative form of cDNA clone 2:2. Exons 1 and 1a also constitute the 5' exons of clone 2:2 and can both be spliced to exon 2 of clone 2:2. The presence of exon 1a in clone 2:2 and EST w05790 introduces a stop codon at position 44 of the postulated open reading frame (chapter 5). The function of cDNA clone 2:2 is as yet unknown. cDNA 2:2 is expressed at a higher level than cDNA clone 1 as it is detectable on Northern blot analysis in testis, lymphoid tissues and small intestine (Liu *et al.*, 1997). Liu *et al.* identified 3 patients with very small regions of heterozygous deletion clustered around exon 1 of Leu 1 (cDNA 2:2). However,

when they sequenced the other allele in these patients, no mutations of this gene could be found. If the longer version of this gene, cDNA clone 1, has significance in the pathogenesis of B-cell CLL then mutations could have existed further downstream in the rarer transcript, explaining the discrepancies in their results.

Other spliced versions seen in cDNA clone 1 all involved the first five 5' exons as shown in figure 6.9. As described above, exon 1a was spliced out in many of the 5- RACE clones isolated. Alternative forms with or without exons 3 or 4 were also identified. In all RT PCR transcripts exons 3 or 4 were always present. Exon 5 was frequently spliced out of the 3' RACE PCR clones that were isolated and sequenced. Comment cannot be made on the existence of transcripts without exon 2 as primers from this exon were the starting point of the RACE reactions. This strategy should be borne in mind when interpreting the various open reading frames seen with these splice variants.

Although it is possible that there are other exons (as yet unidentified), that belong to this mRNA, the multiple RT PCR and RACE PCR experiments performed would suggest that the eight exons identified are the most frequently spliced within this transcript. Some of the RACE PCR sequences identified were not characterised and may represent additional exons. However, these novel sequences were not amplified again when normal lymphocyte cDNA was amplified by RT PCR across exon-exon boundaries from exon 2 to exon 6.

Alternative splicing is a normal, common phenomenon seen in many genes as discussed in chapter 3. The dystrophin gene, mutated in Duchenne muscular dystrophy, exists as many different functional and non functional splice forms and new variants continue to be described (Surono *et al.*, 1997). The reason for the existence of so many alternative splice versions of cDNA clone 1 is unclear. The possible mechanisms of function of this transcript are discussed below. If clone 1 is a non functional RNA then there will be little or no pressure for it to be correctly spliced. It is conceivable that, in the activated environment of the cell nucleus, combinations of transcription factors, DNA binding elements and enzymes could lead to spurious RNA splicing that was 'consistent' and reproducible if the consequences were negligible. Alternatively, one of the transcripts may be functional and the others may exist as regulatory elements of the same transcript.

6.10.4. What is the significance of the transposon sequences isolated with cDNA clone 1 sequence ?

The L1 transposon element found in the original 5'-RACE PCR sequence was identified on several occasions. Whilst all the RACE PCR sequences identified the L1 element as being spliced to the 5' border of exon 2, in the cDNA clone 11:2 it was spliced to the 5' border of exon 3. An EST was deposited in the BLAST database in March 1999 (no. AI 337241), which was 100% homologous to this original 5' RACE PCR clone i.e. the sequence read from the L1 element into exons 2, 3 and 4 (refer to figure 6.5). The sequence of the EST then read into the 3' intron of exon 4. Consequently, this transposon has now been isolated from 2 cDNA libraries (the EST and cDNA clone 11.2), as well as from thymic cDNA by RACE PCR. As we have not yet identified the 5' intron - exon boundary of exon 2, it is possible that the L1 element is within this intronic sequence and is easily amplified by RACE PCR. However, this would not explain its appearance in cDNA clone 11:2. It is more feasible that this L1 element does exist as part of a further splice version of cDNA clone 1.

Transposons were described in chapter 1 because of their rare role in disruption of cell growth controlling genes and the subsequent emergence of malignant potential within a cell. As described, there are 50,000 copies of these L1 transposon elements comprising 5-15% of the human genome (Sassaman *et al.*, 1997; Kazazian and Moran, 1998). They can be of any length, up to 6.4kb (Lewin, 1994) and are reverse transcribed by a reverse transcriptase encoded for by any complete transposon sequences within the same genome (Sassaman *et al.*, 1997). The full L1 element contains a 5' untranslated region with an internal promoter, two open reading frames and one 3' untranslated region which terminates in a poly A tail (Lodish *et al.*, 1995).

Processed pseudogenes related to L1 elements have been identified. These pseudogenes resemble a cDNA copy of a fully processed mRNA in that they always include the 3' terminal poly A tract and lack any introns present in the parental gene (Weiner *et al.*, 1986). They are characteristically found at a location removed from the original gene e.g. β globin processed pseudo genes, as they originate from the original gene but are carried to different sites of the genome by transposons. Repetitive sequences, derived from the transposon, flank these pseudogenes and the original promoter is usually missing.

Our gene, cDNA clone 1, does not have these features of a processed pseudogene. Other functional genes containing L1 elements have been described, as discussed below. The L1 element appears to only exist in some forms of our gene, and it may be that it is spliced out of the functional transcript. When the gene was amplified from normal lymphoid cDNA between exons 1 and 2 the transposon elements were not amplified. The L1 element does not appear to be part of this message. It may, however, be part of another version of this RNA that is too long for PCR amplification to be successful due to the length of the L1 element.

It is also possible that cDNA clone 1 is a sterile transcript and it may have been inactivated during evolution because of an insertion of an L1 element.

6.10.5. Possible functions of cDNA clone 1

Despite clarification of the existence of this clone and the demonstration of multiple splice versions, the function of this transcript remains unclear. The lack of an extensive open reading frame in any of the splice versions suggests that this transcript may not encode for a peptide. If the majority of the 5' exons are noncoding then a postulated open reading frame of 297bp does exist (with the ATG at position 121 of exon 6), i.e. encoding for a 99 amino acid peptide. As shown, this open reading frame exists in all splice versions identified.

Other similar transcripts have initially been reported to be sterile but further experimentation has shown that they do, indeed, encode for a small peptide. For example, an RNA originally described as a functional non coding transcript is the Drosophila *hsr* omega gene. Despite the fact that the RNA is spliced and polyadenylated, only very short open reading frames exist within the sequence. Work done by Fini *et al.* showed that one of these open reading frames is actually translated (Fini *et al.*, 1989). Similarly, it is possible that one of the short open reading frames of cDNA clone 1 is translated.

When sequenced in patients with heterozygous deletion of 13q14.3, no mutations could be found either in this potential open reading frame or any of the other upstream exons. Technically there are some theoretical problems with the way in which this was done: Some of the analysis was carried out on amplified cDNA and other parts of the gene were analysed directly from DNA. Not all the intron-exon boundaries had been identified and some exons could not be analysed at the DNA level. Consequently, the full sequence of cDNA clone 1 was amplified

and analysed in overlapping sections as described. Sequencing from cDNA only, as was the case for exons 2, 3 and 5 has the following problem; if the retained allele in these heterozygously deleted patients does not give rise to a transcript for cDNA clone 1 due to abnormalities within it, then cDNA from normal contaminating lymphocytes will have been amplified in preference, as cDNA from the tumour cells will not exist. These normal cells obviously would not be expected to carry any defects at 13q14.3. Amplification and sequencing directly from DNA is more certain to represent tumour cells as, in all samples, they were present in excess. Therefore, although it is unlikely that this gene is mutated in these patients, we have not conclusively proven this as yet.

If the transcript for clone 1 does not encode for a peptide then two other possibilities remain. Firstly, the transcript may be non functional. Many consistently spliced, non functioning polyadenylated RNAs exist as has already been discussed in this chapter. The lack of an open reading frame in cDNA clone 1 and the lack of species conservation of the exons is compelling evidence that this transcript is another one. It should be noted that lack of species conservation within the mouse is not conclusive of lack of conservation in other species. Although there is more homology between human coding sequences and those of the mouse than other species, some human genes are conserved in other genomes and not in the mouse. For example, the Kallman gene is conserved in the horse and pig but not in the mouse (Legouis *et al.*, 1991).

The second possibility is that the mRNA has a function that is not related to transcription of a peptide i.e. it is a non coding RNA. An ever increasing body of literature is available describing these RNAs and their functions would appear to be every bit as diverse as those for RNAs that do encode proteins. One very interesting non coding RNA that has similarities to cDNA clone 1 is a transcript known as BORG. BORG is a target gene of bone morphogenetic proteins (BMP). It's expression is dramatically induced in a mouse myeloblast cell line when the cells are treated with bone morphogenetic proteins. The RNA of this gene was identified by differential display techniques (Takeda *et al.*, 1998). The cDNA of the gene was identified with a combination of 5'-RACE PCR and cDNA library screening techniques. When the gene was cloned and sequenced some of its features bore a resemblance to our transcript, cDNA clone 1. Following a data base search for homologies using the BLAST database, no human cDNA or RNA sequence homologies for BORG were found. However, Takeda *et al.* identified several interspersed repetitive sequences in the middle part of the cDNA. Three of

these sequences had 70%, 78% and 62% homology respectively to known L1 elements. Another similar feature of this cDNA was the lack of any extensive open reading frame. Although the cDNA was 2812bp in length to the poly-A addition site, the longest open reading frame was only 363bp. The investigators were able to demonstrate that BORG induced the expression of alkaline phosphatase activity in myeloblast cell lines despite not appearing to code for a protein. They suggested that their gene functioned as a non coding RNA. The significance of the presence of L1 elements within BORG is unknown

Other genes that function as non coding RNAs have also been described. Two of the first to be reported were H19 and Xist. They resemble BORG and cDNA clone 1 in that they are spliced and polyadenylated but lack any extensive open reading frames. Xist RNA is confined to the nucleus and acts on the X chromosome inactivation centre. It is essential for inactivation of most genes along the inactivated X chromosomes in female mammals. The RNA does not encode for a protein but the Xist mRNA combines with the inactivated chromosome, suggesting a structural role for the mRNA (Brockdorff *et al.*, 1992; Brown *et al.*, 1992; Kuroda and Meller, 1997).

Perhaps of greater interest to this project is the gene H19. This gene is located close to the insulin-2 and insulin-like growth factor 2 genes at 11p15.5 and, like them, is paternally imprinted (Leighton *et al.*, 1995). Despite being spliced and polyadenylated, H19 lacks any extensive open reading frame and does not appear to encode for a protein. H19 is expressed in differentiating foetal cells and the RNA transcript has been demonstrated to have tumour suppressor activity in embryonal tumour cell lines (Hao *et al.*, 1993). The mechanism of tumour suppressor function of H19 remains unknown as the normal function of the transcript is at present unknown.

Unlike cDNA clone 1, all three of these non coding RNAs have homologues in the mouse demonstrating evolutionary conservation and suggesting a function for the transcripts. However, the possibility of cDNA clone 1 having a function as an RNA has yet to be investigated.

Chapter 7. Mapping the minimal region of deletion within our B-cell CLL patients

As for any positional cloning project, defining the abnormal genomic region associated with a disease forms an essential part of the project design. Prior to the start of this work Dr Jabbar from Dr Foroni's group analysed 44 patients with B-cell CLL for deletion at 13q14.3 using probes for the microsatellite markers D13S25 and D13S319. Using Southern blotting techniques, she showed that 50% of patients had either a heterozygous or homozygous deletion of these probes. She also noted that when one probe was deleted the other was also deleted in the patients tested. She concluded that the minimal region of deletion in these patients spanned these two markers. A physical map of this region was then constructed as discussed in chapter 3. As candidate tumour suppressor gene cDNAs were identified from this region and mapped to 13q14.3, new probes from the area were isolated and it was possible to refine and add to the information from our patient cohort using Southern blotting techniques. The data discussed in this chapter results from the on-going mapping of the deleted region that occurred throughout this project. As a consequence, the minimal area of deletion at 13q14.3 in our patients with B-cell CLL has been defined.

None of the work discussed here was carried out prior to the start of this thesis.

7.1. Methods

7.1.1. Preparation of DNA samples

The DNA for analyses was extracted from peripheral lymphocytes. The diagnosis of B-cell CLL was made for a new patient in the clinic based on clinical history, lymphocyte morphology and immunophenotype as described in chapter 1. A minimum of 20ml of blood was collected in heparin from a patient with confirmed B-cell CLL. The blood was diluted 1:1 with sterile Hanks media (GibcoBRL) and the mononuclear layer was then separated by centrifugation on an equal volume of Lymphoprep (Nycomed) at 1400 rpm for 30 minutes in a Beckman GS-6R centrifuge. The mononuclear layer was collected with a pipette and washed in 10ml sterile Hanks media. The cells were collected by centrifugation at 1400 rpm for 10 minutes and the wash was repeated. The cells

were then resuspended in 30ml Dulbecco's media (Gibco BRL) with 10% foetal calf serum (Gibco BRL) and incubated at 37°C for 90 minutes in a plastic tissue culture flask. The cells collected from the mononuclear layer consisted of lymphocytes and monocytes. Due to the adherent properties of monocytes the monocyte free lymphocytes were decanted off, washed again in Hanks media and collected by centrifugation as before. The cells were then resuspended in Hanks media in an appropriate volume for the pellet collected. A cell count and differential was performed using a Coulter counter. A CD 2 count (T cell marker) was also performed to estimate the percentage of T lymphocytes within the sample. To do this, 1 million cells were taken in suspension and collected by centrifugation at 2000 rpm for 2 minutes. The pellet was then resuspended in 10µl of CD 2 antibody conjugated with FITC (Becton-Dickinson) and incubated at 4°C for 30 minutes. The B cell : T cell ratio was calculated using a FACs scanner. If the T cell count was below 10% of the sample, then the cells were used to make DNA directly. If the count was above 10% then the T cells were removed by rosetting with sheep red blood cells, commercially available from TCS.

The sheep cells were prepared by separating 6ml of cells into 3 x 50ml Falcon tubes. The cells were spun at 2400 rpm for 10 minutes, washed in 50ml of sterile Hanks solution and collected again by centrifugation. This wash step was then repeated. 10ml of Hanks solution was then added to the cells with 0.5ml neuramidase-1 unit/ml stock (Sigma). The cells were then incubated at 37°C for 45 minutes, pelleted by centrifugation as before and washed for a final time in 50ml Hanks media. The cells were resuspended in 50ml Hanks media and stored at 4°C until use.

To remove the T lymphocytes from the patients sample by rosetting, the patients total lymphocytes were resuspended in RPMI (GibCo BRL) and 10% foetal calf serum at a concentration of 4 x 10^6 cells/ml. 30ml of cells were taken and added to 10ml of prepared sheep red blood cells in a 50ml Falcon tube. The cells were gently mixed and then incubated together at 4°C overnight. The incubated cells were layered onto an equal volume of Lymphoprep and the gradient was centrifuged at 2000 rpm for 30 minutes at 4°C. During centrifugation the B cells remain at the interface whilst the T cells that have bound to the sheep red cells pass through the Lymphoprep gradient and pellet at the bottom of the tube. The B cells were then removed from the interface with a plastic pipette and washed in 50ml of Hanks media. The cells were collected by centrifugation as before and washed a second time in sterile Hanks solution. The B cells were

pelleted and were then ready to be used for DNA preparation. A further assessment of purity of the B cell clone was performed after Southern blotting as described below.

DNA was made from the final purified B lymphocyte pellet using a Puregene DNA extraction kit (catalog No.12-5001). The genomic DNA was resuspended to a final concentration of 1μg/μl in TE. 10μg of DNA was digested with the appropriate enzyme in a 50μl reaction and was electrophoresed on a large 0.8% agarose, 1 x AGB gel with 10μg/ml of ethidium bromide at 35-40 volts for Southern blotting as previously described.

7.1.2. Assessment of clonality of the purified cell sample.

The proportion of the purified lymphocytes that were from the malignant B cell clone was estimated by hybridising lymphocyte DNA from each patient to a probe for the J_H locus of the immunoglobulin heavy chain. This method has been described previously by Foroni *et al.* (Foroni *et al.*, 1991). As discussed in chapter 1, the immunoglobulin heavy chain is in germline configuration in immature B cells but then undergoes a unique rearrangement within each maturing B cell. When the JH probe is hybridised to DNA from a normal lymphocyte sample, only the germline configuration will be identified as the unique rearrangements cannot be identified by this method. However, if the B cells all belong to one clone, as in CLL, then the rearrangement of the JH locus within those cells will be seen on Southern blot hybridisation as a different band from the germline (figure 7.1a). The amount of signal from the rearranged band/s was compared to the residual germline signal by eye. Only those samples with a clonal population of >85% malignant B lymphocytes were assessed as informative for loss of markers from 13q14.3. If normal contaminating lymphocytes were present in over 15% of the sample then interpretation of the signals from hybridisations was deemed to be inaccurate.

7.1.3. Assessment of deletion at 13q14.3

Control probes from 13q14.3 were hybridised to blots from all patients to quantify the amount of DNA loaded in each lane. The control probes used were from the human TCRD gene (chromosome 14q11) (Baer *et al.*, 1988) and the human tumour suppressor gene, APC (chromosome 5q) (Jabbar *et al.*, 1995). Chromosome 5 has not been documented to be involved in B-cell CLL and

involvement of chromosome 14 in B-cell CLL is usually at 14q32 (chapter 1). At least 2 normal, non CLL DNA samples were included in each gel to control for the hybridisation efficiency of the probes.

All the test probes used are described below. For the location of each probe at 13q14.3, refer to figure 7.5 at the end of this chapter.

1. D13S25 -a Bluescript clone of a 2kb Hind III genomic fragment containing the D13S25 sequence (Jabbar *et al.*, 1995). This hybridised to a 2.5kb fragment on Hind III digested DNA and a 18kb fragment on EcoR I digested DNA. See figure 7.1b.
2. D13S319 -a Bluescript clone of a 4kb EcoR I repeat free fragment from a cosmid MGG15, kindly provided by Liu *et al.* (Liu *et al.*, 1993). This probe hybridised to a 4kb fragment on EcoR I digested DNA. See figure 7.1c.
3. cDNA 2:2 -a Bluescript clone of a 900bp cDNA insert, containing 2 exons as described in chapter 4 (corresponding to the gene Leu 1). This cDNA hybridised to a 8.5kb and a 20kb fragment on EcoR I digested DNA. On a Hind III digested blot, it hybridised to a 5.5kb and a 8.5kb fragment. See figure 7.2 b.
4. Exon 6 -a Bluescript clone of a 199bp probe as described in chapter 3 (TA 6.35). This probe hybridised to a 6kb fragment on EcoR I digested DNA and a 6kb fragment on Hind III digested DNA. See figure 7.2 c.
5. Race 1a -a Bluescript clone of a 500bp repeat free genomic fragment sub cloned from a λ Get clone containg the sequence of exon 4 of cDNA clone 1. This probe identified a 700bp fragment on a Hind III digest and a 10kb fragment on an EcoR I digest. See figure 7.2a.

Figure 7.1a: Southern blot analysis of the immunoglobulin heavy chain probe, J_H. 10μg of human genomic DNA from a normal control (N) and from patients with B-cell CLL were digested with EcoR I, blotted and hybridised in Church buffer to the human J_H probe. After a medium stringency wash (3 x SSC/ 0.1% SDS at 65°C) the filter was exposed to Fuji Medical autoradiography film for 6 days. The unrearranged 20kb germline band can be seen in the normal control. The patient samples show varying percentages of rearrangement depending on the clonality of the purified lymphocytes from which the DNA was prepared e.g. patient 75 shows 60% rearrangement and patient 39 has 100% rearrangement.

Figure 7.1b: Southern blot analysis of the control probe APC and the 13q14.3 probe, D13S25. 10μg of human genomic DNA from normal controls (N) and from patients with B-cell CLL were digested with Hind III, blotted and hybridised in Cambridge buffer to the human control probe APC and the test probe D13S25. After a high stringency wash (0.5 x SSC/ 0.1% SDS at 65°C) the filter was exposed to Fuji Medical autoradiography film for 3 days. The APC probe hybridised to a 4kb fragment and the D13S25 probe hybridised to a 2kb fragment. There is homozygous loss of the D13S25 band in samples 24 and 3. The arrow marks the diminished signal seen with this probe in patient 54 who had deletion of this marker (see text for full description).

Figure 7.1c: Southern blot analysis of the control probe APC and the 13q14.3 probe, D13S319. 10μg of human genomic DNA from normal controls (N) and from patients with B-cell CLL were digested with EcoR I, blotted and hybridised in Cambridge buffer to the human control probe APC and the test probe D13S319. After a high stringency wash (0.5 x SSC/ 0.1% SDS at 65°C) the filter was exposed to Fuji Medical autoradiography film for 11 days. The APC probe hybridised to a 6.5kb fragment and the D13S319 probe hybridised to a 4kb fragment. There is homozygous loss of the D13S319 marker in patient 23 and the arrow marks the diminished signal seen in patient 69 who had heterozygous loss of this marker.

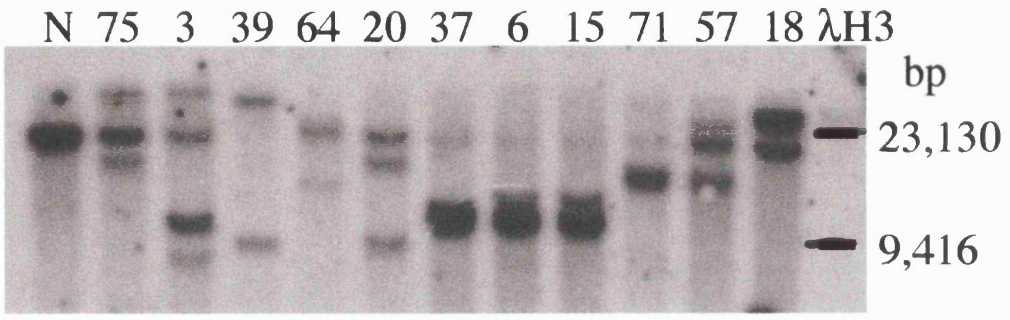


Figure 7.1a.

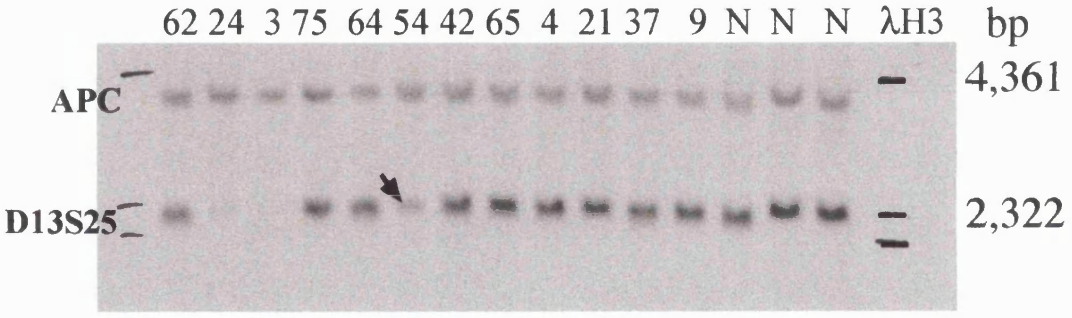


Figure 7.1b.

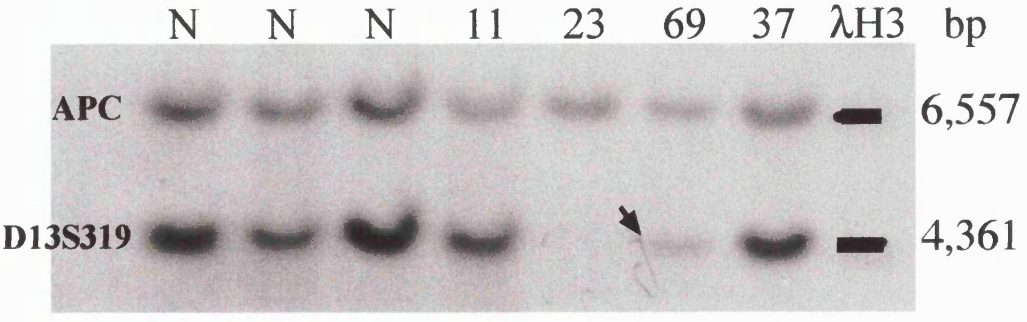


Figure 7.1c

Figure 7.2a: Southern blot analysis of the human probe Race 1a. 10μg of human genomic DNA from both normal controls (N) and patients with B-cell CLL were digested with Hind III, blotted and hybridised in Cambridge buffer. After a high stringency wash (0.5 x SSC/ 0.1 % SDS at 65°C) the filter was exposed to Fuji Medical autoradiography film for 14 days. The probe hybridised to a 700bb fragment.
When compared to a signal from the control probe APC, hybridised to the same gel, only patient 44 had a diminished signal compatible with heterozygous deletion of this probe (shown with an arrow). Although the signals are weak for patients 45 and 60, the signal from the APC probe was also weak, due to underloading of DNA on the gel.

Figure 7.2b: Southern blot analysis of the human cDNA probe 2:2. 10μg of human genomic DNA from both normal controls (N) and patients with B-cell CLL were digested with EcoR I, blotted and hybridised in Cambridge buffer. After a high stringency wash (0.5 x SSC/ 0.1 % SDS at 65°C) the filter was exposed to Fuji Medical autoradiography film for 14 days. As previously described, the probe hybridised to an 8.5kb fragment and a 20kb fragment, representing the two exons of this cDNA.
Patients 46 and 73 show homozygous deletion of both bands.

Figure 7.2c: Southern blot analysis of the human probe exon 6 and the control probe APC. 10μg of human genomic DNA from both normal controls (N) and patients with B-cell CLL were digested with EcoR I, blotted and hybridised in Cambridge buffer. After a high stringency wash (0.5 x SSC/ 0.1 % SDS at 65°C), the filter was exposed to Fuji Medical autoradiography film for 11 days. The exon 6 probe identified a 6kb fragment and the APC probe identified a 3.8kb fragment.
Patients 54, 69 and 23 were shown on this gel to have heterozygous deletion of this probe and patient 53 was shown to have homozygous deletion of this probe as indicated by arrows.

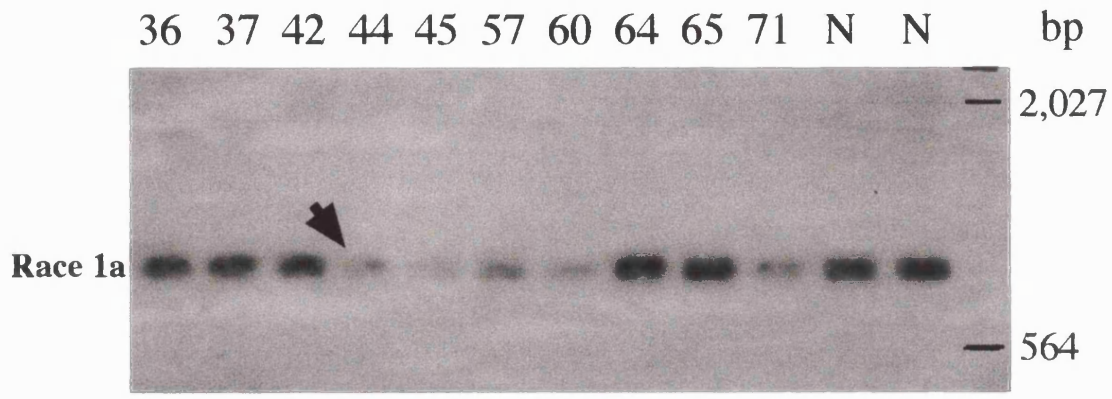


Figure 7.2a

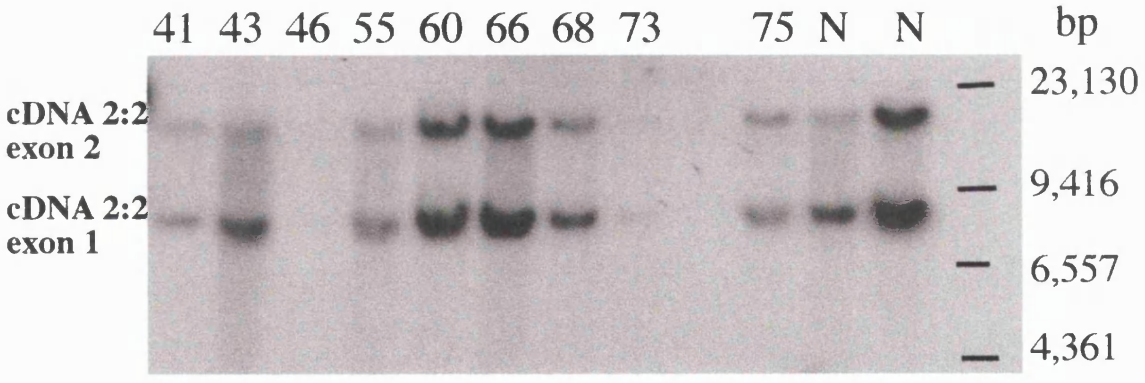


Figure 7.2b

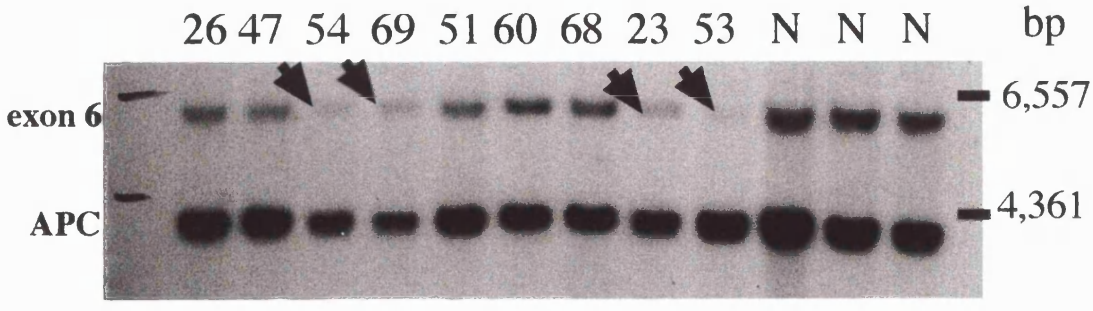


Figure 7.2c

Control probes used:

1. TCRD -a Bluescript clone of a EcoR I fragment from the TCRD constant region, kindly donated by Dr TH Rabbits (Baer *et al.*, 1988). This probe identified a 3.5kb band on an EcoR I digest.
2. APC -from 5q31 as previously described (Jabbar *et al.*, 1995). See figures 7.1b and c. This probe identified a 6kb band on an EcoR I digest and a 3.9kb band on a Hind III digest.

7.1.4. Analysis of signal intensity.

A Molecular Dynamics densitometer was used to quantitate the hybridisation signals from each probe and the data was analysed with Molecular Dynamics Image Quant software. Where possible, the same hybridisations were repeated with different blots. For each test probe a value was recorded and this was compared to the value documented for the control probe hybridisation to that same sample. The two values were expressed as a ratio (test probe / control probe). The same calculation was performed for each normal DNA sample on the gel. The patient ratio was then expressed as a percentage of the ratio seen in the normal samples. Where readings differed for the same patient between gels, an average was taken if the readings were similar. If the readings were persistently very different for a particular probe then the patient was noted as uninformative for that locus. There was a noticeable variation within the ratios seen for the normal samples. Therefore, a patient may have a percentage of 75% of the averaged normal ratio and yet have a ratio similar to one of the individual normal samples. In a situation like this the patient would be recorded as not having a deletion of that probe.

7.2. Results.

Table 7.1 shows the results of the hybridisations obtained for each probe in the patients tested.

Of the 76 patients looked at in this study, the DNA purified from 12 was shown not to be from a clonal lymphocyte population of over 85%. In the

Patient No	Clonality	cDNA 2:2	D13S319	Race 1a	Exon 6	D13S25	Status
1	>95%	50%	60%	nd	94%	90%	DG
2	100%	100%	100%	nd	100%	100%	GG
3	>85%	0%	0%	nd	0%	0%	DD
4	90%	80%	93%	86%	100%	100%	GG
5	100%	33%	nd	27%	33%	45%	DG
6	>95%	100%	100%	100%	100%	100%	GG
7	>90%	31%	30%	nd	94%	100%	DG
8	>95%	88%	100%	100%	100%	93%	GG
9	>85%	100%	100%	100%	100%	100%	GG
10	95%	0%	0%	0%	0%	75%	DD
11	>85%	50%	100%	100%	nd	nd	DG
12	100%	80%	100%	nd	92%	100%	GG
13	>90%	nd	0%	nd	0%	0%	DD
14	100%	18%	nd	nd	22%	25%	DG
15	>95%	80%	90%	nd	98%	100%	GG
16	85%	100%	100%	100%	100%	nd	GG
17	>75%	100%	nd	nd	100%	100%	GG
18	100%	50%	nd	42%	40%	50%	DG
19	>90%	46%	nd	46%	52%	40%	DG
20	60%	44%	51%	nd	50%	45%	NI
21	60%	90%	85%	76%	100%	80%	NI
22	60%	80%	100%	nd	100%	100%	NI
23	>85%	0%	0%	0%	48%	100%	DD
24	>85%	0%	0%	nd	0%	0%	DD
25	100%	nd	0%	50%	31%	0%	DG/DD
26	90%	100%	88%	nd	80%	100%	GG
27	100%	100%	nd	nd	90%	100%	GG
28	>95%	100%	100%	53%	100%	nd	DG
29	85%	0%	0%	nd	0%	0%	DD
30	100%	100%	83%	nd	100%	100%	GG
31	>85%	58%	53%	52%	60%	70%	DG
32	75%	55%	30%	42%	nd	32%	NI
33	>95%	28%	36%	33%	42%	38%	DG
34	90%	44%	66%	nd	50%	47%	DG
35	75%	100%	nd	nd	100%	34%	NI
36	75%	nd	85%	100%	100%	100%	NI
37	>95%	100%	100%	100%	89%	100%	GG
38	>85%	100%	90%	nd	100%	nd	GG

nd - not done or uninterpretable
GG - germline for both alleles
DG - heterozygously deleted
DD - homozygously deleted
R - rearranged allele
NI - not informative (non clonal population)

Patient No	Clonality	cDNA 2:2	D13S319	Race 1a	Exon 6	D13S25	Status
39	100%	100%	nd	nd	100%	100%	GG
40	100%	100%	100%	nd	100%	100%	GG
41	>90%	100%	nd	nd	92%	100%	GG
42	85%	100%	95%	100%	100%	100%	GG
43	100%	70%	70%	100%	100%	nd	?GG
44	>85%	42%	55%	45%	100%	100%	DG
45	>85%	100%	100%	87%	100%	92%	GG
46	90%	0%	0%	nd	0%	0%	DD
47	85%	100%	80%	100%	100%	nd	GG
48	85%	36%	34%	57%	45%	33%	DG
49	100%	53%	nd	47%	53%	50%	DG
50	>90%	33%	54%	61%	85%	100%	DG
51	100%	98%	nd	73%	100%	100%	GG
52	>85%	100%	nd	100%	95%	100%	GG
53	>85%	R/D	0%	0%	0%	26%	DD
54	>85%	13%	nd	6%	20%	21%	DG
55	100%	60%	43%	51%	33%	50%	DG
56	>90%	35%	nd	20%	34%	45%	DG
57	60%	nd	nd	92%	90%	100%	NI
58	100%	37%	nd	36%	46%	55%	DG
59	75%	38%	nd	42%	54%	40%	NI
60	90%	100%	100%	80%	90%	80%	GG
61	>90%	51%	43%	nd	100%	100%	DG
62	100%	73%	93%	nd	95%	92%	GG
63	95%	41%	50%	nd	nd	40%	DG
64	85%	nd	100%	79%	100%	100%	GG
65	>90%	nd	90%	100%	100%	90%	GG
66	>95%	100%	nd	nd	100%	100%	GG
67	80-85%	50%	nd	nd	nd	73%	NI
68	85%	90%	nd	71%	75%	88%	GG
69	85%	nd	50%	36%	55%	27%	DG
70	60%	100%	100%	nd	100%	100%	NI
71	100%	72%	nd	79%	92%	97%	GG
72	70%	100%	nd	nd	100%	100%	NI
73	100%	0%	0%	nd	0%	0%	DD
74	>95%	100%	68%	90%	100%	nd	GG
75	60%	100%	nd	nd	100%	100%	NI
76	>90%	31%	nd	46%	nd	30%	DG

Table 7.1: Data from Southern blots of 76 patients with B-cell CLL. The status of the alleles at 13q14.3 can only be interpreted if the malignant population of lymphocytes is >85% (clonality). See text for explanation of data.

remaining 64 patients DNA was purified from a clonal population of lymphocytes and could be analysed further;

31 patients showed germline configuration (GG) of both alleles for all probes tested from 13q14.3 (48%). 1 patient (number 43) probably has no deletion i.e. was germline for both alleles, but values were persistently difficult to interpret for D13S319 and cDNA 2:2 (table 7.1).

9 patients had homozygous deletion (DD) of the region tested at 13q14.3 (14%). 23 patients were shown to have heterozygous deletion (DG) of the region tested at 13q14.3 (36%).

The remaining 1 patient (25) had a non-contiguous homozygous deletion and will be discussed in further detail below. Including patient 25, 11 patients had deletions that were informative for the minimally deleted region at 13q14.3 and will be discussed below. The pattern of these deletions is represented in figure 7.4.

patients 1, 7 and 61 - no deletion of D13S25 or exon 6
Race 1a uninformative
heterozygous deletion of D13S319 and cDNA 2:2

patients 44 and 50 - no deletion of D13S25 or exon 6
heterozygous deletion of Race 1a, D13S319 and cDNA 2:2.

patient 11 - no deletion of Race 1a or D13S319
exon 6 and D13S25 not done / uninformative
heterozygous deletion of cDNA 2:2

Although patient 11 has deletion of only one marker, this must be regarded with caution as this result was not repeated on a second blot. Whilst interesting, this patient is not included in the data defining the minimal region of deletion at present as further work needs to be done.

patient 10 - no deletion of D13S25
homozygous deletion of all other markers tested

patient 25 - homozygous deletion of D13S25 and D13S319
heterozygous deletion of Race 1a and exon 6
cDNA 2:2 uninformative

This patient appears to have a contiguous deletion at 13q14.3 in one allele and a non-contiguous deletion of the other allele resulting in this pattern (figure 7.4).

patient 23 - no deletion of D13S25
heterozygous deletion of exon 6
homozygous deletion of Race 1a, D13S319 and cDNA 2:2

Along with patients 44 and 50, this patient defines the telomeric boundary of the minimal region of deletion as lying between Race 1a and exon 6.

patient 28 - no deletion of exon 6, D13S319 and cDNA 2:2
heterozygous deletion of Race 1a
D13S25 uninformative

These data for patient 28 were confirmed on 2 blots. However, to be certain that this patient has a small deletion identified only by this probe, further probes from the region of exon 4 (including exon 4 itself) should be tested on this patient.

patient 53 - heterozygous deletion of D13S25
homozygous deletion of exon 6, Race 1a and D13S319
rearrangement of the band identified by exon 1 of cDNA 2:2, with deletion of the band identified by exon 2 (figure 7.3).

This patient defines the centromeric border of deletion in our patients as lying between exon 1 and exon 2 of cDNA 2:2 (Leu 1).

7.3. Conclusion

In conclusion, 33 of the 64 of patients with clonal malignant B cells tested were shown to have a deletion of 13q14.3 (52%). Of these 9 (14%) had homozygous deletion of the region and 23 (36%) had heterozygous deletion of the region. A further 1 patient (2%) had a non-contiguous homozygous deletion of 13q14.3. These data are in keeping with those published by other groups (Corcoran *et al.*, 1998) (Stilgenbauer *et al.*, 1998). The patients defined a minimally deleted region with the centromeric boundary between exons 1 and 2

Figure 7.3. Southern blot analysis of the probe from cDNA 2:2 in patient 53 with B-cell CLL.

(A) 10μg of genomic DNA was digested with BamHI, blotted and hybridised to the human cDNA 2:2 probe. The blot was washed at a final stringency 0.2 x SSC/ 0.1% SDS at 65°C. The filter was then exposed to Fuji Medical autoradiography film for 3 days. A deletion and rearrangement of cDNA 2:2 can be seen in patient 53.

(B) 10μg of genomic DNA was digested with Hind III, blotted and hybridised to the human cDNA 2:2 probe. The blot was washed at a final stringency of 0.2 x SSC/ 0.1% SDS at 65°C. The filter was then exposed to Fuji Medical autoradiography film for 7 days. A deletion and rearrangement of cDNA 2:2 can be seen in patient 53.

(C) 10μg of genomic DNA was digested with EcoRI, blotted and hybridised to a PCR probe for exon 1 of cDNA 2:2. The blot was washed at a final stringency of 0.2 x SSC/ 0.1% SDS at 65°C and was exposed to Fuji Medical autoradiography film for 3 days. A rearrangement of the band identified by exon 1 can be seen in patient 53.

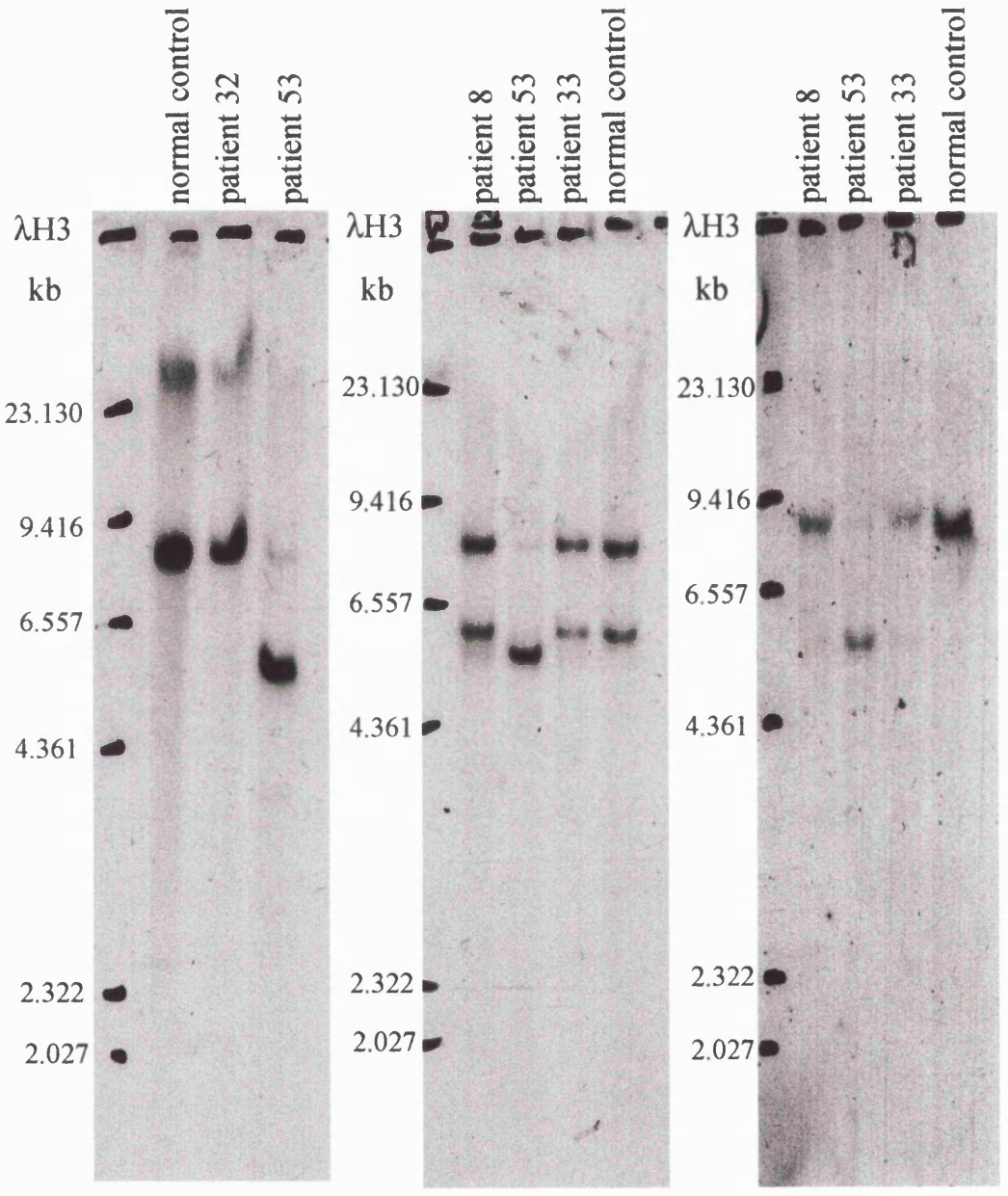


Figure 7.3. A **Figure 7.3. B** **Figure 7.3. C**

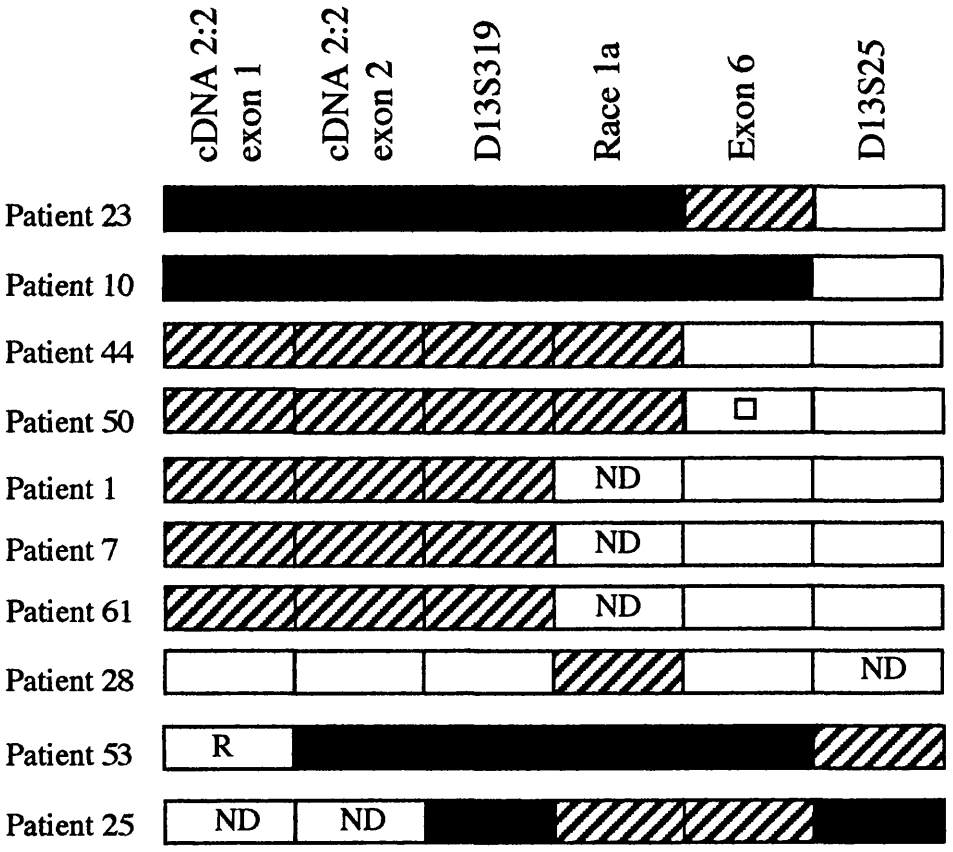


Figure 7.4: Schematic showing the minimal region of deletion (MRD), as definedby a few critical patients.The MRD in our patients is has a centromeric border between exons 1 and 2 of gene 2:2 and a telomeric border between Race 1a and exon 6. See text.
Note patient 25 has a discontinous homozygous deletion.

of cDNA 2:2, and a telomeric boundary between Race 1a and exon 6. This minimal region of deletion overlaps with those published by others (Bullrich *et al.*, 1996; Kalachikov *et al.*, 1997; Corcoran *et al.*, 1998) as shown in figure 7.5.

As described in patient 28, deletion of probe Race 1a only was detected suggesting a much smaller minimal region of deletion than described. However, densitometry is a variable technique, even when normal controls are used for comparison and the experiments are repeated with several blots. To draw firm conclusions, the same result should be obtained for more than one probe from a region. When defining small regions of deletion, the probes used may need to be close together. Further work is required on patient 28 in order to clarify the extent of this apparent small deletion. Interestingly, this patient was documented as GG for 13q14.3 until they were tested with the probe for Race 1a, highlighting the need to keep testing the patients that appear to have no deletion at 13q14 as new probes are obtained.

Patient 25 has a non-contiguous deletion of one allele at 13q14.3, causing homozygous deletion of D13S319 and D13S25 with heterozygous retention of the probes in between i.e. Race 1a and exon 6. Patients with similar, non-contiguous deletions have been reported by other groups (Bullrich *et al.*, 1996; Kalachikov *et al.*, 1997; Stilgenbauer *et al.*, 1998). One group used FISH analysis to show that a similar non-contiguous deletion was due to a complex rearrangement secondary to a translocation with chromosomal loss (Stilgenbauer *et al.*, 1998). The pattern of deletion in Patient 25 would suggest that Race 1a is outside the region of deletion and defines the telomeric border of minimal deletion in our patients. However, due to the complex nature of the deletion in this patient, caution is needed in interpreting these results. It should be noted that such discontiguous deletions would affect exons from cDNA clone 1 at both the telomeric and centromeric loci.

Many of the CLL samples analysed showed clear homozygous or heterozygous loss of alleles at 13q14.3 i.e. values for all probes approximating to 50% or 0%. This is in keeping with previously published data (Jabbar *et al.*, 1995). However, in some patients the densitometry measurements of band intensities consistently gave ratios of 20-30% e.g. patients 5, 14 and 54. This suggests that the malignant clone in these patients consists of a mixture of cells, some with heterozygous deletion and others with a homozygous deletion of the same region. This assumption has been shown to be correct in patients tested by

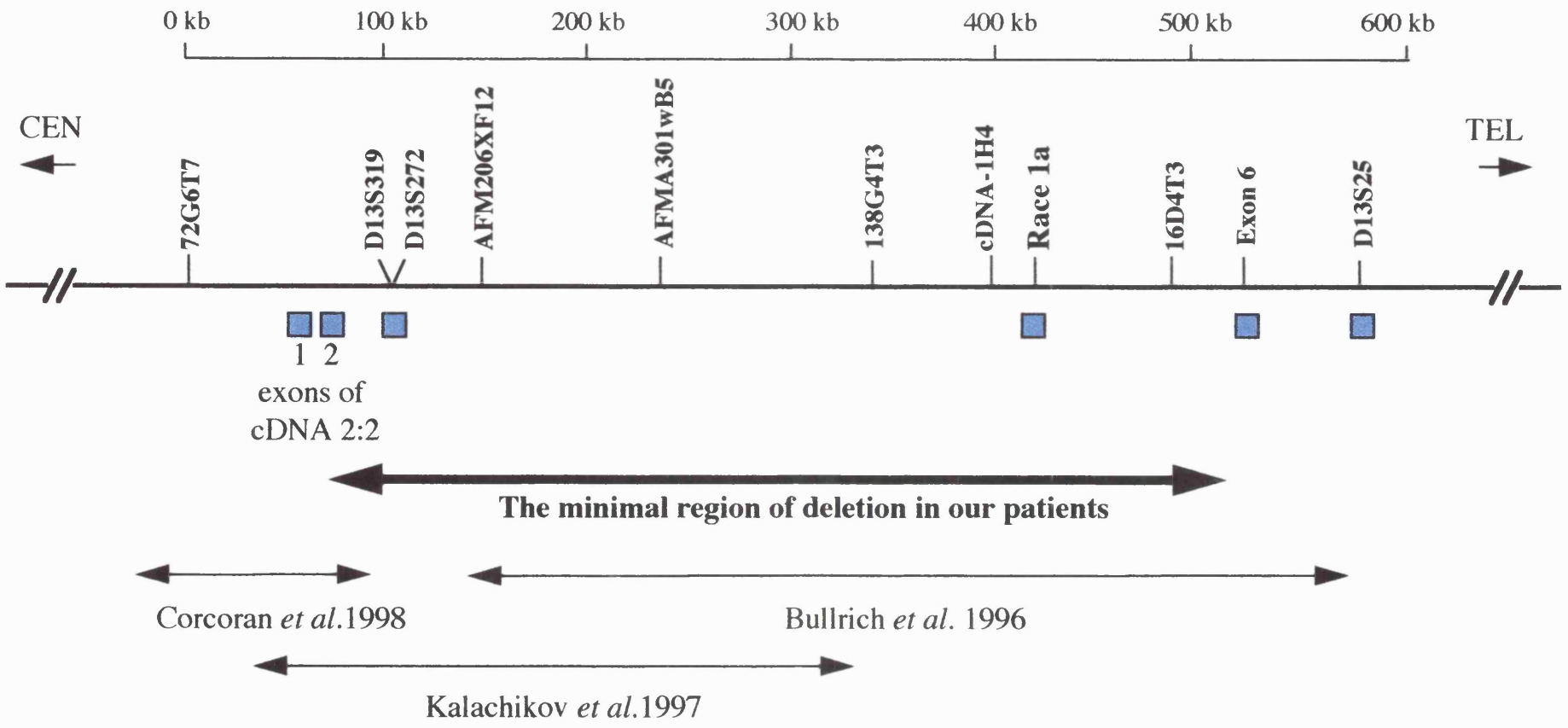


Figure 7.5: Comparison of our own region of minimal deletion with those published by other groups.

other groups using interphase FISH analysis (Avet-Loiseau *et al.*, 1996; Merup *et al.*, 1998). Some of the patients with heterozygous deletion of all probes show variation in the percentage of retained signal e.g. patients 69, 48, 33 and 34. These variations may represent subclones within the malignant lymphocytes that have small areas of homozygous deletion resulting in further loss of some markers. Densitometry is an unreliable method for analysing the presence of such subclones, as values vary significantly for the same DNA between different blots and different hybridisations. FISH remains the mainstay of this type of analysis.

In order to answer some of the questions raised above and to continue redefining the region of deletion, further work needs to be done. More unique probes from within the region of deletion, particularly between D13S319 and Race 1a (figure 7.5), need to be identified. In this way the telomeric border of the minimal region of deletion can be more clearly defined. Complementation of Southern blotting techniques with interphase FISH studies would be advantageous, both in answering the question of the existence of suclones and for the analysis of interesting patients such as patient 25.

Chapter 8. Concluding discussions

Positional cloning remains a powerful tool for the identification of regions of the human genome that contain genes whose functions are related to disease, and for the subsequent isolation of the particular gene of interest. This thesis has described a positional cloning project, designed to identify candidate tumour suppressor gene cDNAs from the region of chromosome 13q14.3 deleted in B-cell chronic lymphocytic leukaemia.

8.1. Mapping of the region of deletion at 13q14.3.

Two markers at 13q14.3 telomeric of the RB 1 gene, D13S319 and D13S25, had been identified by others as being frequently deleted in B-cell CLL (Brown *et al.*, 1993; Liu *et al.*, 1993; Liu *et al.*, 1995). A comprehensive physical map was constructed of this deleted region, including a contig of 6 YAC and 6 PAC clones, giving complete coverage of the area. Some parts of the map were further covered in cosmid and bacteriophage clones. Thirteen unique markers were positioned within this map. The physical distance between D13S25 and D13S319 was estimated at 600kb, in keeping with published data (Bullrich *et al.*, 1996; Kalachikov *et al.*, 1997; Stilgenbauer *et al.*, 1998).

Using this physical map the minimal region of deletion was defined in a cohort of our patients with B-cell CLL. 52% of patients were shown to have homozygous or heterozygous allelic deletion of 13q14.3. Similar figures have been demonstrated by other groups (Corcoran *et al.*, 1998; Stilgenbauer *et al.*, 1998). Although most patients had deletion values consistent with clonal loss of 13q14.3, a small number repeatedly showed 70-80% loss of all markers tested, suggesting a mixture of heterozygous and homozygous loss within the same tumour. This finding has also been reported by others (Avet-Loiseau *et al.*, 1996; Merup *et al.*, 1998). A minimal region of deletion was defined by a centromeric boundary at exon 1 of cDNA clone 2:2, and a telomeric boundary at exon 6. This minimal region of deletion overlaps with others published (Bullrich *et al.*, 1996; Kalachikov *et al.*, 1997; Corcoran *et al.*, 1998) as previously discussed (figure 7.5).

As illustrated in chapter 7, the true minimal region of deletion in these patients is likely to be considerably smaller than that described here. However, further work needs to done as previously indicated.

8.2. Expressed transcripts.

Using a combination of exon trapping, cDNA library screening and RACE PCR, two candidate tumour suppressor gene cDNAs were identified from the mapped region at 13q14.3.

8.2.1. cDNA clone 1.

The transcript cDNA clone 1 was initially isolated from a peripheral leucocyte cDNA library following identification of a putative exon by λ Get exon trapping from the region of deletion. Further exons from this transcript were subsequently discovered by 5'-RACE PCR analysis. The full length transcript consisted of 9 exons spliced together to form a 2kb, polyadenylated mRNA. No sequence homologies or peptide homologies with other characterised genes were found for cDNA clone 1. This lack of homology does not exclude cDNA clone 1 from being a candidate tumour suppressor gene. The tumour suppressor gene, BRCA 2, implicated in some cases of familial breast cancer, is one of an increasing number of functional genes to be discovered that do not contain any sequence or protein homologies to other described genes (Tavtigian *et al.*, 1996).

Genomic organisation of the transcript of clone 1 showed this gene to cover a genomic distance of approximately 500 kb, with the 5' exons positioned centromeric of D13S319 and the 3' exons positioned close to, but centromeric of D13S25.

Expression of this transcript could not be demonstrated by Northern blot analysis but was proven by RT PCR amplification of both normal and malignant lymphocyte cDNA. Several alternative splice forms of this transcript were shown to exist using RACE PCR on normal thymus cDNA and RT PCR amplification on cDNA from malignant and normal lymphocytes. None of these splice forms were shown to have extensive open reading frames, the longest being 297 base pairs in length. Because of this finding it was postulated that this mRNA may not encode for a peptide but may have a direct function as an RNA molecule (discussed in chapter 6).

Only one patient with heterozygous deletion of 13q14.3 was shown to have a mutation in the sequence of cDNA clone 1 in all clones sequenced and this appeared to be a silent point mutation. Some of the technical problems with the methods employed to look for mutations have already been presented and

discussed. Although most pathological lesions in human disease lie within the coding regions of genes, mutations in regulatory elements of genes have also been shown to cause disease. Most of the mutations affecting such regions described so far have been documented in the promoter regions of genes (Vogelstein and Kinzler, 1998). For example, mutations in the 5' promoter region of the β-globin gene have been shown to reduce globin chain synthesis resulting in β-thalassaemia (Antonarakis *et al.*, 1984; Orkin *et al.*, 1984; Meloni *et al.*, 1992). Occasionally, mutations in remote promoter regions have also been shown to cause disease. For example, a form of β-thalassaemia major, known as γδβ-thalassaemia, has been described arising as a result of a deletion in the gene cluster 25kb upstream of the β-globin gene (Curtin *et al.*, 1985). This deletion resulted in absent expression of β-globin RNA. The promoter region of cDNA clone 1 has not been identified and it is possible that the retained copy in patients with heterozygous loss of this locus may be non functional due to mutations within the promoter.

A CpG island has been documented close to D13S319 at 13q14.3 (Corcoran *et al.*, 1998). These rare, unmethylated regions of the human genome are often involved in regulating gene expression (Bird, 1986). Methylation of CpG islands may cause down regulation of expression of associated genes. Such alteration of gene expression by methylation changes has been shown to be relevant in malignancy. Whilst widespread hypomethylation of the genome has been reported in a variety of malignancies, including CLL (Feinberg and Vogelstein, 1983; Wahlfors *et al.*, 1992), selective hypermethylation of CpG islands in human malignancy resulting in inactivation of genes has been described (Robertson and Jones, 1997). DNA methylation mediated transcription inhibition has been proposed to be a mechanism employed by malignant cells in the inactivation of tumour suppressor genes such as the retinoblastoma gene and BRCA 1 (Robertson and Jones, 1997). The methylation status of the CpG island at D13S319 in CLL patients with heterozygous deletion of 13q14.3 has not been evaluated and may be relevant in the control of the expression of cDNA clone 1.

As illustrated here there are ways, other than by mutation or deletion of the reading frame, in which a CLL cell could inactivate cDNA clone 1. These mechanisms have yet to be investigated and this gene, therefore, remains a candidate tumour suppressor gene from the region at 13q14.3.

8.2.2. cDNA clone 2:2.

The second candidate tumour suppressor gene cDNA isolated from the region of deletion at 13q14.3 was cDNA clone 2:2. This spliced, polyadenylated transcript consisted of 3 exons existing as two alternatively spliced RNA forms. The original cDNA clone identified by cDNA library screening was formed by exons 1 and 2, and had a postulated open reading frame of 78 amino acids. This clone was 940 base pairs in length and identified a 1.1kb message on Northern blot analysis. Expression of this gene was demonstrated by Northern blot in all lymphoid tissues tested.

When cDNA clone 2:2 was amplified by RT PCR from normal lymphocyte and thymus cDNA, a second version of this gene was also identified. This alternative version had a third, 53bp exon spliced between exons 1 and 2. The sequence of this additional exon introduced a stop codon into the postulated open reading frame of this gene. We propose that this alternative version may exist as a non functional transcript, regulating expression of the shorter functional transcript. Similar regulatory alternative splice forms have been demonstrated in other genes such as the APC tumour suppressor gene (Horii *et al.*, 1993) (Thliveris *et al.*, 1994).

cDNA clone 2:2 was positioned on the physical map of 13q14.3 centromeric of D13S319. Exon 1 of this gene defined the centromeric border of the minimal region of deletion in our patients with B-cell CLL. As for cDNA clone 1, no sequence or peptide homology with known genes was demonstrated for clone 2:2. This gene has been demonstrated by others not to show species conservation (Liu *et al.*, 1997).

No mutations of this gene were found in B-cell CLL patients with heterozygous deletion of 13q14.3. This is in keeping with data from other groups (Liu *et al.*, 1997; Rondeau *et al.*, 1999). As discussed for cDNA clone 1, the promoter region of this gene has not been identified and mutations within it that may alter its expression have consequently, not been excluded. The methylation status of the CpG island at D13S319 may also be relevant for the control of this gene. Therefore, cDNA clone 2:2 also remains a candidate tumour suppressor gene from the region at 13q14.3.

8.2.3. Relationship between cDNA clone 2:2 and cDNA clone 1.

The most intriguing fact about these two candidate tumour suppressor gene transcripts is that they share their first two 5' exons i.e. exons 1 and 1a. Clone 1 exists as a longer version of clone 2:2 and it is, therefore, likely that their functions are related. As indicated by Northern blot data, the mRNA from clone 1 is less abundant than the mRNA from clone 2:2. The possible non coding role for clone 1 has already been discussed. We postulate that clone 1 acts to regulate expression of clone 2:2. In this way, loss of clone 1 e.g. by deletion, may have a profound effect on the expression of clone 2:2. If clone 2:2 encodes for a peptide with tumour suppressor function then homozygous loss of either clone may be sufficient for the emergence of malignant potential within that cell. Although no mutations have been found in the sequence of these genes in patients with heterozygous loss of 13q14.3, other mechanisms of inactivation of expression may be of relevance and these have been discussed.

If the functions of these two clones are related, and in combination they have tumour suppressor activity within normal lymphocytes, these data described in this thesis are compatible with that published by others. Some of the most compelling data in the publication by Liu *et al.* suggesting that Leu 1 (cDNA 2:2) was the gene involved in B-cell CLL, was the description of the minimally deleted region in their patients (Liu *et al.*, 1997). They narrowed the minimal region of deletion to 10kb including exons 1 of Leu 1 and Leu 2. However, they were unable to demonstrate that either of these genes were mutated on the retained allele in patients with heterozygous 13q14.3 loss. If cDNA clone 1 acts to regulate expression of Leu 1, then some of their patients may have had mutations outside of their minimally deleted area resulting in inactivation of clone 1. The net result would be down regulation of expression of Leu 1 without any detectable mutation in its genomic sequence.

Other groups have defined a more telomeric region of deletion (Bullrich *et al.*, 1996; Bouyge-Moreau *et al.*, 1997), again including many exons of cDNA clone 1. These deletions would inactivate exons of clone 1 and potentially alter expression of Leu 1 using the model described above.

8.3. Further work

The aims of this thesis were i) to define the deleted region at 13q14.3 and produce a map of the deleted region, and ii) to isolate and characterise candidate tumour suppressor gene cDNAs from within the region of deletion.

Whilst the first objective was achieved, there is further work to be done in defining the minimally deleted region. As discussed in chapter 3, there were discrepancies between some of the distances described in our physical map of 13q14.3 and those published by others. Cloning of the deleted region into smaller clones such as cosmids would help to resolve these discrepancies of distance. In addition, the identification of more STS markers between exon 4 and exon 6 of cDNA clone 1 is imperative to improve the accuracy of our map. This would allow for more detailed characterisation of the minimal region of deletion in our patients and would help to resolve the potentially informative deletions in some of the patients discussed in chapter 7 who appear to have small deletions excluding cDNA 2:2 and D13S319.

As demonstrated, patient data are vital in positional cloning projects. Evaluation of patients such as patient 25, shown to have a discontinuous deletion at 13q14.3, using additional techniques such as FISH analysis would be highly informative in further defining the minimal region of deletion. Additional patient material should be collected and analysed as described, in order to refine current data regarding the minimal region of deletion and the incidence of deletion.

The second objective of this thesis was achieved in that 2 candidate tumour suppressor gene cDNAs were isolated from the deleted region at 13q14.3. However, cDNA clone 1 has not been fully characterised. The intron-exon boundaries for exons 2, 3 and 5 have yet to be sequenced. Completion of this is currently underway. In addition, there is still doubt as too whether all the exons that belong to this transcript have been identified. Further RACE PCR experiments starting from templates other than exon 2 would help to clarify this.

We have postulated that cDNA clone 1 may belong to an ever expanding group of RNA molecules known as non coding RNA transcripts. A feature that is common to many of these transcripts is that they show nuclear restriction within the cell i.e. they are not transported into the cytoplasm and do not associate with the translational machinery of the cell (Brockdorff *et al.*, 1992; Brown *et al.*, 1992; Meller *et al.*, 1997; Bortolin and Kiss, 1998). H19 is an exception to this in that it is predominantly cytoplasmic. However, like its nuclear counterparts it has not

been shown to associate with the cellular translational machinery (Brannan *et al.*, 1990). Experiments are planned to isolate both lymphoid nuclear RNA and lymphoid cytoplasmic RNA to assess each for expression of cDNA clone 1. The demonstration that cDNA clone 1 is predominantly or exclusively expressed in the nucleus would strengthen the argument that this transcript may act as a functional non-coding RNA.

Ultimately a knowledge of the action of the RNA / peptide from both clone 1 and 2:2 will be essential for demonstration of tumour suppressor activity. Various coding and non coding RNA molecules have, however, been proven to have tumour suppressor activity without their function being known. The BRCA 1 gene has been implicated as the mutated tumour suppressor gene involved in some cases of hereditary breast cancer without its normal function being understood. Antisense oligos complimentary to BRCA 1 were transfected into cells from a breast cancer cell line, MCF-7. Following transfection the growth rate of the cells increased suggesting an inhibitory role for the functioning RNA product (Thompson *et al.*, 1995). Further experiments showed that MCF-7 tumours induced in nude mice could be inhibited by the presence of wild type BRCA 1. The growth of the tumours were unaffected by mutated BRCA 1 (Holt *et al.*, 1996). In addition, the tumour suppressor action of the non coding RNA molecule H19 was suggested when malignant potential was reversed by transfection of an expression construct of H19 into an embryonic tumour cell line (Hao *et al.*, 1993).

In order to demonstrate tumour suppressor activity of our cDNA clones similar techniques need to be explored. One of the problems facing this project is that CLL cell lines do not exist due to the non dividing nature of the mature, malignant B cells. Possibilities of looking at other cell lines such as an acute lymphocytic leukaemia cell line with 13q14.3 deletion are being investigated. The assumption that the same tumour suppressor gene is involved in all B cell malignancies with 13q14 deletion may, of course, be incorrect.

As demonstrated, whilst the initial aims of this thesis have been met, there remains more work to be done before the tumour suppressor gene located at 13q14.3 is identified and characterised. Only then will steps be made to further the understanding of the biology of this fascinating leukaemia. The hope is ultimately to translate this information into the clinical setting, leading to advances in the treatment of patients with B-cell CLL.

Appendix 1. Solutions

L-broth (LB) (Luria-Bertani)	10g bactotryptone 5g bacto yeast extract 10g NaCl to 1 litre with double distilled water
L- agar plates	1.5% bacto agar in LB sterilised by autoclaving solid media were melted by microwaving and cooled to 50°C before adding antibiotics
2 x TY	16g bactotryptone 10g yeast extract 5g NaCl to 1 litre with double distilled water
Top agar	0.7% bacto agar in media (LB or 2 x TY) Sterilised by autoclaving Solid media were melted by microwaving and cooled to 50°C before adding to cells
SOB	20g bactotryptone 5g bacto yeast extract 0.5g NaCl 10mls of 250mM KCL adjust to pH 7.0 to 1 litre with double distilled water
GTE	50mM glucose 25mM Tris pH 8.0 1mM EDTA pH 8.0

PCIA	25 parts phenol 24 parts chloroform 1 part isoamyl alcohol
CIA	24 parts chloroform 1 part isoamyl alcohol
TE	10mM Tris pH 8.0 1mM EDTA pH 8.0
20 x SSC	3M NaCl 0.3M sodium citrate
Church Buffer	500mls 1M Na_2HPO_4 pH 7.2 10g Bovine serum albumin 70g SDS 2mls 0.5M EDTA pH 8.0 to 1 litre with double distilled water
2 X Denhardts	4g BSA fraction V 4g Polyvinyl pyrrolidine 4g Ficoll to 1 litre with double distilled water filtered through 3MM paper
Salmon sperm DNA	5mg/ml DNA in distilled water add 10.5mM of NaOH boil for 10 mins
'Cambridge' Buffer	500mls 2 x Denhardts 50g Dextran Sulphate 300mls 20 x SSC 10mls 10% SDS 50mls 5mg/ml Salmon Sperm DNA to 1 litre with double distilled water

Depurination solution	20mls of HCl in 1 litre of double distilled water
Denaturing solution	1.5M NaCl 0.5M NaOH
Neutralising solution	1.5M NaCl 0.5M Tris. pH 7.2 1mM EDTA
5 x TBE	27g Trisma base 13.75g Boric acid 10mls 0.5M EDTA pH 8.3 to 500ml with double distilled water
50 X AGB	242g Trisma base 57.1ml glacial acetic acid 100ml 0.5M EDTA pH 8.0 to 1 litre with double distilled water
'Carlo's' buffer	330µl 1M Tris-acetate pH 7.5 660µl 2M K acetate pH 7.5 10µl 1M Mg acetate 1 mg/ml BSA fraction V
0.5 x TBE acrylamide solution	75ml 40% acrylamide/bis-acrylamide 50ml 5 x TBE 230g Urea to 500ml with double distilled water
2.5 x TBE acrylamide solution	37.5ml 40% acrylamide/bis-acrylamide 125ml 5 x TBE 115g Urea to 250ml with double distilled water

20%PEG 2.5M NaCl	40g PEG 29.22g NaCl dissolve in 100ml double distilled water, filter and make up to 200ml final volume.
AHC	1.7g yeast nitrogen base w/o amino acids 5g $(NH_4)_2SO_2$ 10g caseine hydrosylate 10ml 2mg/ml adenine sulphate to 900ml with double distilled water After autoclaving, add 100ml of filter sterilised 20% glucose
Sequencing buffer	200mM Tris. pH 7.5 100mM $MgCl_2$ 250mM NaCl
λ Diluent	10mM Tris. pH 7.5 10mM $MgSO_4$ filter sterilised
SM buffer	5.8g NaCl 2.0g $MgSO_4$ 50ml 1M Tris. pH 7.5 5ml 2% gelatin to 1 litre with double distilled water filter sterilised
SCE	1M Sorbitol 0.1M citrate 60mM EDTA pH 7
YAC Lysis buffer	0.5M Tris. pH 9 3% Sarkosyl 0.2M EDTA

Sucrose solutions	made up in 0.8 M NaCl 20mM Tris. pH 8 10mM EDTA pH 8.0
1M DTT	3.09g DTT in 20ml double distilled sterile water pH 5.2 with 0.01M Na Acetate
0.1M DTT used for all reactions	dilute with double distilled water
Loading buffer	50% glycerol 0.2% bromophenol blue 0.2% xylene cyanol 0.1M EDTA pH 8.0
TFB	0.5M(2(N-morpholino)ethane sulphonic acid) to pH 6.3 with potassium hydroxide 100mM potassium chloride 45mM manganese chloride 10mM calcium chloride 3mM hexamincobaltic chloride.
DnD	1M DTT 90% DMSO 10mM Potassium Acetate
5 x sequence reaction buffer	200mM Tris. pH 7.5 100mM $MgCl_2$ 250mM NaCl
2.5 x Sequencing buffer	200mM Tris. 5 mM $MgCl_2$ pH 9 at room temperature

Stop solution	95% formamide 20mM EDTA 0.05% bromophenol blue 0.05% xylene cyanol
NZY Broth	5gNaCl 2g MgSO4.7H2O 5g yeast extract 10g casein hydrolysate pH 7.5 up to 1 litre with double distilled water.
GITC	4M guanidium isothiocyanate 5mM sodium citrate pH7.0 0.1M β-mercaptoethanol 0.5% sarkosyl

References

Antonarakis S.E., Orkin S.H., Cheng T.C., Scott A.F., Sexton J.P., Trusko S.P., Charache S. and Kazazian H.H. (1984). "β-thalassemia in American Blacks. Novel mutations in the TATA box and acceptor splice site." Proceedings of the National Academy of Science, USA 81: 853-856.

Avet-Loiseau H., Devilder M.C., Garand R., Bouyge I., Rapp M.J., Milpied N., Harousseau J.L., Moisan J.P. and Bataille R. (1996). "13q14 deletions are not primary events in B-cell chronic lymphocytic leukemia: A study of 100 patients using fluorescence *in situ* hybridization." Clinical Cancer research 2: 1673-1677.

Baens M., Aerssens J., Van Zand K., Ven den Berghe H. and Meryen P. (1995). "Isolation and regional assignment of human chromosome 12p cDNAs." Genomics 29: 44-52.

Baer R., Boehm T., Spits L. and Rabbits T.H. (1988). "Complex rearrangements in human Jδ-Cδ/Jα-Cα locus. Aberrant recombination between Jα segments." EMBO Journal 7: 1661-1668.

Barr F.G. (1998). "Translocations, cancer and the puzzle of specificity." Nature Genetics 19: 121-124.

Benham F., Hart K., Crola J., Bobrow M., Francavilla M. and Goodfellow P.N. (1989). "A method for generating hybrids containing nonselected fragments of human chromosomes." Genomics 4: 509-517.

Bezieau S., Devilder M., Rondeau G., Cadoret E., Moisan J. and Moreau I. (1998). "Assignment of 48 ESTs to chromosome 13 band q14.3 and expression pattern for ESTs located in the core region deleted in B-CLL." Genomics 52: 369-373.

Binet J.L., Auquier A., Dighiero G., Chastang C., Piguet H., Goasguen J., Vaugier G., Potroni G., Colona P., Oberling F., Thomas M., Tchernia G., Jacquillat C., Boivin P., Letsy C., Duault M.T., Monconduit M., Belabbes S. and Gremy F.

(1981). "A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis." Cancer 48: 198-206.

Bird A.P. (1986). "CpG-rich islands and the function of DNA methylation." Nature 321: 209-213.

Bishop J.O. (1974). "The gene numbers game." Cell 2: 81-85.

Bortolin M.L. and Kiss T. (1998). "Human U19 intron-encoded snoRNA is processed from a long primary transcript that possesses little potential for protein coding." RNA 4: 445-454.

Boumsel L., Coppin H., Pham D., Raynal B., Lemerle J., Dausset J. and Bernard A. (1980). "An antigen shared by a human T cell subset and B cell chronic lymphocytic leukemic cells." Journal of Experimental Medicine 152: 229-234.

Bouyge-Moreau I., Rondeau G., Avet-Loiseau H., Andre M.T., Bezieau S., Cherel M., Saleun S., Cadoret E., Shaikh T., De Angelis M.M., Arcot S., Batzer M., Moisan J.P. and Devilder M.C. (1997). "Construction of a 780-kb PAC, BAC, and cosmid contig encompassing the minimal critical deletion involved in B cell chronic lymphocytic leukemia at 13q14.3." Genomics 46: 183-190.

Brannan C.I., Dees E.C., Ingram R.S. and Tilghman S.M. (1990). "The product of the H19 gene may function as an RNA." Molecular Cell Biology 10: 28-36.

Brennand J., Chinault A.C., Konecki D.S., Melton D.W. and Caskey C.T. (1982). "Cloned cDNA sequences of the hypoxanthine / guanine phosphoribosyltransferase gene from a mouse neuroblastoma cell line found to have amplified genomic sequences." Proceedings of the National Academy of Science, USA 79: 1950-1954.

Brockdorff N., Ashworth A., Kay G.F., McCabe V.M., Norris D.P., Cooper P.J., Swift S. and Rastan S. (1992). "The product of the mouse Xist gene is a 15kb inactive X-specific transcript containing no conserved ORF and located in the nucleus." Cell 71: 515-526.

Brown A.G., Ross F.M., Eimer N.D., Steel C.M. and Weir-Thompson E.M. (1993). "Evidence for a new tumour suppressor locus (DBM) in human B-cell neoplasia telomeric to the retinoblastoma gene." Nature Genetics 3: 67-73.

Brown C.J., Hendrich B.D., Rupert J.L., Lafreniere R.G., Xing Y., Lawrence J. and Willard H.F. (1992). "The human XIST gene: Analysis of a 17kb inactive X-Specific RNA that contains conserved repeats and is highly localized within the nucleus." Cell 71: 527-542.

Buckler A.J., Chang D.D., Graw S.L., Brook D., Haber D.A., Sharp P.A. and Housman D.E. (1991). "Exon amplification: A strategy to isolate mammalian genes based on RNA splicing." Proceedings of the National Academy of Sciences, USA 88: 4005-4009.

Bullrich F., Rasio D., Kitada S., Starostik P., Kipps T., Keating M., Albitar M., Reed J.C. and Croce C.M. (1999). "ATM mutations in B-cell chronic lymphocytic leukaemia." Cancer Research 59: 24-27.

Bullrich F., Veronese M.L., Kitada S., Jurlander J., Caliguri M.A., Reed J.C. and Croce C.M. (1996). "Minimal region of loss at13q14 in B-cell chronic lymphocytic leukemia." Blood 88: 3109-3115.

Caligaris-Cappio F. (1996). "B-chronic lymphocytic leukemia: A malignancy of anti-self B cells." Blood 87: 2615-2620.

Call K.M., Glaser T., Ito C.Y., Buckler A.J., Pelletier J., Haber D.A., Rose E.A., Kral A., Yeger H., Lewis W.H., Jones C. and Housman D.E. (1990). "Isolation and characterization of a zinc finger polypeptide gene at the human chromosome 11 Wilms' tumour locus." Cell 60: 509-520.

Canman C.E., Lim D.S., Cimprich K.A., Taya Y., Tamai K., Sakaguchi K., appella E., Kastan M.B. and Siliciano J.D. (1998). "p53 tumor suppressor gene is activated in response to DNA damage." Science 281: 1677-1679.

Cantwell M., Hua T., Pappas J. and Kipps T.J. (1997). "Acquired CD40-ligand deficiency in chronic lymphocytic leukemia." Nature Medicine 3: 984-989.

Chang H., Bouman D., Boerkoel C.F., Stewart A.K. and Squire J.A. (1999). "Frequent monoallelic loss of D13S319 in multiple myeloma patients by interphase fluorescence *in situ* hybridisation." Leukemia 13: 105-109.

Chapman R.M., Corcoran M.M., Gardiner A., Hawthorn L.A., Cowell J.K. and Oscier D.G. (1994). "Frequent homozygous deletions of the D13S25 locus in chromosome region 13q14 defines the location of a gene critical in leukaemogenesis in chronic B-cell lymphocytic leukaemia." Oncogene 9: 1289-1293.

Chen C.C., Chang C.C., Krieger D.T. and Bardin C.W. (1986). "Expression and regulation of Pro opiomelanocortin-like gene in the ovary and placenta: Comparison with the testis." Endocrinology 118: 2382-2389.

Cheng J., Scully P., Shew J.Y., Lee W.H., Vila V. and Haas M. (1990). "Homozygous deletion of the Retinoblastoma Gene in an Acute Lymphoblastic Leukemia (T) cell line." Blood 75: 730-735.

Cheson B.D., Bennet J.M., Grever M., Kay N., Keating M.J., O'Brien S. and Rai K. (1996). "National Cancer Institute- Sponsored working group guidelines for chronic lymphocytic leukemia: Revised guidelines for diagnosis and treatment." Blood 87: 4990-4997.

Chomczynski P. and Sacchi N. (1987). "Single step method of RNA isolation by acid guanidinium thiocyanate - phenol - chloroform extraction." Analytical Biochemistry 162: 156-159.

Cleary M.L., Smith S.D. and Sklar J. (1986). "Cloning and structural analysis of cDNAs for bcl-2 and a hybrid bcl-2/immunoglobulin transcript resulting from the t(14;18) translocation." Cell 47: 19-28.

Collins F.S. (1995). "Positional cloning moves from perditional to traditional." Nature Genetics 9: 347-350.

Collins S. and Groudine M. (1982). "Amplification of endogenous myc-related DNA sequences in human myeloid leukaemia cell line." Nature 298: 679-681.

Cooper T.A. and Mattox W. (1997). "The regulation of splice-site selection, and its role in human disease." American Journal of Human Genetics 61: 259-266.

Corcoran M.M., Rasool O., Liu Y., Iyengar A., Grander D., Ibbotson R.E., Merup M., Wu X., Brodyansky V., Gardiner A.C., Juliusson G., Chapman R.M., Ivanova G., Tiller M., Gahrton G., Yanovsky N., Zabarovsky E., Oscier D.G. and Einhorn S. (1998). "Detailed molecular delineation of 13q14.3 loss in B-cell chronic lymphocytic leukemia." Blood 91: 1382-1390.

Craig J.M. and Bickmore W.A. (1994). "The distribution of CpG islands in mammalian chromosomes." Nature Genetics 7: 376-381.

Cui H., Hordon I.L., Ohlsson R., Hamilton S.R. and Feinberg A.P. (1998). "Loss of imprinting in normal tissue of colorectal cancer patients with microsatellite instability." Nature Medicine 4: 1276-1280.

Cuneo A., Bigoni R., Negrini M., Bullrich F., Veronese M.L., Roberti M.G., Bardi A., Rigolin G.M., Cavazzini P., Croce C.M. and Castoldi G. (1997). "Cytogenetic and interphase cytogenetic characterization of atypical chronic lymphocytic leukemia carrying BCL1 translocation." Cancer Research 57: 1144-1150.

Curtin P., Pirastu M., Kan Y.W., Gobert-Jones J.A., Stephens A.D. and Lehmann H. (1985). "A distant gene deletion affects β-globin gene function in an atypical γδβ-thalassemia." Journal of Clinical Investigation 76: 1554-1558.

Dameshek W. (1967). "Chronic lymphocytic leukemia-an accumulative disease of immunologically incompetent lymphocytes." Blood 29: 566-584.

Damle R.N., Wasil T., Fais F., Ghiotto F., Valetto A., Allen S.L., Buchbinder A., Budman D., Dittmar K., Kolitz J., Lichtman S.M., Schulman P., Vinciguerra V.P., Rai K., Ferrarini M. and Chiorazzi N. (1999). "Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia." Blood 94: 1840-1847.

Devilder M.C., Francois S., Bosic C., Moreau A., Mellerin M.P., Le Paslier D., Bataille R. and Moisan J.P. (1995). "Deletion cartography around the D13S25

locus in B cell chronic lymphocytic leukemia and accurate mapping of the involved tumor suppressor gene." Cancer research 55: 1355-1357.

Dierlamn J., Michaux L., Criel A., Wlodarska I., Van Den Berghe H. and Hossfeld D.K. (1997). "Genetic abnormalities in chronic lymphocytic leukemia and their clinical and prognostic implications." Cancer Genetics and Cytogenetics 94: 27-35.

Dietz H.C., Cutting G.R., Pyeritz R.E., Maslen C.L., Sakai L.Y., Corson G.M., Puffenberger E.G., Hamosh A., Nanthakumar E.J., Curristin S.M., Stetten G., Meyers D.A. and Francomano C.A. (1991). "Marfan syndrome caused by a recurrent *de novo* missense mutation if the fibrillin gene." Nature 352: 337-339.

Dohner H., Fischer K., Bentz M., Hansen K., Benner A., Cabot G., Diehl D., Schlenk R., Coy J., STilgenbauer S., Volkmann M., Galle P.R., Poustka A., Hunstein W. and Lichter P. (1995). "p53 gene deletion predicts for poor survival and non-response to therapy with purine analogs in chronic B-cell leukemias." Blood 85: 1580-1589.

Dohner H., Stilgenbauer S., James M.R., Benner A., Weilguni T., Bentz M., Fischer K., Hunstein W. and Lichter P. (1997). "11q deletions identify a new subset of B-cell chronic lymphocytic leukemia characterized by extensive nodal involvement and inferior prognosis." Blood 89: 2516-2522.

Dono M., Hashimoto S., Fais F., Trejo V., Allen S.L., Lichtman S.M., Schulman P., Vinciguerra V.P., Sellars B., Gregersen P.K., Ferrarini M. and Chiorazzi N. (1996). "Evidence for progenitors of chronic lymphocytic leukemia B cells that undergo intraclonal differentiation and diversification." Blood 87: 1586-1594.

Duyk G.M., Kim S., Myers R.M. and Cox D.R. (1990). "Exon trapping: A genetic screen to identify candidate transcribed sequences in cloned mammalian genomic DNA." Proceedings of the National Academy of Science, USA 87: 8995-8999.

Ebling S.B., schutte M.E.M. and Logtenberg T. (1993). "Molecular analysis of V_H and V_L regions expressed in IgG-bearing chronic lymphocytic leukemia

(CLL): Further evidence that CLL is a heterogenous group of tumours." Blood 82: 1626-1631.

Elkahloun A.G., Krizman D.B., Wang Z., Hofman T.A., Roe B. and Meltzer P.S. (1997). "Transcript mapping in a 46-kb sequenced region at the core of 12p13.3 amplification in human cancers." Genomics 42: 295-301.

Evan G. and Littlewood T. (1998). "A matter of life and cell death." Science 281: 1317-1322.

Fearon E. (1997). "Human cancer syndromes: Clues to the origin and nature of cancer." Science 278: 1043-1050.

Feil R. and Kelsey G. (1997). "Insights from model systems. Genomic imprinting: A chromatin connection." American Journal of Human Genetics 61: 1213-1219.

Feinberg A.P. and Vogelstein B. (1983). "Hypomethylation distinguishes genes of some human cancers from their normal counterparts." Nature 301: 89-92.

Fini M.E., Bendena W.G. and Pardue M.L. (1989). "Unusual behavior of the cytoplasmic transcript of hsr omega: An abundant, stress-inducible RNA that is translated but yields no detectable protein product." Journal of Cellular Biology 108: 2045-2057.

Fishel R., Lesco M.K., Rao M.R., Copeland N.G., Jenkins N.A., Garber J., Kane M. and Kolodner R. (1993). "The human mutator gene homolog *MSH2* and its association with hereditary nonpolyposis colon cancer." Cell 75: 1027-1038.

Fitchett M., Griffiths M.J., Oscier D.G., Johnson S. and Seabright M. (1987). "Chromosome abnormalities involving band 13q14 in hematologic malignancies." Cancer Genetics and Cytogenetics 24: 143-150.

Foo N.C., Carter D., Murphy D. and Ivell R. (1991). "Vasopressin and oxytocin gene expression in rat testis." Endocrinology 128: 2118-2128.

Foroni L., Mason P. and Luzzato L. (1991). Immunoglobulin and T-cell receptor gene analysis and investigation of lymphoproliferativedisorders. Leukaemic Cell Methods in Haematology. 2nd, ed. 339-391.

Franco B., Guioli S., Pragiola A., Incerti B., Bardoni B., Tonlorenzi R., Carrozzo R., Maestrini E., Pieretti M., Taillon-Miller P., Brown C.J., Willard H.F., Lawrence C., Persico M.G., Camerino G. and Ballabio A. (1991). "A gene deleted in Kallmann's syndrome shares homology with neural cell adhesion and axonal path-finding molecules." Nature 353: 529-535.

Friend S.H., Bernards R., Rogelj S., Weinberg R.A., Rapaport J.M., Albert D.M. and Dryja T.P. (1986). "A human segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma." Nature 323: 643-646.

Frohman M.A., Dush M.K. and Martin G.R. (1988). "Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer." Proceedings of the National Academy of Science, USA 85: 8998-9002.

Gahn B., Schafer C., Neef J., Troff C., Feuring-Buske M., Hiddemann W. and Worman B. (1997). "Detection of trisomy 12 and Rb-deletion in CD34+ cells of patients with B-cell chronic lymphocytic leukemia." Blood 89: 4275-4281.

Gaidano G., Hauptschien R.S., Parsa N.Z., Offit K., Rao P.H., Lemoir G., Knowles D.M., Chagnati R.S.K. and Dalla-Favera R. (1992). "Deletions involving two distinct regions of 6q in B-cell non-Hodgkin lymphoma." Blood 80: 1781-1787.

Gaidano G., Newcomb E.W., Gong J.Z., Tassi V., Neri A., Cortelezzi A., Calori R., Baldini L. and Dalla-Favera R. (1994). "Analysis of alterations of oncogenes and tumor suppressor genes in chronic lymphocytic leukemia." American Journal of Pathology 144: 1312-1319.

Garcia-Marco J., Matutes E., Morilla R., Ellis J., Oscier D., Fantes J., Catovsky D. and Price C.M. (1994). "Trisomy 12 in B-cell chronic lymphocytic leukaemia: Assessment of lineage restriction by simultaneous analysis of immunophenotype

and genotype in interphase cells by fluorescent *in situ* hybridisation." British Journal of Haematology 87: 44-50.

Garcia-Marco J., Price C. and Catovsky D. (1997). "Interphase cytogenetics in chronic lymphocytic leukemia." Cancer Genetics and Cytogenetics 94: 52-58.

Garcia-Marco J.A., Caldas C., Price C.M., Wiedermann L.M., Ashworth A. and Catovsky D. (1996). "Frequent somatic deletion of the 13q12.3 locus encompassing BRCA2 in chronic lymphocytic leukemia." Blood 88: 1568-1575.

Garcia-Marco J.A., Nouel A., Navarro B., Matutes E., Oscier D., Price C.M. and Catovsky D. (1998). "Molecular cytogenetic analysis in splenic lymphoma with villous lymphocytes: Frequent allelic imbalance of the *RB1* gene but not the D13S25 locus on chromosome 13q14." Cancer Research 58: 1736-1740.

Gitschier J., Wood W.I., Goralka T.M., Wion K.L., Chen E.Y., eaton D.H., Vehar G.A., Capon D.J. and Lawn R.M. (1984). "Characterization of the human factor VIII gene." Nature 312: 326-330.

Greenblatt M.S., Bennett W.P., Hollstein M. and Harris C.C. (1994). "Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis." Cancer Research 54: 4855-4878.

Gyapay G., Morisette J., Vignal A., Dib C., Fizames C., Millaseau P., Marc S., Bernardi G., Lathrop M. and Weissenbach J. (1994). "The 1993-1994 Genethon human genetic linkage map." Nature Genetics 7: 246-307.

Haber D.A. and Fearon E. (1998). "The promise of cancer genetics." Lancet 351 (suppl II): 1-8.

Hamagucchi M., O'Connor E.A., Chen T., Parnell L., McCombie R.W. and Wigler M.H. (1998). "Rapid isolation of cDNA by hybridisation." Proceedings of the National Academy of Science, USA 95: 3764-3769.

Hamblin T.J., Davis Z., Gardiner A., Oscier D.G. and Stevenson F. (1999). "Unmutated Ig V_H genes are associated with a more aggressive form of chronic lymphocytic leukemia." Blood 94: 1848-1854.

Hamblin T.J., Davis Z., Oscier D.G. and Stevenson F.K. (1998). "In chronic lymphocytic leukemias germline configuration of immunoglobulin heavy chain genes is associated with a more aggressive form of the disease." Blood 92: Abstract 2119.

Hao Y., Crenshaw T., Moulton T., Newcomb E. and Tycko B. (1993). "Tumour-suppressor activity of H19 RNA." Nature 365: 764-767.

Hauptschein R.S., Gamberi B., Rao P.H., Frigeri F., Scotto L., Venkatraj V.S., Gaidano G., Rutner T., Edwards Y.H., Chagnanti R.S.K. and Dalla-Favera R. (1998). "Cloning and mapping of human chromosome 6q26-q27 deleted in B-cell non-Hodgkin lymphoma and multiple tumour types." Genomics 50: 170-186.

Hawthorn L. and Cowell J.K. (1995). "Integration of the physical and genetic linkage map for human chromosome 13." Genomics 27: 399-404.

Hawthorn L., Roberts T., Verlind E., Kooy R.F. and Cowell J.K. (1995). "A yeast artificial chromosome contig that spans the RB1-D13S31 interval on human chromosome 13 and encompasses the frequently deleted region in B-cell chronic lymphocytic leukemia." Genomics 30: 425-430.

Hawthorn L.A., Chapman R., Oscier D. and Cowell J.K. (1993). "The consistent 13q14 translocation breakpoint seen in chronic B-cell leukaemia (B CLL) involves deletion of the D13S25 locus which lies distal to the retinoblastoma predisposition gene." Oncogene 8: 1415-1419.

Heiskanen M. (1996). "Visual mapping by high resolution FISH." Trends in Genetics 12: 379-382.

Hochgeschwender U. and Brennan M.B. (1991). "Identifying genes within the genome: New ways for finding the needle in a haystack." BioEssays 13: 139-144.

Hoffbrand A.V. and Pettit J.E. (1994). Essential Haematology. Oxford, Blackwell Scientific Publications.

Holt J.T., Thompson M.E., Szabo C., Robinson-Benion C., Arteaga C.L., King M.C. and Jensen R.A. (1996). "Growth retardation and tumour inhibition by BRCA 1." Nature Genetics 12: 298-302.

Horii A., Nakatsuru S., Ichii S., Nagase H. and Nakamura Y. (1993). "Multiple forms of the APC gene transcripts and their tissue-specific expression." Human Molecular Genetics 2: 283-287.

Humphries C.G., Shen A., Kuziel W.A., Capra J.D., Blattner F.R. and Tucker P.W. (1988). "A new human immunoglobulin VH family preferentially rearranged in immature B-cell tumours." Nature 331: 446-449.

Hunter T. (1991). "Cooperation between oncogenes." Cell 64: 249-270.

Ibbotson R.E., Chapman R.M., Corcoran M.M. and Oscier D.G. (1996). "PCR analysis of polymorphisms at the D13S25 locus." Leukemia 10: 1712-1714.

Ionov Y., Peinado M.A., Malkhosyan S., Shibata D. and Perucho M. (1993). "Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carncinogenesis." Nature 363: 558-561.

Ivell R. (1992). "'All that glisters is not gold'-common testis transcripts are not always what they seem." International Journal of Andrology 15: 85-92.

Jabbar S.A.B., Ganeshaguru K., Wickremasinghe R.G., Hoffbrand A.V. and Foroni L. (1995). "Deletion of chromosome 13 (band q14) but not trisomy 12 is a clonal event in B-chronic lymphocytic leukaemia." British Journal of Haematology 90: 476-478.

Jackson A., Panayiotidis P. and Foroni L. (1998). "The human homologue of the *Drosophila tailless* gene (TLX): Characterization and mapping to a region of common deletion in human lymphoid leukemia on chromosome 6q21." Genomics 50: 34-43.

Jin Y., Dietz H.C., Montgomery R.A., Bell W.R., McIntosh I., Coller B. and Bray P.F. (1996). "Glazmann thrombasthenia - cooperation between sequence variants in cis during splice site selection." Journal of Clinical Investigation 98: 1745-1754.

Johnson T.A., Rassenti L.Z. and Kipps T.J. (1997). "Ig V_H1 genes expressed in B cell chronic lymhocytic leukemia exhibit distinctive molecular features." Journal of Immunology 158: 235-246.

Jongsma A., Van Someren H., Westerweld A., Hagemijer A. and Pearson P. (1973). "Localisation of genes on human chromosomes by studies of human-Chinese hamster somatic cell hybrids." Human Genetics 20: 195-202.

Juliusson G. and Merup M. (1998). "Cytogenetics in chronic lymphocytic leukemia." Seminars in oncology 25: 19-26.

Juliusson G., Oscier D. and Gahrton G. (1991). "For the International Working Party on Chromosomes in Chronic Lymphocytic Leukemia (IWCCLL): Cytogenetic findings and survival in B-cell chronic lymphocytic leukemia. Second international compilation of data on 662 patients." Leukemia and Lymphoma 5 (suppl): 21-25.

Juliusson G., Oscier D.G., Fitchett M., Ross F.M., Stockdill G., Mackie M.J., Parker A.C., Castoldi G.L., Cuneo A., Knuutila S., Elonen E. and Gahrton G. (1990). "Prognostic subgroups in B-cell chronic lymphocytic leukemia defined by specific chromosomal abnormalities." New England Journal of Medicine 323: 720-724.

Kalachikov S., Migliazza A., Cayanis E., Francchiolla N.S., Bonaldo M.F., Lawton L., Jelenc P., Ye X., Qu X., Chien M., Hauptschein R., Gaidano G., Vitolo U., Saglio G., Resegotti L., Brodjansky V., Yanovsky N., Zhang P., Soares M.B., Russo J., Edelman I.S., Efstratiadis A., Dalla-Favera R. and Fischer S.G. (1997). "Cloning and gene mapping of the chromosome 13q14 region deleted in chronic lymphocytic leukaemia." Genomics 42: 369-377.

Kamoun M., Kadin M.E., Martin P.J., Nettleton J. and Hansen J.A. (1981). "A novel human T cell antigen preferentially expressed on mature T cells and shared by both well and poorly differentiated B cell leukemias and lymphomas." Journal of Immunology 127: 987-991.

Kapanadze B., Kashuba V., Baranova A., Rasool O., van Everdink W., Liu Y., Syomov A., Corcoran M., Poltaraus A., van den Berg A., Gizatullin R., Fedorova L., Sulimova G., Zelenin A., Deaven L., Lehrach H., Grander D., Buys C., Oscier D., Zabarovsky E.R., Einhorn S. and Yanovsky N. (1998). "A cosmid and cDNA fine physical map of a human chromosome 13q14 region frequently lost in B-cell chronic lymphocytic leukaemia and identification of a new putative tumor suppressor gene, Leu 5." Federation of European Biochemical Societies 426: 266-270.

Kazazian H.H.J. and Moran J.V. (1998). "The impact of L1 retrotransposons on the human genome." Nature Genetics 19: 19-24.

Kerem B., Rommens J.M., Buchanan J.A., Markiewicz D., Cox T.K., Chakravarti A., Buchwald M. and Tsui L. (1989). "Identification of the cystic fibrosis gene: Genetic analysis." Science 245: 1073-1080.

Kilpatrick D.L., Borland K. and Jin D.F. (1987). "Differential expression of opiod peptide genes by testicular germ cells and somatic cells." Proceedings of the National Academy of Science, USA 84: 5695-5699.

Kipps T.J. (1998). "Chronic lymphocytic leukemia." Current Opinion in Haematology 5: 244-253.

Kipps T.J. and Carson D.A. (1993). "Autoantibodies in chronic lymphocytic leukemia and related systemic autoimmune diseases." Blood 81: 2475-2487.

Kitada S., Anderson J., Akar S., Zapata J.M., Takayama S., Krajewski S., Wang H.-G., Zhang X., Bullrich F., Croce C.M., Rai K., Hines J. and Reed J.C. (1998). "Expression of apoptosis-regulating proteins in chronic lymphocytic leukemia: Correlations with in vitro and in vivo chemoresponses." Blood 91: 3379-3389.

Knudson A.G. (1971). "Mutation and cancer: Statistical study of retinoblastoma." Proceedings of the National Academy of Sciences, USA 68: 820-823.

Knudson A.G. (1993). "Antioncogenes and human cancer." Proceedings of the National Academy of Science, USA 90: 10914-10921.

Koenig M., Hoffman E.P., Bertelson C.J., Monaco A.P., Feener C. and Kunkel L.M. (1987). "Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals." Cell 50: 509-517.

Korenberg J.R. and Rykowski M.C. (1988). "Human genome organization: Alu, LINES, and the molecular structure of metaphase chromosome bands." Cell 53: 391-400.

Kuroda M.I. and Meller V.H. (1997). "Transient *Xist*-ence." Cell 91: 9-11.

La Starza R., Wlodarska I., Aventin A., Falzetti D., Crescenzi B., Martelli M.F., Van den Berghe H. and Mecucci C. (1998). "Molecular delineation of 13q deletion boundaries in 20 patients with myeloid malignancies." Blood 91: 231-237.

Leach F.S., Nicolaides N.C., Papadopoulos N., Liu B., Jen J., Parsons R., Peltomaki P., Sistonen P., Aaltonen L.A. and Nystrom-Lahti M.e.a. (1993). "Mutations of a *mut*S homolog in hereditary nonpolyposis colorectal cancer." Cell 75: 1215-1225.

Legouis R., Hardelin J., Levilliers J., Claviere J., Compain S., Wunderle V., Millasseau P., Le Paslier D., Cohen D., Caterina D., Bougueleret L., Delemarre-Van de Waal H., Lutfalla G., Weissenbach J. and Petit C. (1991). "The candidate gene for the X-linked Kallmann syndrome encodes a protein related to adhesion molecules." Cell 67: 423-435.

Leighton P.A., Ingram R.S., Eggenschwiler J., Efstradtiadis A. and Tilghman S.M. (1995). "Disruption of imprinting caused by deletion of the H19 gene region in mice." Nature 375: 34-39.

Levine A. (1993). "The tumour suppressor genes." Annual Review of Biochemistry 62: 623-651.

Lewin B. (1994). Genes V. New York, Oxford University Press.

Liu Y., Corcoran M., Rasool O., Ivanova G., Ibbotson R., Grander D., Iyengar A., Baranova A., Kashuba V., Merup M., Wu X., Gardiner A., Mullenbach R., Poltaraus A., Hultstrom A.L., Julisson G., Chapman R., Tiller M., Cotter F., Gahrton G., Yankovsky N., Zabarovsky E., Einhorn S. and Oscier D. (1997). "Cloning of two candidate tumour suppressor genes within a 10 kb region on chromosome 13q14, frequently deleted in chronic lymphocytic leukaemia." Oncogene 15: 2463-2473.

Liu Y., Hermanson M., Grander D., Merup M., Wu X., Heyman M., Rasool O., Juliusson G., Gahrton G., Detlofsson R., Nikiforova N., Buys C., Soderhall S., Yanovsky N., Zabrovsky E. and Einhorn S. (1995). "13q deletions in lymphoid maligancies." Blood 86: 1911-1915.

Liu Y., Szekely L., Grander D., Soderhall S., Juliusson G., Gahrton G., Linder S. and Einhorn S. (1993). "Chronic lymphocytic leukemia cells with allelic deletions at13q14 commonly have one intact *RB1* gene: Evidence for a role of an adjacent locus." Proceedings of the National Academy of Science, USA. 90: 8697-8701.

Llewellyn D.H., Scobie G.A., Urquhart A.J., Whatley S.D., Roberts A.G., Harrison P.R. and Elder G.H. (1996). "Acute intermittent porphyria caused by defective splicing of porphobilinogen deaminase RNA - a synonymous codon mutation at -22bp from the 5' splice site causes skipping of exon 3." Journal of Medical Genetics 33: 437-438.

Lodish H., Baltimore D., Berk A., Zipursky S.L., Matsudaira P. and Darnell J. (1995). Molecular cell biology. New York, Scientific American Books.

Look A.T. (1997). "Oncogenic transcription factors in the human acute leukemias." Science 278: 1059-1064.

Malkin D., Li F.P., Strong L.C., Fraumeni J.F., Nelson C.E., Kim D.H., Kassel J., Gryka M.A., Bischoff F.Z., M.A. T. and Friend S.H. (1990). "Germ line p53 mutations in a family syndrome of breast cancer, sarcomas and other neoplasms." Science 250: 1233-1238.

Marshall C.J. (1991). "Tumour Suppressor Genes." Cell 64: 313-326.

Matutes E., Oscier D., Garcia-Marco J., Ellis J., Copplestone A., Gillingham R., Hamblin T., Lens D., Swansbury G.J. and Catovsky D. (1996). "Trisomy 12 defines a group of CLL with atypical morphology: Correlation between cytogenetic, clinical and laboratory features in 544 patients." British Journal of Haematology 92: 382-388.

Meller V.H., K.H. W., Roman G., Kuroda M.I. and Davis R.L. (1997). "roX1 RNA paints the X chromosome of male Drosophila and is regulated by the dosage compensation system." Cell 88: 445-457.

Meloni A., Rosatelli M.C., Faa V., Sardu R., Saba L., Murru S., Sciarratta P., Baldi M. and Tannoia N. (1992). "Promoter mutations producing mild β-thalassemia in the Italian population." British Journal of Haematology 80: 222-226.

Merup M., Jansson M., Corcoran M., Liu Y., Rasool O., Stellan B., Hermansson M., Juliusson G., Gahrton G. and Einhorn S. (1998). "A FISH cosmid 'cocktail' for detection of 13q deletions in chronic lymphocytic leukemia - comparison with cytogenetics and Southern hybridisation." Leukemia 12: 705-709.

Merup M., Juliusson G., Hammarstrom L., Smith C.I.E. and Gahrton G. (1994). "T-cell receptor β gene rearrangements in leukaemic B-cells from patients with chronic lymphocytic leukaemia: Association with chromosome 6 deletions." British Journal of Haematology 86: 291-297.

Monaco A.P., Neve R.L., Colletti-Feener C., Bertelson C.J., Kurnit D.M. and Kunkel L.M. (1986). "Isolation of candidate cDNAs for portions of the Duchenne muscular dystrophy gene." Nature 323: 646-650.

Mould S., Gardiner A. and Oscier D.G. (1996). "Trisomy 12 and structural abnormalities of 13q14 occurring in the same clone in chronic lymphocytic leukaemia." British Journal of Haematology 92: 389-392.

Navarro B., Garcio-Marco J.A., Jones D., Price C.M. and Catovsky D. (1998). "Association and clonal distribution of trisomy 12 and 13q14 deletions in chronic lymphocytic leukamia." British Journal of Haematology 102: 1330-1334.

Nehls M., Pfeifer D. and Boehm T. (1994). "Exon amplification from complete libraries of genomic DNA using a novel phage vector with automatic plasmid excision facility: Application to the mouse neurofibromatosis-1 locus." Oncogene 9: 2169-2175.

O'Brien S., Giglio A. and Keating M. (1995). "Advances in the biology and treatment of B-cell chronic lymphocytic leukemia." Blood 85: 307-318.

Offit K., Louie D.C., Parsa N.Z., Filippa D., Gangi M., Siebert R. and Chaganti R.S.K. (1994). "Clinical and morphological features of B-cell small lymphocytic lymphoma with del (6)(q21q23)." Blood 83: 2611-2618.

Offit K., Parsa N.Z., Gaidano G., Filippa D.A., Louie D., Pan D., Jhanwar S.C., Dalla-Favera R. and Chagnati R.S.K. (1993). "6q deletions define distinct clinico-pathologic subsets of non-Hodgkin's lymphoma." Blood 82: 2157-2162.

Ogawa O., Eccles M.R., Szeto J., McNoe L.A., Yun K., Maw M.A., Smith P.J. and Reeve A.E. (193). "Relaxation of insulin-like growth factor II gene imprinting implicated in Wilms' tumour." Nature 362: 749-751.

Ohno H., Takimoto G. and McKeithan T.W. (1990). "The candidate proto-oncogene bcl-3 is related to genes implicated in cell lineage determination and cell cycle control." Cell 60: 991-997.

Orkin S.H., Antonarakis S.E. and Kazazian H.H. (1984). "Base substitution at position-88 in a β-thalassemic globin gene. Further evidence for the role of distal promoter element ACACCC." Journal of Biology and Chemistry 259: 152-154.

Oscier D.G., Stevens J., Hamblin T.J., Pickering R.M., Lambert R. and Fitchett M. (1990). "Correlation of chromosome abnormalities with laboratory features and clinical course in B-cell chronic lymphocytic leukaemia." British Journal of Haematology 76: 352-358.

Oscier D.G., Thompsett A., Zhu D. and Stevenson F.K. (1997). "Differential rates of somatic hypermutation in V_H genes among subsets of chronic lymphocytic leukemia defined by chromosomal abnormalities." Blood 89: 4153-4160.

Panayiotidis P., Ganeshaguru K., Hoffbrand A.V., Rowntree C., Jabbar S.A.B. and Foroni L. (1997). "Deletion of 13q14.3 and not 13q12 is the most common genetic abnormality detected in chronic lymphocytic leukemia cells." Blood 89: 734-735.

Parimoo S., Patanjali S.R., Shukla H., Chaplin D.D. and Weissman S.M. (1991). "cDNA selection: Efficient PCR approach for the selection of cDNAs encoded in large chromosomal DNA fragments." Proceedings of the National Academy of Science, USA 88: 9623-9627.

Perucho M. (1996). "Microsatellite instability: The mutator that mutates the other mutator." Nature Medicine 2: 630-631.

Perucho M., Goldfarb M.P., Shimizu K., Lama C., Pogh J. and Wigler M. (1981). "Human-tumor-derived cell lines contain common and different transforming genes." Cell 27: 467-471.

Pierce J.C., Sauer B. and Sternberg N. (1992). "A positive selection vector for cloning high molecular weight DNA by the becteriophage P1 system: Improved cloning efficiency." Proceedings of the National Academy of Science, USA 89: 2056-2060.

Ponder B. (1988). "Gene losses in human tumours." Nature 335: 400-402.

Rabbitts T.H. (1991). "Translocations, master genes, and the differences between the origins of acute and chronic leukaemias." Cell 67: 641-644.

Rabbitts T.H. (1994). "Chromosomal traslocations in human cancer." Nature 372: 143-149.

Rai K.R., Kipps T.J. and Barlogie B. (1996). "Chronic lymphocytic leukemia and myeloma: Update on the biology and management." American Society of Haematology - educational program : 62-73.

Rai K.R., Sawitsky A., Cronkite E.P., Chanana A.D., Levy R.N. and Pasternack B.S. (1975). "Clinical staging of chronic lymphocytic leukemia." Blood 46: 219-234.

Randhawa G.S., Cui H., Barletta J.A., Strichman-Almashanu L.Z., Talpaz M., Kantarijian H., Deisseroth A.B., Champlin R.C. and Feinberg A.P. (1998). "Loss of imprinting in disease progression in chronic myelogenous leukemia." Blood 91: 3144-3147.

Ranheim E.A. and Kipps T.J. (1993). "Activated T cells induce expression of B7/BB1 on normal or leukemic B cells through a CD40-dependent signal." Journal of Experimental Medicine 177: 925-935.

Rassenti L.Z. and Kipps T.J. (1993). "Lack of extensive somatic mutations in the VH5 genes used in common B cell chronic lymphocytic leukemia." Journal of Experimental Medicine : 1039-1046.

Ray P.N., Belfall B., Duff C., Logan C., Kean V., Thompson M.W., Sylvester J.E., Gorski J.L., Schmickel R.D. and Worton R.G. (1985). "Cloning of the breakpoint of an X;21 translocation associated with Duchenne muscular dystrophy." Nature 318: 672-675.

Riley J., Butler R., Ogilivie D., Finniear R., Jenner D., Powell S., Anand R., Smith J.C. and Markham A.F. (1990). "A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones." Nucleic Acids Research 18: 2887-2890.

Riordan J.R., Rommens J.M., Kerem B., Alon N., Rozmahel R., Grzelczak Z., Zielenski J., Lok S., Plavsic N., Chou J., Drumm M.L., Iannuzzi M.C., Collins

F.S. and Tsui L. (1989). "Identification of the cystic fibrosis gene: Cloning and characterisation of complementary DNA." Science 245: 1066-1073.

Robertson K.D. and Jones P.A. (1997). "Chromatin dynamics '97. Dynamic interrelationships between DNA replication, methylation, and repair." American Journal of Human Genetics 61: 1220-1224.

Rommens J.M., Iannuzzi M.C., Kerem B., Drumm M.L., Melmer G., Dean M., Rozmahel R., Cole J.L., Kennedy D., Hidaka N., Zsiga M., Buchwald M., Riordan J.R., Tsui L. and Collins F.S. (1989). "Identification of the cystic fibrosis gene: Chromosome walking and jumping." Science 245: 1059-1065.

Rondeau G., Moreau I., Bezieau S., Cadoret E., Moisan J.P. and Devilder M.C. (1999). "Exclusion of *Leu 1* and *Leu 2* genes as tumor suppressor genes in 13q14.3-deleted B-CLL." Leukemia 13: 1630-1632.

Royer-Pokora B., Kunkel L.M., Monaco A.P., Goff S.C., Newburger P.E., Baehner R.L., Cole F.S., Curnutte J.T. and Orkin S.H. (1986). "Cloning the gene for an inherited human disorder-chronic granulomatous disease-on the basis of its chromosomal location." Nature 322: 32-37.

Sambrook J., Fritsch E.F. and Maniatis T. (1989). Molecular Cloning: A Laboratory Manual. Cold Sping Harbour Laboratory Press.

Sanger F., Nicklen S. and Coulsen A.R. (1977). "DNA sequencing with chain terminating inhibitors." Proceedings of the National Academy of Science, USA 74: 5463-5467.

Sassaman D.M., Bombroski B.A., Moran J.V., Kimberland M.L., Naas T.P., DeBerardinis R.J., Gabriel A., Swergold G.D. and Kazazian H.H. (1997). "Many human L1 elements are capable of retrotransposition." Nature Genetics 16: 37-43.

Schaffner C., Stilgenbauer S., Dohner H. and Lichter P. (1998). "Mutation analysis of ATM in lymphoproliferative diseases reveals a pathogeneic role for ATM not only in T-PLL but also in B-CLL." Blood 92: Abstract 1274.

Schlessinger D. (1990). "Yeast artificial chromosome: Tools for mapping and analysis of complex genomes." Trends in Genetics 6: 248-258.

Schwartz D.C. and Cantor C.R. (1984). "Separation of yeast chromosome-sized DNAs by pulsed field gel electrophoresis." Cell 37: 67-75.

Shaw S.H., Farr J.E., Thiel B.A., Matise T.C., Weissenbach J., Chakaravarti A. and Richard III C.W. (1995). "A radiation hybrid map of 95 STSs spanning human chromosome 13q." Genomics 27: 502-510.

Sherr C.J. (1996). "Cancer cell cycles." Science 274: 1672-1677.

Sherrington R., Rogaev E.I., Liang Y., Rogaeva E.A., Levesque G., Ikeda M., Chi H., Lin C., Li G., Holman K., Tsuda T., Mar L., Foncin J.F., Bruni A.C., Montesi M.P., Sorbi S., Rainero I., Pinessi L., Nee L., Chumakov I., Pollen D., Brookes A., Sanseau P., Polinsky R.J., Wasco W., Da Silva H.A.R., Haines J.L., Rommens J.M. and St George-Hyslop P.H. (1995). "Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease." Nature 375: 754-760.

Shizuya H., Birren B., Kim U., Mancino V., Slepak T., Tachiri Y. and Simon M. (1992). "Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector." Proceedings of the National Academy of Science, USA 89: 8794-8797.

Southern E.M. (1975). "Detection of specific sequences among DNA fragments by gel electrophoresis." Journal of Molecular Biology 98: 503-.

Stankovic T., Weber P., Stewart G., Bedenham T., Murray J., Byrd P.J., Moss P.A.H. and Taylor A.M.R. (1999). "Inactivation of ataxia telangiectasia mutated gene in B-cell chronic lymphocytic leukaemia." Lancet 353: 26-29.

Starostik P., Manshouri T., O'Brien S., Freireich E., Kantarijian H., Haidar M., Lerner S., Keating M. and Albitar M. (1998). "Deficiency of the ATM protein expression defines an aggressive subgroup of B-cell chronic lymphocytic leukaemia." Cancer Research 58: 4552-4557.

Stilgenbauer S., Liebisch P., James M.R., Schroder M., Schlegelberger B., Fischer K., Bentz M., Lichter P. and Dohner H. (1996). "Molecular cytogenetic delineation of a novel critical genomic region in chromosome bands 11q22.2-23.1 in lymphoproliferative disorders." Proceedings of the National Academy of Science, USA 93: 11837-11841.

Stilgenbauer S., Nickolenko J., Wilhem J., Wolf S., Weitz S., Dohner K., Boehm T., Dohner H. and Lichter P. (1998). "Expressed sequences as candidates for a novel tumor suppressor gene at band 13q14 in B-cell chronic lymphocytic leukemia and mantle cell lymphoma." Oncogene 16(14): 1891-7.

Stilgenbauer S., Schaffner C., Litterst A., Liebisch P., Gilad S., Bar-Shira A., James M.R., Lichter P. and Dohner H. (1997). "Biallelic mutations in the ATM gene in T-prolymphocytic leukemia." Nature Medicine 3: 1155-1159.

Straughen J., Ciocci S., Ye T., Lennon D.N., Proytcheva M., Alhadeff B., Goodfellow P., German J., Ellis N.A. and Groden J. (1996). "Physical mapping of the Bloom syndrome region by identification of YAC and P1 clones from human chromosome 15 band q26.1." Genomics 35: 118-128.

Surono A., Takeshima Y., Wibawa T., Pramono Z.A.D. and Matsuo M. (1997). "Six novel transcripts that remove a huge intron ranging from 250 to 800 kb are produced by alternative splicing of the 5' region of the dystrophin gene in human skeletal muscle." Biochemical and Biophysical Research Communications 239: 895-899.

Takeda K., Ichijo H., Fujii M., Mochida Y., Saitoh M., Nishitoh H., Sampath T.K. and Miyazano K. (1998). "Identification of a novel bone morphogenetic protein-responsive gene that may function as a noncoding RNA." The Journal of Biological Chemistry 273: 17079-17085.

Tavtigian S.V., Simard J., Romens J., Couch F., Shattuck-Eidens D., Neuhasen S. *et al.* (1996). "The complete BRCA 2 gene and mutations in chromosome 13q-linked kindreds." Nature Genetics 12: 333-337.

Thibodeau S.N., Bren G. and Schaid D. (1993). "Microsatellite instability in cancer of the proximal colon." Science 260: 816-819.

Thliveris A., Samowitz W., Matsunami N., Groden J. and White R. (1994). "Demonstration of promoter activity and alternative splicing in the region 5' to exon 1 of the APC gene." Cancer Research 54: 2991-2995.

Thompson M.E., Jensen R.A., Obermiller P.S., Page D.L. and Holt J.T. (1995). "Decreased expression of BRCA 1 accelerates growth and is often present during sporadic breast cancer progression." Nature Genetics 9: 444-450.

Trask B., Pinkel D. and van den Engh G. (1989). "The proximity of DNA sequences in interphase cell nuclei is correlated to genomic distance and permits ordering of cosmids spanning 250 kilobase pairs." Genomics 5: 710-717.

Trask B.J. (1991). "Fluorescent in situ hybridisation: Applications in cytogenetics and gene mapping." Trends in Genetics 7: 149-154.

Trofatter J.A., MacCollin M.M., Rutter J.L., Murrell J.R., Duyao M.P., Parry D.M., Eldridge R., Kley N., Menon A.G., Pulaski K., Haase V.H., Ambrose C.M., Munroe D., Bove C., Haines J.L., Martuza R.L., MacDonald m.E., Seizinger B.R., Short M.P., Buckler A.J. and Gusella J.F. (1993). "A novel moesin-, ezrin-, radixin-like gene is a candidate for the neurofibromatosis 2 tumor suppressor." Cell 72: 791-800.

Valetto A., Ghiotto F., Fais F., Hashimoto S., Allen S.L., Lichtman S.M., Schulman P., Vinciguerra V.P., Messmer B.T., Thaler D.S., Ferrarini M. and Chiorazzi N. (1998). "A subset of IgG+ B-CLL cells expresses virtually identical antigen receptors that bind similar peptides. Evidence for antigen selection in the leukemogenic process." Blood 92: Abstract 1784.

Vaux D.L., Cory S. and Adams J.M. (1988). "Bcl-2 gene promotes haemopoietic cell survival and co-operates with c-*myc* to immortalize pre-B cells." Nature 335: 440-442.

Vogelstein B., Fearon E.R., Hamilton S.R., Kern S.E., Preisinger A.C., Leppert M., Nakamura Y., White R., Smits A.M. and Bos J.L. (1988). "Genetic alterations during colorectal-tumor development." New England Journal of Medicine 319: 525-32.

Vogelstein B. and Kinzler K.W. (1998). The genetic basis of human cancer. McGraw-Hill.

Wahlfors J., Hiltunen K., Heinonen N., Hamalainen E., Alhonen L. and Janne J. (1992). "Genomic hypomethylation in human chronic lymphocytic leukemia." Blood 80: 2074-2080.

Wang Q. and Keating M.T. (1994). "Isolation of P1 insert ends by direct sequencing." BioTechniques 17: 282-284.

Watson J.D., Giman M., Witowski J. and Zoller M. (1992). Recombinant DNA. New York, Scientific American Books.

Weinberg R.A. (1991). "Tumor Suppressor Genes." Science 254: 1138-1146.

Weiner A.M., Deininger P.L. and Efstratiadis A. (1986). "Nonviral retroposons: Genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information." Annual Review of Biochemistry 55: 631-661.

Wickremasinghe R.G. and Hoffbrand A.V. (1999). "Biochemical and genetic control of apoptosis: Relevance to normal hematopoiesis and hematological malignancies." Blood 93: 3587-3600.

Willis T.G., Zalcberg I.R., Coignet L.J.A., Wlodorska I., Stul M., Jadayel D.M., bastard C., Treleaven J.G., Catovsky D., Silva M.L.M. and Dyer M.J.S. (1998). "Molecular cloning of translocation t(1;14)(q21;q23) defines a novel gene (BCL 9) at chromosome 1q21." Blood 91: 1873-1881.

Yuille M.R., Houlston R.S. and Catovsky D. (1998). "Anticipation in familial chronic lymphocytic leukemia." Leukemia 12: 1696-1698.

Zhao N., Stiffel A., Wang P.W., Eisenbart J.D., Espinosa III R., Larson R.A. and Le Beau M.M. (1997). "Molecular delineation of the smallest commonly deleted region of chromosome 5 in malignant diseases to 1-1.5 Mb and preparation of a PAC-based physical map." Proceedings of the National Academy of Sciences, USA 94: 6948-6953.