

Open Research Online

The Open University's repository of research publications and other research outputs

A molecular genetic investigation of rhabdomyosarcoma

Thesis

How to cite:

Chalk, Jeremy (1998). A molecular genetic investigation of rhabdomyosarcoma. PhD thesis The Open University.

For guidance on citations see [FAQs](#).

© 1998 The Author

Version: Version of Record

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

UNRESTRICTED

A MOLECULAR GENETIC INVESTIGATION OF RHABDOMYOSARCOMA

by

Jeremy Chalk

A thesis submitted for the degree of

Doctor of Philosophy,

The Open University.

December, 1997.

The Institute of molecular medicine

Department of Paediatrics

John Radcliffe Hospital

Headington

Oxford OX3 9DU.

Author no. P9275976
Date of award 27th May 1998

ABSTRACT

Alveolar Rhabdomyosarcoma is characterised by a t(2;13)(q35;q14) chromosome translocation, which leads to the fusion of the PAX3 and the FKHR genes. The resulting fusion gene encodes a chimeric protein which has aberrant transcriptional activity. The data here describes the molecular definition of the genomic breakpoints on both derivative chromosomes in one case and the derivative chromosome 13 breakpoints in two other cases. The DNA sequences adjacent to the breakpoints on the derivative chromosome 13 are remarkable for their resemblance to recognition sequences for the protein translin. Electrophoretic mobility shift studies (EMSA) confirm that these sequences bind translin. These findings suggest that translin may not only be important in the genesis of chromosomal translocations in lymphoid malignancy, but also in translocations found in solid tumours.

Mutation analysis of tumour samples and cell lines from patients with embryonal and alveolar rhabdomyosarcoma suggests that there are no subtle disease associated mutations within the PAX3 gene that could contribute towards the neoplastic state.

ACKNOWLEDGEMENTS

I would first of all like to thank my supervisor and director of studies Dr. Chris Mitchell for his advice, support, enthusiasm and patience throughout the last few years. Thanks also to my other supervisors Dr. Ann Harris and Dr. Sue Malcolm, both of whom provided important objective and critical suggestions. I would also like to thank Dr. Rick Womer (Children's Hospital of Philadelphia) for providing tumour samples and the idea that led to the work described in chapter five of this thesis.

At a more technical level I am grateful to Dr. Tarra McDowell and Dr. Karl Morton for providing insight into the idiosyncrasies of several techniques that I used. Thanks to Dr. Fred Barr (University of Pennsylvania) for providing mapping data and genomic DNA for a number of t(2;13) bearing cell lines and to Dr. Janet Shipley for sending me an RMS cell line.

Finally, thanks to OUCCC, MRCFC and my friends for helping me to get here.

TABLE OF CONTENTS

Title	1
Abstract	2
Acknowledgements	3
Contents	4
List of figures	9
List of tables	10
Abbreviations	11
Publications	12
Chapter 1 Introduction	13
1.1 Cancer genetics	13
1.2 Chromosome instability	14
1.3 Rhabdomyosarcoma	15
1.4 Cytogenetic abnormalities in rhabdomyosarcoma	16
1.5 Identification of genetic loci disrupted by the (2;13)	19
1.6 The pax gene family	21
1.6.1 pax1 and pax9	23
1.6.2 pax2,pax5 and pax8	23
1.6.3 pax4 and pax6	24
1.6.4 pax3 and pax7	25
1.7 The forkhead gene family	28
1.8 PAX3/FKHR fusion protein function	30

1.9 Causes of translocations	34
1.10 Aims of this project	37
Chapter 2. Materials and Methods	38
2.1. Chemical reagents	38
2.2 Enzymes	38
2.3 Nucleotides and Oligonucleotides	38
2.4 DNA size markers	39
2.5 Hybridisation membranes	39
2.6 X-ray film	39
2.7 Bacterial hosts	39
2.8 Cell lines	39
2.9 Cosmid libraries	41
2.10 Media	41
2.11 Solutions	42
2.12 Restriction endonuclease digestion of DNA	46
2.12.1 Plasmid and Cosmid DNA	46
2.12.2 Genomic DNA	46
2.13 Preparation of DNA from agarose gel slices	47
2.14 Polymerase chain reaction (PCR) methodology	48
2.14.1 Oligonucleotide synthesis	48
2.14.2 Basic PCR protocol	48
2.14.3 Cloning of PCR products	49
2.14.4 SSCP-PCR	51

2.15	Radiolabelling probes	51
2.16	Cosmid library plating and filter lifts	52
2.17	Southern blotting	53
2.18	Screening of DNA-bound filters by hybridisation	53
2.19	Isolation of plasmid DNA	54
2.20	Isolation of cosmid DNA	55
	2.20.1 Small scale	55
	2.20.2 Large scale	55
2.21	Sequencing	56
2.22	Preparation of genomic DNA	57
2.23	Whole cell protein extract preparation	58
2.24	Protein concentration assay	58
2.25	End-labelling of oligonucleotides	59
2.26	Electrophoretic mobility shift assay (EMSA)	59

**Chapter 3. Mutation analysis in the PAX3 gene in Embryonal
and Alveolar Rhabdomyosarcoma** **61**

3.1	Mutation detection methods	62
-----	----------------------------	----

**Chapter 4. Cloning of chromosome translocation breakpoints
using cosmids** **69**

4.1	Isolation of RHF14 cosmids	70
4.2	Isolation of normal chromosome 2 and 13 cosmids	72
4.3	Cosmid to cosmid hybridisation	72

Chapter 5. Cloning translocation breakpoints by vectorette

PCR 73

- 5.1 Cloning of the derivative chromosome 13 translocation
breakpoint in Rh28 73
- 5.2 Cloning of chromosome 13 sequences which span the
translocation breakpoint in Rh28 75
- 5.3 Cloning of the derivative chromosome 2 translocation breakpoint
in Rh28 76
- 5.4 Cloning the derivative 13 breakpoints in other t(2;13) bearing
cell lines 77
- 5.5 Cloning of normal chromosome 13 sequences spanning the
breakpoints in cell lines Rh5 and TTC487 79
- 5.6 Analysis of sequences around the translocation breakpoints 80

Chapter 6. Protein binding studies 83

- 6.1 Binding of protein from Rhabdomyosarcoma cell lines to
oligonucleotides 85
- 6.2 Specificity of oligonucleotide-protein binding 87
- 6.3 Sizing of protein that binds to oligonucleotides 88
- 6.4 Effect of oligonucleotide mutation on oligonucleotide-protein
complex formation 89
- 6.5 Effect of anti-translin antibody on oligonucleotide-protein
complex formation 90

Chapter 7. Discussion	92
7.1 Cloning chromosome translocation breakpoints	92
7.2 Interaction of proteins with translocation breakpoint sequences	94
7.3 Double-strand break (DSB) repair and a model for the generation of chromosome translocations	95
7.4 The t(2;13) chromosome translocation and position effect mutations	98
7.5 Mutation analysis in the PAX3 gene in Embryonal and Alveolar Rhabdomyosarcoma	100
7.6 Further studies	101

Chapter 8. References

GLOSSARY OF ABBREVIATIONS.

ATP	adenosine 5'-triphosphate
arms	alveolar rhabdomyosarcoma
bp	base pair
cpm	counts per minute
CTP	cytidine 5'-triphosphate
DNA	deoxyribonucleic acid
dNTPs	deoxyribonucleoside 5'-triphosphate
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EMSA	Electrophoretic mobility shift assay
erns	Embryonal rhabdomyosarcoma
FKHR	Forkhead in Rhabdomyosarcoma
GTP	guanosine 5'-triphosphate
Ig	immunoglobulin
kb	kilobase pair
LINEs	long interspersed repetitive elements
OD	optical density
PAX	Paired box
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PMSF	phenylmethyl sulfonylflouride
rpm	revolutions per minute
SDS	sodium dodecyl sulphate

SSC	standard saline citrate
SSCP	single strand conformation polymorphism
TAE	tris-acetate/EDTA
TBE	tris-borate/EDTA
TE	tris/EDTA
Tris	tris(hydroxymethyl)aminomethane
TTP	thymidine 5'-triphosphate
uv	ultraviolet

PUBLICATIONS

Chalk,J., Barr,F.G., Mitchell,C.D.(1997). Translin recognition site sequences flank chromosome translocation breakpoints in alveolar rhabdomyosarcoma cell lines.

Oncogene, in press.

CHAPTER ONE INTRODUCTION

1.1 CANCER GENETICS

“Each cancer is the result of an independent evolutionary process at the somatic cell level, involving the mechanisms of mutation and selection but without the intervention of a sexual process. The accumulation of the successive mutations that result in a cancer must take place stepwise, with each mutation having a selective advantage leading to an expanded cellular population within which the next mutation, giving rise to a further selective advantage and further expansion, takes place. Thus, it is the selection rather than the mechanism of mutation which is important in the genesis of cancer. Once a cellular population has escaped from the constraints of normal growth control, malignancy is established. The stepwise nature of this model for cancer indicates that it is a time-dependent process, and so the incidence of cancer increases with age. The clue to understanding cancer is to define the set of individual genetic steps that constitute the somatic evolutionary process and, through understanding their functions, to define the nature of the selective advantages associated with each step” (Bodmer et al, 1994).

There are two classes of genes whose mutation may result in cancer;

The first class are the *Oncogenes* which are defined as genes which, when acting alone or in combination with other genes, can induce the transformation of cells in tissue culture. The first oncogenes to be discovered were those found to be homologous to the genes of viruses which are capable of inducing tumourigenesis in normal animal and human cells (Bishop 1981, 1987). Oncogenes code for proteins such as growth factors, growth factor receptors, protein kinases or transcription factors. The discovery of proto-oncogenes has been followed by the discovery of a number of mechanisms which result in their aberrant activation. These include gene amplification and DNA rearrangements such as chromosome translocation (see below).

The second class of genes are termed the *tumour suppressor genes*. Both alleles of a tumour suppressor gene have to be inactivated to allow tumourigenesis. The study of inherited predisposition to particular tumours has led to an understanding of what role tumour suppressor genes play in cancer. For example, retinoblastoma is a childhood malignancy

which occurs in both sporadic and familial forms. It was postulated by Knudsen (Knudsen, 1973) that in either case development of a tumour depends on two mutation events; in the familial form one mutation occurs in the germline and the other is a somatic mutation, whereas in the sporadic form both are somatic mutations. Studies on the tumour suppressor gene *Rb* in patients with retinoblastoma have shown that this hypothesis is correct (Friend et al, 1986). Loss of function of tumour suppressor genes are caused by point mutations, deletions, insertions or loss of whole chromosomes.

1.2 CHROMOSOME INSTABILITY.

Row in MO-ADINE MSA

The relationship between 'chromosome instability' and cancer has been apparent since the development of chromosome banding techniques in the sixties (Cassperson et al, 1968). It became clear that the majority of cancer cells had aberrant chromosome numbers and/or structural chromosome abnormalities and also that there was an association between specific abnormalities and particular types of cancer. The main observable cytogenetic changes are deletions and translocations/inversions. Many translocations have been cloned and some general principles have emerged from the study of the associated genes. Deletions often result in loss of a tumour suppressor gene.

There are two consequences of translocations;

Firstly, if a gene for a T-cell receptor or an immunoglobulin is juxtaposed to a proto-oncogene, the latter becomes activated. The c-MYC gene translocation in Burkitt's lymphoma is a good example of the activation of a proto-oncogene by proximity to a T-cell receptor or immunoglobulin gene. T-cell receptor or immunoglobulin genes are frequently involved in chromosome aberrations because they are naturally rearranged to generate active antigen-receptor genes. This process occasionally, in error, leads to an interchromosomal translocation or intrachromosomal inversion, as in Burkitt's lymphoma. c-MYC can dimerize with another protein, MAX, to become transcriptionally active. MAX can also dimerize with other proteins so that an equilibrium in the amount of each of these proteins is found in the normal cell. After chromosome translocation, this equilibrium is disrupted by

overexpression of c-MYC and results in transcription of down stream target genes, leading to oncogenesis (Blackwood and Eisenman, 1991; Prendergast et al, 1991; Amati et al, 1993).

Secondly, if the breaks occur within a gene on each chromosome involved, a fusion gene is created encoding a chimeric protein. The genes involved often encode transcription factors, indicating that altered transcription plays a major part in tumourigenesis. The first constant rearrangement to be identified was the Philadelphia chromosome in chronic myeloid leukaemia (CML), subsequently characterised as a reciprocal translocation between the long arms of chromosomes 9 and 22, t(9:22)(q33;q21) (Nowell and Hungerford, 1960; Rowley, 1973). The fusion of the BCR and c-ABL genes on the Philadelphia chromosome typifies the situation in which breakage on each chromosome occurs within the introns of genes, producing fusion genes and subsequently leading to expression of a fusion protein. Sequences from both genes are essential for the oncogenic activity of the fusion protein and results in enhanced tyrosine kinase activity (Sawyers et al, 1991). The BCR-ABL fusion protein is thought to exist in a complex with a protein, GRB-2. The BCR-ABL-GRB-2 complex is thought to be required for the activation of the Ras signalling pathway, a highly conserved pathway which transmits signals from extracellular growth factors to the nucleus and which is often aberrantly activated in neoplasia (Prendergast et al, 1991).

1.3 RHABDOMYOSARCOMA

Rhabdomyosarcoma (RMS) is the commonest form of soft-tissue sarcoma in childhood and represents between 5% and 8% of all cases of childhood cancer. The histological classification of RMS is based on their resemblance to normal fetal skeletal muscle (Horn and Enterline, 1958). The two main histological variants of RMS can be discerned, the embryonal type (ERMS), so called because of its resemblance to fetal striated muscle, and the alveolar type (ARMS), so called because of the presence of open spaces within the tumour which resemble the alveolar spaces of the lung (Raney, 1989). Typically, ERMS, which accounts for approximately 60% of cases, arises in younger children in central sites which include the head and neck region, genito-urinary tract and orbit. Histologically, the

tumour is characterised by variable numbers of malignant spindle and primitive round cells that may contain cross-striations typical of skeletal muscle. The alveolar subtype arises in older children or adolescents and primarily affects peripheral sites such as the limbs. This histological variant is characterised by the presence of fibrovascular septa that form alveolar-like spaces filled with primitive, poorly cohesive, monomorphous, malignant cells. Patients with tumours of alveolar histology are generally considered to have a worse clinical prognosis compared to those with embryonal rhabdomyosarcoma (Crist et al, 1990).

In addition to their resemblance to fetal skeletal muscle, rhabdomyosarcomas have been consistently shown to express transcripts for the MyoD family of muscle-specific regulatory factors (Davis et al, 1987; Olson, 1990; Weintraub et al, 1991; Li and Olson, 1992; Olson and Klein, 1994). Members of this family of transcription factors such as MyoD, myogenin and myf5 direct the co-ordinated expression of the skeletal myogenic programme which leads to terminal differentiation (Tapscott and Weintraub, 1991). It is, therefore, surprising that expression of these factors occurs in undifferentiated rhabdomyosarcomas. However, cell fusion studies suggest that endogenous rhabdomyosarcoma MyoD protein is not able to function as a transcription activator due to a deficiency in the tumour cells of an unknown co-operating co-factor (Tapscott et al, 1993). Although no correlation between tumour histology and the expression pattern of these transcription factors has been reported, the restriction of their expression to cells of myogenic lineage has proven a useful diagnostic tool to differentiate between rhabdomyosarcomas and other paediatric neoplasm's (Dias et al, 1990).

1.4 CYTOGENETIC ABNORMALITIES IN RHABDOMYOSARCOMA

Another important technique for the characterisation of childhood tumours has been cytogenetic analysis. Although random chromosomal aberrations are common to all neoplastic cells, it has been possible to identify non-random chromosomal rearrangements, particularly translocations, which are specific to certain tumour types. Cytogenetics not

only provides a reliable way of grouping seemingly dissimilar forms of the same disease, but has created a starting point for the positional cloning of important genes located at the sites of the rearrangements.

Cytogenetic analysis of rhabdomyosarcoma has demonstrated chromosomal abnormalities in both embryonal and alveolar histologies. The translocation $t(2;13)(q35;q14)$ was first described by Seidal in 1982 and was noted to occur specifically in cases of alveolar rhabdomyosarcoma (see figure 1.1; Seidal et al, 1982; Douglass et al, 1987). Subsequent reports have confirmed this association in over 70% of successfully karyotyped alveolar rhabdomyosarcomas and it is now considered specific for tumours of this histology (Whang-Peng et al, 1992). Additionally, a less frequent variant translocation $t(1;13)(p36;q14)$ has been reported, which occurs in 10-15% of tumours of alveolar histology (Beigel et al, 1991; Douglass et al, 1991; Whang-Peng et al, 1992). Both translocations appear to involve the same region on chromosome 13q14. Evidence has been found to suggest that patients with the $t(1;13)$ have different clinical and pathological features from those with the common $t(2;13)$ translocation (personal communication, Dr. Rick Womer).

In addition to structural abnormalities, flow cytometry studies have indicated that approximately two thirds of alveolar tumours have a near tetraploid DNA content, whereas this finding is rarely observed in embryonal rhabdomyosarcoma (Shapiro et al, 1991; Pappo et al, 1993; Wijnaendts et al, 1993; Niggli et al, 1994).

A number of differing chromosomal abnormalities have been reported for embryonal rhabdomyosarcoma such as deletion of chromosome 1p with hyperdiploidy and ring chromosome 13 (Potluri and Gilbert, 1985; Voullaire et al 1991; Olegard et al, 1992; Whang-peng et al, 1992). Perhaps the only consistently observed abnormality is trisomy of chromosome 2 (Biegel et al, 1995). Approximately two-thirds of embryonal rhabdomyosarcomas have hyperdiploid DNA content; the remaining embryonal

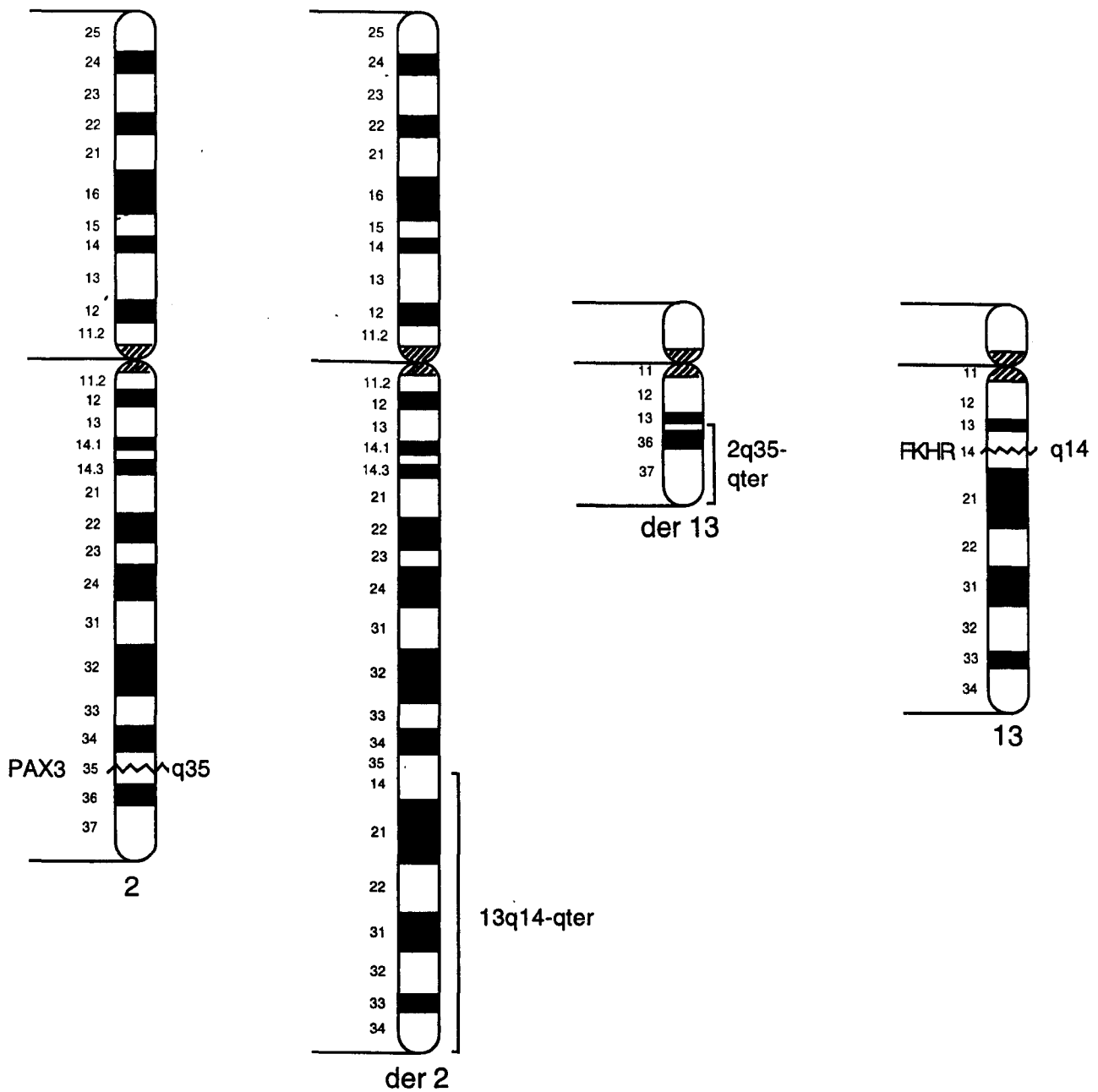


Figure 1.1 Ideograms of the normal chromosomes 2 and 13, and of the derivative chromosomes der2 and der13 specifically found in alveolar rhabdomyosarcoma. The PAX3 gene locus is indicated on chromosome 2 at position q35 and the FKHR gene locus is indicated on chromosome 13 at position q14. Chromosome translocation between the normal chromosomes 2 and 13 at these loci generate the derivative chromosomes and result in the fusion of the PAX3 and FKHR genes.

rhabdomyosarcomas are diploid and appear to respond less well to chemotherapy (Shapiro et al, 1991; Pappo et al, 1993; Wijnaendts et al, 1993; Niggli et al, 1994).

Unlike alveolar rhabdomyosarcoma, no consistent translocations or inversions have been reported for embryonal rhabdomyosarcoma which could contribute to tumourigenesis (Potluri and Gilbert, 1985; Olegard et al, 1992). However, restriction fragment length polymorphism analysis of embryonal rhabdomyosarcoma has shown consistent loss of heterozygosity (LOH) through mitotic recombination for loci on chromosome 11p, thus implying the location of a tumour suppressor gene. This loss has not been observed for alveolar rhabdomyosarcoma (Scrabble et al, 1987, 1989). The smallest region affected in these cases encompasses 11p15.5-pter and includes the loci for the haemoglobin β gene cluster (HBBC), tyrosine hydroxylase (TH), H19, insulin (INS), insulin-like growth factor 2(IGF2) and the Harvey RAS sarcoma virus oncogene (HRAS) (Scrabble et al, 1990; Newsham et al, 1991). This region has also been implicated in the development of other embryonal tumours, including Wilm's tumour, adrenal carcinoma and hepatoblastoma, as well as tumours of the lung, bladder, ovary and breast (Koufos et al, 1985). Furthermore, the Beckwith-Wiedemann syndrome, an autosomal dominant syndrome consisting of generalized somatic hyperplasia and a predisposition for the development of embryonal tumours (including Wilm's tumour and rhabdomyosarcoma), has been mapped to the same region of LOH as that found in embryonal rhabdomyosarcoma (Henry et al, 1991,1993).

Evidence that a tumour suppressor gene is expressed in this region has been provided by the transfer of an intact chromosome 11, using microcell hybridisation, into an embryonal rhabdomyosarcoma having LOH for this region (Loh Jr et al, 1992). By selective retention of either the short or long arms of chromosome 11 in the microcell hybrids, loss of proliferative capacity was observed. Thus, in addition to the known suppressor locus at 11p15.5, these results indicate the existence of another tumour suppressor on 11q (Weissman et al, 1987).

The location of the 11p15.5 suppressor locus has been refined by microcell fusion with subchromosomal fragments containing small segments of chromosome 11p (Koi et al, 1993). These studies have localised the suppressor locus to a 4.5Mb region between the anonymous markers D11S719 and D11S724.

Studies investigating the parental origin of alleles in this genomic region in familial and sporadic cases of embryonal rhabdomyosarcoma have shown that isodisomic chromosome 11p alleles are consistently of paternal origin (Scrabble et al, 1989). Thus, genomic imprinting of the paternal allele in embryonal rhabdomyosarcoma may be an alternative first step to inactivation by mutation at the 11p15.5 suppressor locus. Similar preferential inheritance of parental alleles have been reported for a variety of other tumours including retinoblastoma, Wilm's tumour, and osteosarcoma (Sakai et al, 1991; Rainier et al, 1993; Tycko, 1994)

1.5 IDENTIFICATION OF GENETIC LOCI DISRUPTED BY THE t(2;13)

Standard positional cloning techniques were used initially to attempt to identify loci disrupted by the t(2;13) chromosome translocation in alveolar rhabdomyosarcoma. Studies sought initially to refine the location of the chromosome 2 and 13 breakpoints on physical maps of these chromosomes. The breakpoint in Rhabdomyosarcoma cell lines was positioned proximally on chromosome 13 to the markers D13S10, esterase D (ESD) and RB1, and distally to the markers D13S6 and D13S1 by *in situ* hybridisation to metaphase chromosomes (Valentine et al, 1989). Further refinement was possible by mapping studies using somatic cell hybrids retaining either the derivative chromosome 2 or derivative chromosome 13¹,

1. Derivative chromosomes are created by the chromosome translocation. In this case the derivative chromosome 2 is composed of the p-arm, centromere and proximal q-arm of chromosome 2 and the distal q-arm of chromosome 13. The derivative chromosome 13 is composed of the p-arm, centromere and proximal q-arm of chromosome 13 and the distal q-arm of chromosome 2.

and by the analysis of long-range physical maps of these regions by pulsed-field gel electrophoresis (Barr et al, 1991a, b,1992; Mitchell et al, 1991; Shapiro et al, 1992). No rearrangements in 14 candidate genes which mapped to this region could be detected, but mapping of the breakpoint region could be refined to a two megabase interval between markers D13S29 and TUBBP2 (Barr et al, 1991b; Shapiro et al, 1992). On chromosome 2 the breakpoint region was found to be flanked proximally by inhibin-alpha (INHA) and distally by intestinal alkaline phosphatase (ALP1), a genetic distance of at least 5cM (Barr et al, 1992).

The physical map of the breakpoint region on chromosome 2 was less detailed than that of the breakpoint region on chromosome 13. However, the murine *pax3* gene (see below), was mapped by linkage analysis to the region between the inhibin-alpha and intestinal alkaline phosphatase loci on proximal mouse chromosome 1, a segment which is syntenically conserved with the distal human 2q region (Schurr, 1990).

The human homologue of the murine *pax3* gene was therefore a candidate for involvement in the rearrangement caused by the t(2;13) translocation. Using a chromosome 2 somatic cell hybrid mapping panel, human PAX3 was positioned within the same interval on chromosome 2q as the alveolar rhabdomyosarcoma-specific t(2;13) breakpoint and was shown to be rearranged in tumour DNA from cell lines containing this translocation (Barr et al, 1993). Fine-mapping showed that all rearrangements occur, albeit in differing positions, within the same 20kb intron located between the last two exons of the PAX3 gene. In all cases these rearrangements result in the translocation of the 5' region of PAX3 to the derivative chromosome 13, and the 3' region of PAX3 to the derivative chromosome 2.

Isolation and characterisation of fusion cDNA clones has shown that the PAX3 rearrangement results in the creation, on the derivative chromosome 13, of a novel chimeric gene composed of 5' PAX3 sequences juxtaposed to 3' sequences derived from a gene

designated FKHR (ForKHead in Rhabdomyosarcoma) which is a member of the forkhead family of transcription factors, and which is located on chromosome 13q14 (see figure 1.2). On the reciprocal derivative chromosome 2, 5' FKHR sequences are juxtaposed to 3' PAX3 sequences. Analysis of several independently derived RMS cell lines by RT-PCR indicates that the t(2;13) translocation results in chimeric transcripts which are uniform in structure, consistent with the occurrence of translocation breakpoints within specific PAX3 and FKHR introns (Galili et al, 1993; Shapiro et al, 1993).

Although the t(2;13)(q35;q14) translocation has been found in the majority of cases of alveolar rhabdomyosarcoma, several cases with the variant (1;13)(p36;q14) translocation have been reported (Beigel et al, 1991; Douglass et al, 1991; Whang-Peng et al, 1992). The (1;13) has been shown to rearrange the PAX7 gene on chromosome 1 and fuse it to identical FKHR sequences on chromosome 13 (Davis et al, 1994). The PAX3 and PAX7 genes have very similar expression patterns and functional elements (see below) so the respective PAX3-FKHR and PAX7-FKHR fusion proteins are likely to play similar roles in rhabdomyosarcoma tumourigenesis.

With the characterisation of the loci affected by the t(2;13) translocation in alveolar rhabdomyosarcoma complete, studies could be directed towards an understanding of how the PAX3, FKHR and novel chimeric PAX3-FKHR genes could contribute to the aetiology of alveolar rhabdomyosarcoma.

1.6 THE PAX GENE FAMILY

To date nine PAX genes have been characterised which constitute a family of developmental control genes which encode transcription factors containing the characteristic DNA binding domain termed the paired-box (Gruss and Walther, 1992). The nine mouse (*pax1* to *pax9*) and nine human (PAX1 to PAX9) genes have had their chromosomal locations mapped (Walther et al, 1991; Wallin et al, 1993; Stapleton et al, 1993). The paired box was first found in five *Drosophila* genes *paired*, *gooseberry* and *gooseberry neuro* which are expressed in early development, and *pox meso* and *pox neuro* which have a tissue restricted

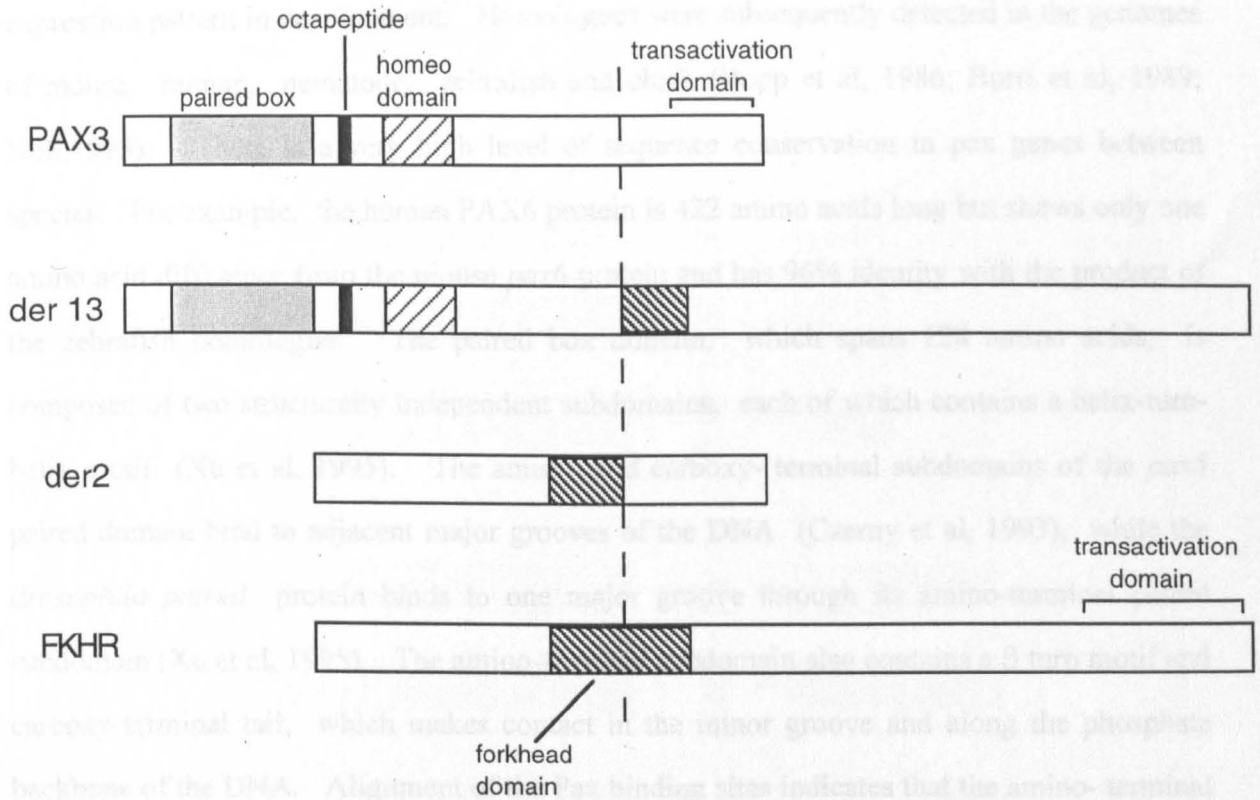


Figure 1.2 Schematic diagram depicting the consequences of the t(2;13)(q35;q14) chromosome translocation in alveolar rhabdomyosarcoma. The PAX3 gene contains two DNA-binding motifs, the paired box and the paired type homeodomain, an octapeptide motif of unknown function and a C-terminal transactivation domain. The FKHR gene contains the 'winged' helix-loop-helix forkhead DNA-binding domain and a C-terminal transactivation domain. The chimeric PAX3/FKHR gene on the derivative chromosome 13 contains the two PAX3 DNA-binding motifs and a truncated forkhead DNA-binding domain from the FKHR gene. The other chimeric gene on the derivative chromosome 2 does not contain an intact DNA-binding domain so is unlikely to have any activity. The important difference between the PAX3 gene and the PAX3/FKHR chimeric gene on the derivative chromosome 13 maybe the transactivation domains. The chimeric protein may bind to the same DNA sequences as PAX3 but could transactivate different transcriptional machinery.

expression pattern in development. Homologues were subsequently detected in the genomes of mouse, human, nematode, zebrafish and chick (Bopp et al, 1986; Burri et al, 1989; Noll,1993). There is a very high level of sequence conservation in pax genes between species. For example, the human PAX6 protein is 422 amino acids long but shows only one amino acid difference from the mouse *pax6* protein and has 96% identity with the product of the zebrafish homologue. The paired box domain, which spans 128 amino acids, is composed of two structurally independent subdomains, each of which contains a helix-turn-helix motif (Xu et al, 1995). The amino- and carboxy- terminal subdomains of the *pax5* paired domain bind to adjacent major grooves of the DNA (Czerny et al, 1993), while the *drosophila paired* protein binds to one major groove through its amino-terminal paired subdomain (Xu et al, 1995). The amino-terminal subdomain also contains a β turn motif and carboxy-terminal tail, which makes contact in the minor groove and along the phosphate backbone of the DNA. Alignment of the Pax binding sites indicates that the amino- terminal subdomain contacts a conserved recognition sequence whereas the more divergent carboxy-terminal subdomain appears to play a role in site-specific DNA recognition (Czerny et al, 1993).

A number of PAX genes also contain a homeobox. This 'paired-type' homeodomain is similar to classic homeodomains featuring a helix-turn-helix motif. In addition, most of the genes contain a third highly conserved sequence, the octapeptide, which is located between the two DNA-binding domains. The function of the octapeptide motif has yet to be established.

The nine paired box genes in mouse and man are dispersed around the genome, unlike the clustering of the HOX genes. The genes can be classed into four subgroups based on their structural similarities and expression patterns (see figure 1.3).

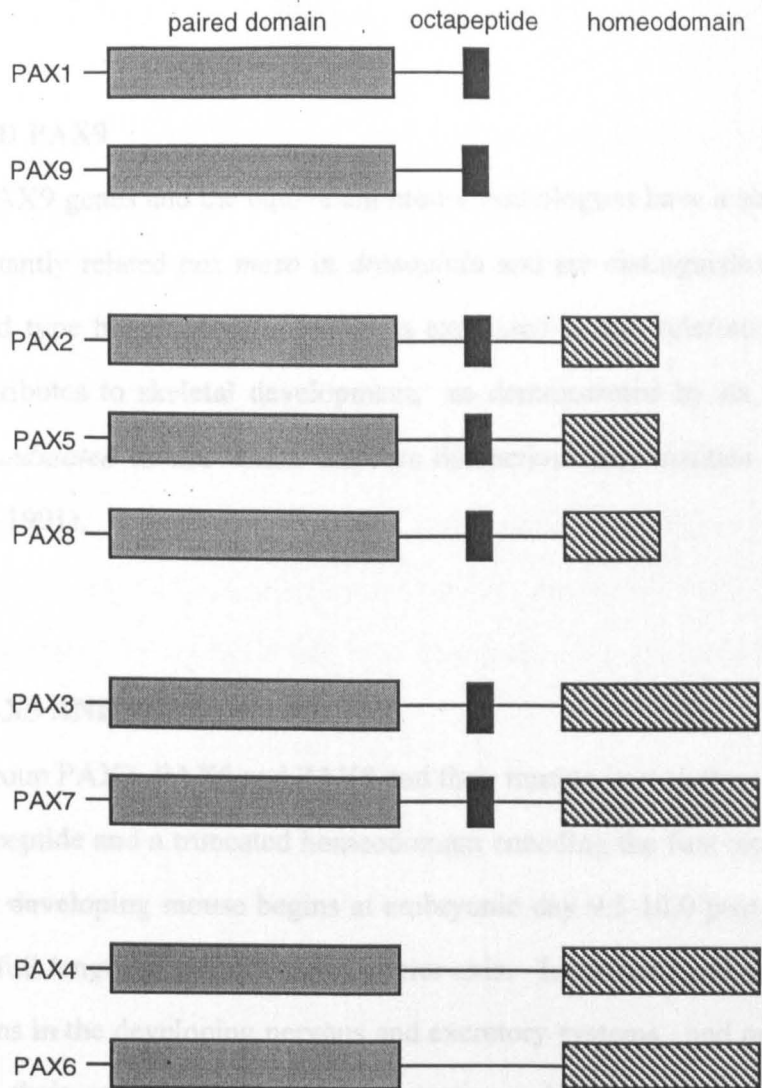


Figure 1.3 The PAX gene family. The nine known members of the PAX gene family can be divided into four subcategories depending on their structural motifs. The PAX1 and PAX9 genes have the paired domain DNA-binding motif and the octapeptide but lack the homeodomain. The PAX2, PAX5 and PAX8 genes contain the paired domain DNA-binding motif, the octapeptide and a truncated homeodomain. The PAX3 and PAX7 genes contain the full complement of motifs and the PAX4 and PAX6 genes contain the two DNA-binding motifs but no octapeptide.

1.6.1 PAX1 AND PAX9

The PAX1 and PAX9 genes and the equivalent mouse homologues have a similar expression pattern to the distantly related *pox meso* in *drosophila* and are distinguished by a complete lack of the paired type homeodomain. PAX1 is expressed in the sclerotomal parts of the somites and contributes to skeletal development, as demonstrated by its loss of function mutation in the *undulated* mouse which displays distinctive abnormalities of the vertebrae (Chalepakis et al, 1991).

1.6.2 PAX2, PAX5 AND PAX8

The second subgroup PAX2, PAX5 and PAX8 and their murine homologues possess a paired domain, an octapeptide and a truncated homeodomain encoding the first alpha-helix. Their expression in the developing mouse begins at embryonic day 9.5-10.0 post coitum (pc) and occurs along the full length of the anterior-posterior axis. Later *pax2* and *pax8* show similar expression patterns in the developing nervous and excretory systems, and *pax2* and *pax5* are distinguished by their expression in the adult testis. Also, *pax2* is expressed in the developing ear and eye, *pax5* in B lymphocytes in both embryo and adult and *pax8* in the developing and adult thyroid.

Of these genes PAX5 is the best studied. It is expressed in the developing central nervous system, in B-cells and in testis (Adams et al, 1992; Asano and Gruss, 1992). PAX5 (also known as BSAP) is an important regulator of B-cell development and some PAX5 target genes have been identified (Kozmik et al, 1992; Rothman et al, 1991; Singh and Birshstein, 1993; Waters et al, 1989). Transgenic mice homozygous for *pax5* have disturbed midbrain morphogenesis, abnormal spleen and lymphoid architecture and arrested B-cell development (Urbanek et al, 1994).

PAX5 expression has been demonstrated in certain human tumours including medulloblastoma and highly malignant astrocytomas (glioblastoma) where expression

appears to correlate with progression of malignancy (Kozmik et al, 1995; Stuart et al, 1995). Expression of PAX5 in astrocytomas coincided with the expression of recognised oncogenes (myc, fos, jun) but in glioblastomas expression inversely correlated with the tumour suppressor gene p53. Following these observations a direct role for PAX5 in the regulation of p53 transcription has been identified. PAX genes bind to a sequence within the 5' regulatory region of the human p53 gene and are capable of repressing p53 promoter activity in cell culture. The site to which PAX5 binds appears to be crucial for basal transcriptional activity and it is likely that PAX proteins interact or compete with a component of the basal transcriptional machinery for access to this site, thus controlling p53 transcription (Stuart et al, 1995).

A recurring translocation, $t(9;14)(p13;q32)$, has been identified in approximately 2% of non-Hodgkin lymphomas (lymphoplasmacytoid immunocytomas) which involves the PAX5 gene. In a cell line bearing this translocation the intact PAX5 gene has been juxtaposed to an immunoglobulin heavy chain gene promoter resulting in deregulation of its expression (Busslinger et al, 1996).

Transgenic mice which overexpress *pax2* have severe kidney abnormalities, suggesting that repression of *pax2* is required for normal kidney development (Dressler et al, 1993). Subsequently, loss-of-function mutations in the human PAX2 gene have been found in patients with kidney and retinal defects (Sanyanusin et al, 1995)

1.6.3 PAX4 and PAX6

A third subgroup of pax genes, PAX4 and PAX6, are distinguished by a lack of the octapeptide motif. However, it has not been possible to demonstrate expression of PAX4 in humans or its homologue in the mouse, suggesting that it may be a pseudogene. Murine *pax6* transcripts are detectable at embryonic day 8-8.5 pc and are found particularly in the developing central nervous system, in the nose and the developing eye. Expression in the adult is confined to the cerebellum. Mutations in *pax6* have been found in the mouse mutant

smalleye (sey). The variable phenotype in differing *sey* mutant alleles provides an indication of the importance of PAX gene dosage in normal development. In *sey* a point mutation in the coding region of *pax6* leads to a truncation of the protein product upstream of the homeodomain. Heterozygotes have eye malformations whereas homozygotes are neonatal lethal with a complete absence of eyes, nose and olfactory bulb. Mutations in the human homologue, PAX6, have been identified in several patients with aniridia, a condition which is characterised by the complete or partial absence of the iris, and also in patients with Peters anomaly who have defects in the cornea (Glaser et al, 1992; Jordan et al, 1992; Davis and Cowell, 1994; Hanson et al, 1993; Hanson et al, 1994). Overexpression of PAX6 in transgenic mice also appears to result in severe eye abnormalities (Schedl et al, 1996).

The apparent semidominance of PAX gene mutations suggests that these gene products are required in distinct concentrations. The transcriptional activity from a single allele is apparently below a critical threshold, leading to phenotypic alterations (Gruss and Walther, 1992).

1.6.4 PAX3 and PAX7

The fourth and final subgroup of the PAX gene family includes the PAX3 and the PAX7 genes. Both genes encode the full complement of conserved sequence motifs, the paired domain, the octapeptide and the complete paired type homeodomain. Expression of murine *pax3* transcripts can be detected in 8-8.5 pc embryos and occurs in the dorsal ventricular zone of the neural tube including roof plate and neural crest cells, as well as in dermomyotome, limb buds and some cranio-facial structures. Expression of *pax7* has a similar but not identical spatial and temporal pattern to that of *pax3*. Expression of both genes cannot normally be detected in adult tissue.

The *pax3* protein can bind *in vitro* to the e5 sequence from the promoter of the *drosophila even-skipped* gene. The *pax3* recognition site contains two DNA sequence blocks encompassing 18bp. An ATTA motif, recognised by homeodomain proteins in general,

forms part of the *pax3* homeodomain recognition sequence, and the second element includes GTTCC which was originally identified as the core motif for DNA binding by the *pax1* protein (Chalepakis et al, 1991; Chalepakis et al, 1994). Both recognition elements are required for high affinity *pax3* binding (Goulding et al, 1991). Although the wild-type *pax3* protein apparently does not form homodimers, truncated *pax3* proteins can dimerize, and the region essential for dimerization has been mapped to the interval between the paired domain and the homeodomain (Chalepakis et al, 1994)

A number of mutations in the *pax3* gene have been identified in *spotch* mice which have neural crest deficiencies, neural tube defects and limb defects. As with *pax6*, *spotch* is semidominant. The phenotype is called *spotch* because heterozygotes have white spotting on their coats. Homozygotes die by day 13-14 of gestation with neural tube defects, deficiency of melanocytes, schwann cells, dorsal root ganglia, thyroid, thymus, parathyroid and limb muscle deficiencies. In heterozygotes with the *spotch retarded* allele there is retarded growth in addition to the other abnormalities (Goulding et al, 1993; Epstein et al, 1993; Vogan et al, 1993).

Mutations in the PAX3 human homologue have been identified in patients with Waardenburg syndrome, a pleiotropic, dominantly inherited condition involving sensorineural hearing loss and pigmentary anomalies of the iris, hair and skin. Waardenburg syndrome has been subclassified into three types on the basis of the presence or absence of the clinical feature dystopia canthorum, an outward displacement of the inner corners of the eyes. Types 1 and 3 have dystopia and show linkage to PAX3, whereas none of the type 2 patients show linkage to PAX3. Type 2 patients have been linked to mutations in the *microphthalmia* gene (Tassabehji et al, 1994; Steingrimsson et al, 1994). To date, more than forty mutations in the PAX3 gene have been reported which involve changes in the paired and homeo domains of the gene (Epstein et al, 1991; Tassabehji et al, 1992; Baldwin et al, 1992; Morell et al, 1992; Tsukamoto et al, 1992; Hoth et al, 1993; Pasteris et al, 1993; Butt et al, 1994; Tassabehji et al, 1994). Mutational heterogeneity does not account for the substantial

phenotypic variation that exists between and within families carrying the same mutation. However, the phenotypic concordance reported in a case of monozygotic twins with Waardenburg syndrome suggests that variation in the expression of the syndrome is genetically mediated by the polygenic background or by interactions with specific genetic modifiers (Pandya et al, 1996).

Several lines of evidence suggest that mutations in *pax3*, as well as mutations in other *pax* genes, cause loss-of-function of the mutant allele, thus causing haplo-insufficiency. First, missense mutations and small in-frame deletions in an allele which allow expression of the protein product have a similar phenotype to those resulting in whole gene deletion or a frame shift early in the reading frame. This finding suggests that expression of the mutant allele in heterozygotes does not interfere with the function of the wildtype allele, giving a dominant negative effect. Second, *spotch* heterozygotes bearing one wildtype allele, have a less severe phenotype than homozygotes bearing only mutant alleles. Cases homozygous for Aniridia and Waardenburg syndrome support this observation (Hodgson and Saunders, 1980; Zlotogora et al, 1995). Finally, *in vitro* DNA-binding studies with mutant *pax3* and PAX3 alleles support a loss-of-function model (Chalepakis et al, 1994).

Studies involving the overexpression of several of the murine *pax* genes in tissue culture cells and nude mice demonstrate that they can promote oncogenesis and should thus be classified as a novel group of proto-oncogenes. The induction of tumour formation in mice is dependent on a functional paired domain but does not require the presence of a homeodomain (Maulbecker and Gruss, 1993)

1.7 THE FORKHEAD GENE FAMILY

The first forkhead gene (*fkh*) was isolated from *drosophila*, a homeotic protein that is essential for gut differentiation. Mutations within the *fkh* gene result in the duplication of head structures at the posterior portion of the gut (Weigel et al, 1989). Subsequently, hepatocyte nuclear factor 3 (HNF-3), which mediates liver-specific transcription of the transthyretin (TTR) gene, was isolated from rat liver (Costa et al, 1990). The gene coding for this product was cloned and found to have significant homology to the *fkh* gene of *drosophila*. HNF-3 is involved in regulating the expression of several liver-specific genes, such as alpha-1-antitrypsin, albumin, phosphoenolpyruvate carboxykinase (PEPCK), tyrosine aminotransferase and apolipoprotein B (Ang et al, 1993). Truncated polypeptides produced by deletion of HNF-3 cDNA templates have been used in DNA-binding studies to define the DNA-binding domain (Lai et al, 1990). This 300 nucleotide region, in the middle of the gene, has the highest homology with the *drosophila fkh* gene, providing evidence that these two genes were the prototypes of a new family of transcription factors.

HNF-3 has been renamed HNF-3 α following the cloning of two related liver factors, HNF-3 β and HNF-3 γ , from liver cDNA libraries which were probed with the conserved forkhead DNA-binding domain (Lai et al, 1991). X-ray crystallography studies of the HNF-3 γ protein bound to DNA indicate that family members bind as a monomer inducing a 13 degree bend in the DNA molecule. The DNA binding element has a 'winged helix' motif consisting of a helix-turn-helix and DNA interactive loops (Clark et al, 1993).

In addition to the conserved DNA-binding domain, the HNF-3 family of proteins possess homology in two C-terminal regions consisting of 19 and 14 amino acids and two larger N-terminal regions (Lai et al, 1991). Analysis of a series of HNF-3 β deletion mutants in transfection assays define these regions as transcription activation domains. The C-terminal domain requires the retention of the two conserved regions for activity, whereas the N-terminal domain requires the presence of the C-terminal domain, suggesting a physical interaction between these motifs (Pani et al, 1992). There is less homology between HNF-3

genes in these domains than in the forkhead domain, suggesting that they may interact with distinct targets of the transcriptional machinery.

Murine homologues of the HNF-3 α and HNF-3 β genes are expressed initially at embryonic day 6.5 pc during gastrulation and at day 8.5 pc in the node, notochord and floorplate of the neurotube. Later in development HNF-3 α and HNF-3 β are expressed in the gut, pancreas, liver and lung primordium, with HNF-3 γ expression restricted to gut, pancreas and liver (Sasaki and Hogan, 1993; Monaghan et al, 1993). Between days 9.5-15 pc HNF-3 gene expression occurs in the developing adrenal gland, thyroid gland and oral cavity, in bone tissue of the vertebrae, limbs, ribs, jaws and skull (Monaghan et al, 1993). Forkhead genes in the adult tends to show a tissue specific expression pattern

To date there are more than 40 members of the forkhead family of transcription factors based on the presence of the forkhead DNA-binding domain in species ranging from yeast to man. The winged-helix DNA-binding domain sequence is highly variable between family members, with amino acid identity to the HNF-3 α sequence as little as 31%. This heterogeneity may provide each protein with distinct DNA-binding properties and account for the wide range of tissue specific functions of forkhead genes in development and in the adult (Hromas and Costa, 1995).

As stated above, the novel FKHR gene sequences fused to PAX3 gene sequences in the t(2;13) translocation in alveolar rhabdomyosarcoma showed significant homology to the forkhead family of transcription factors. Northern analysis with FKHR sequences detect a 6.5kb wildtype sequence which is ubiquitously expressed in adult tissues (Galili et al, 1993). The forkhead DNA-binding domain of FKHR lacks an N-terminal KPPY amino acid sequence common to most forkhead genes and contains a novel five amino acid insert in the middle of the forkhead domain. Additionally, FKHR shows no region of homology with other forkhead genes at the C- and N-terminus. Other potential motifs based on sequence analysis include a 10 amino acid proline rich domain, which is similar to an SH3 binding site

(Ren et al, 1993). An alanine rich sequence, which has been associated with transcriptional repression, and proline rich and acidic regions, which may function as transcriptional regulatory domains (Han and Manley, 1993; Mitchell and Tijan, 1989). Expression studies with the chimeric PAX3/FKHR product indicate that there are transcriptional activation domains in the 3' region of the FKHR gene (see below).

Two other examples of a role for forkhead genes in neoplasia have been reported. Firstly, identification of the avian retroviral oncogene *qin* as a member of the forkhead family gave the first indication that the mammalian forkhead genes may have oncogenic potential (Li and Vogt, 1993). Secondly, a t(X;11) chromosome translocation in a cell line established from an infant with acute lymphoblastic leukaemia has been cloned and sequenced (Parry et al, 1994). The gene on the X-chromosome, AFX1, is a novel member of the forkhead gene family and has most homology to the FKHR gene. The fusion point in AFX1 is at exactly the same position in the forkhead domain as in the FKHR gene in the t(2;13) translocation.

Following the characterisation of the t(2;13) translocation, the structure of the wild-type FKHR gene on chromosome 13 was analysed (Davis et al, 1995). The FKHR gene consists of three exons spanning 140kb, the first of which is within a CpG island. All t(2;13) translocation breakpoints must occur within the large 130kb intron 1 of FKHR, which provides a large target for DNA rearrangement.

1.8 PAX3-FKHR AND PAX7-FKHR IN ALVEOLAR RHABDOMYOSARCOMA

The t(2;13)(q35;q14) translocation in alveolar rhabdomyosarcoma causes the fusion of the PAX3 and the FKHR genes (Barr et al, 1993). The variant (1;13)(p36;q14) translocation causes the fusion of the PAX7 gene on chromosome 1 to identical FKHR sequences on chromosome 13 (Davis et al 1994). Functional studies which might lead to an understanding of the role of the chimeric fusion products in alveolar rhabdomyosarcoma have been directed at the common PAX3-FKHR fusion. However, since the spatial and temporal expression

patterns and the structure of both PAX3 and PAX7 are so similar, it would be reasonable to infer that the PAX7-FKHR fusion has a similar role in tumourigenesis as the PAX3-FKHR fusion.

Initially, to establish that a mature chimeric protein product is present in t(2;13) bearing cells, polyclonal antisera specific for the normal PAX3 and FKHR proteins have been prepared (Fredericks et al, 1995). In the RD cell line, which does not contain the t(2;13) translocation, the PAX3 antiserum could detect a 56kD protein and the FKHR antiserum could detect a 72kD protein. These sizes are consistent with the predicted open reading frames of the wildtype genes. Immunoprecipitation of proteins from t(2;13) bearing cell lines showed that both antisera detected a 97kD protein which could not be detected in the RD cell line. The molecular mass of 97kD agrees with the predicted 837 amino acid fusion protein. Low levels of the 56kD wildtype PAX3 protein could also be detected but despite the presence of wildtype FKHR mRNA the 72kD FKHR protein was not detectable.

To confirm that both antisera are reacting with the same chimeric protein sequential immunoprecipitation experiments were performed. These demonstrated that the 97kD protein could be precipitated with one antiserum, dissociated from the complex, and reprecipitated with the other antiserum (Fredericks et al, 1995).

Several observations indicate that it is the derivative chromosome 13, 97kD fusion protein, which is important to the aetiology of alveolar rhabdomyosarcoma. Firstly, the reciprocal derivative chromosome 2 product would lack both the PAX3 paired domain and homeodomains. Secondly, based on functional and structural studies demonstrating the essential nature of the amino terminal of the forkhead domain for DNA binding, it is unlikely that the disrupted carboxyl FKHR forkhead domain present in the fusion protein contributes to sequence-specific DNA interactions (Clark et al, 1993; Clevidence et al, 1993). Thirdly, the derivative chromosome 2 transcript has only been detected by RT-PCR analysis in some of the t(2;13) bearing cell lines, whereas the derivative chromosome 13 product has always

been detected (Galili et al, 1993; Shapiro et al, 1993). These observations are consistent with northern analysis of tumour cell RNA which has indicated that, in those tumours expressing both chimeric transcripts, the derivative chromosome 13 transcript is significantly more abundant. Finally, only the 97kD fusion protein from the derivative chromosome 13 has been detected by immunoprecipitation (Fredericks et al, 1995).

Since many transcription factors bind DNA in a sequence specific manner, studies have been performed to determine whether the PAX3-FKHR fusion protein will bind to similar target sequences as wildtype PAX3 protein. PAX3 is able to bind *in vitro* to a DNA sequence derived from the *drosophila* even-skipped gene promoter, e5, which contains distinct domains that are specifically recognised by the paired box and the homeodomain (Goulding et al, 1991; Chalepakis et al, 1994). The PAX3-FKHR protein has been shown to bind to a radiolabelled e5 probe in an electrophoretic mobility shift assay, but not as effectively as wildtype PAX3 protein, despite the two proteins having identical PAX DNA-binding domains (Fredericks et al, 1995). In similar studies using the synthetic PRS-9 DNA target sequence, which is based on the e5 sequence, two different point mutations in the paired domain or deletion of the homeodomain in PAX3-FKHR abolish this binding (Sublett et al, 1995). These observations are consistent with studies originally performed on binding of the *pax-1* protein to the PRS-9 probe (Chalepakis et al, 1991).

Retention of the intact PAX3 paired box and homeodomain in the fusion protein and its ability to bind PAX3 target sequences suggest that it may bind to normal PAX3 genomic targets but aberrantly regulate transcription by either excessive activation or repression through novel FKHR 3' regulatory sequences. This hypothesis is supported by experiments which demonstrate that the fusion protein is capable of binding to and transcriptionally activating model PAX3 binding sites. Transient cotransfection assays using e5-CAT reporter plasmids indicate that the PAX3-FKHR fusion protein is a more potent transcription activator than PAX3 (Fredericks et al, 1995; Sublett et al, 1995). The principal difference between PAX3 and PAX3-FKHR is the substitution of the carboxyl terminal domain of PAX3 for the

putative transcriptional activation domain of FKHR. Experiments have been performed to demonstrate that it is this substitution that alters the transcriptional activation potential of the fusion protein. GAL4 fusion constructs were designed to express various carboxyl terminal portions of FKHR and were used in transient transfection assays to demonstrate that the carboxyl terminal 60 amino acid of FKHR are sufficient for activation of a reporter gene (Bennicelli et al, 1995; Sublett et al, 1995). In similar experiments, evidence suggests that there are elements in the N-terminal region of PAX3 which can inhibit transactivation of PAX3 but not PAX3-FKHR (Bennicelli et al, 1996).

Finally, data from studies involving the detection of the (2;13) and (1;13) translocations by fluorescence *in situ* hybridisation suggest that in 20% of cases, multiple copies of the fusion gene are detectable (Barr et al, 1996). This finding indicates that translocation and fusion gene amplification can occur to potentiate tumourigenicity by complementary mechanisms.

Although the genes involved in both alveolar rhabdomyosarcoma translocations have been identified, progress in understanding their specific mechanisms of transformation will depend on finding their individual or shared downstream targets. In this regard, specific target genes for both PAX3 and PAX7 are unknown. However, a precedent may have been set through studies of the molecular genetics of Ewings sarcoma. EWS-FLI is a fusion gene that is formed by an 11;22 chromosomal translocation characteristically found in Ewings sarcoma (Delattre et al, 1992). The fusion gene encodes a chimeric protein that consists of the amino terminus of EWS, a putative RNA-binding protein, fused to the carboxyl terminus of FLI-1, a member of the ETS family of transcription factors. As with the PAX-FKHR fusion protein, EWS-FLI can bind DNA in a site specific manner and the amino terminus of EWS can function as a strong transcriptional activator (Bailly et al, 1994; Lessnick et al, 1995; Mao et al, 1994; May et al, 1993; Ohno et al, 1993; Ohno et al, 1994). Representational difference analysis (RDA), a PCR-based technique, has been used to demonstrate that at least in part, EWS-FLI and FLI-1 modulate distinct target genes (Braun et al, 1995). This suggests that not only DNA-binding domains, but also transactivation

domains of transcription factors are important in determining which target genes are activated or repressed. It remains to be seen whether the PAX3-FKHR and PAX7-FKHR fusion proteins modulate different target genes to PAX3 and PAX7.

1.9 MECHANISMS WHICH CAUSE CHROMOSOME TRANSLOCATIONS

Following the cloning and characterisation of a number of genes associated with several different chromosome translocations it has become possible to consider the molecular characteristics of nucleotide sequences adjacent to chromosome breakpoints, and to begin to understand the mechanisms by which translocation occurs.

Malignancies of the B-lymphocyte lineage often exhibit chromosomal rearrangements that involve the cytogenetic locations of the immunoglobulin loci. For example, some cases of chronic lymphocytic leukaemia (CLL) of the B-cell type carry a t(11;14) translocation involving the immunoglobulin heavy chain locus (IgH) at band 14q32. Probes from the joining segment of the heavy chain locus (J_H) have been used to isolate molecular clones containing the site of joining between chromosomes 11 and 14 (Tsujimoto et al, 1984a). Analysis of these clones demonstrates that the breakpoint on chromosome 14 is found precisely within the J_H segment itself. At the nucleotide level breakpoints are clustered at the 5' end of J-segments and are also clustered on chromosome 11 within a gene designated BCL-1 (B-cell leukaemia/lymphoma-1). Around the breakpoints are nucleotides which, when compared with normal sequences from chromosome 11 or 14, derive from neither. Also, on chromosome 11, sequences are observed that have marked homology to the heptamer-nonamer sequences found near immunoglobulin variable (V), diversity (D) or joining (J) segments (Tsujimoto et al, 1985; Tonegawa, 1983).

These features have direct bearing on the likely mechanism of translocation when taken in the context of the physiological recombination which takes place at the immunoglobulin loci during normal B-lymphocyte development. The variable region of the immunoglobulin heavy chain molecule is encoded by separate gene segments, the V_H , D and J_H segments on chromosome 14. During the somatic development of the B-cell, the segments are brought together so that, in an immunoglobulin producing B-lymphocyte, a contiguous V_H -D- J_H exon is formed.

The enzyme involved in this series of rearrangements is called recombinase (Tonegawa, 1984). Recombinase recognises the heptamer-nonamer sequences found 3' of each V_H , on both sides of each D, and 5' of each J_H segment. The heptamer-nonamer comprises seven conserved nucleotides, then a non-conserved spacer of 12 or 23 nucleotides followed by a nine nucleotide conserved segment. During rejoining of the segments to each other nucleotides may also be deleted, substituted or added. Additional nucleotides are called N regions and are present through the action of terminal deoxynucleotide transferase (Desiderio et al, 1984).

At the molecular level, the t(11;14) breakpoints are thus very reminiscent of a physiologically recombining immunoglobulin gene segment. The breakpoints on chromosome 14 are clustered 5' of the J_H segments, as would be expected if breakage is mediated by the recombinase enzyme. The regions on chromosome 11 have heptamer-nonamer like signal sequences and N regions suggesting that translocations have occurred through the mistaken operation of the immunoglobulin recombinase.

Similar sequence motifs have been found at the breakpoints of cases of follicular lymphoma which carry a t(14;18) (Tsujimoto et al, 1984b; Tsujimoto et al, 1985), acute lymphoblastic leukaemia carrying a t(8;14) and Burkitt's lymphoma also carrying a t(8;14) (Cleary and Sklar, 1985; Bakhshi et al, 1985; Kleinfeld et al, 1986). In approximately 50% of cases involving immunoglobulin or T-cell receptor loci, however, no signal motif sequences have

been found at the break in the other participating chromosome, arguing against a role for V-D-J recombinase in the rearrangement of the corresponding chromosome. In addition, many of these abnormalities have only heptamer signals, with no accompanying nonamer sequence, at the site of the break (Baer et al, 1988; Rabbitts et al, 1988).

A transcriptional accessibility model has been proposed in which rearrangement is thought only to occur between two regions at which there is transcriptional activity thus allowing recombinase to infiltrate the chromatin of the two chromosomes and carry out the interchromosomal event (Yancopoulos and Alt, 1985). In some cases where the recombinase recognises signal sequences on both chromosomes, it has been demonstrated that rearrangement occurs between two regions at which there is transcriptional activity (Baer et al, 1985, 1987; Denny et al, 1986). There are, however, tumour-specific chromosomal abnormalities where transcriptional activity on both chromosomes has not been observed. Instead, stretches of alternating purine-pyrimidine residues are found within a short distance of the breakpoint region. Alternating nucleotide segments can form left-handed or Z-DNA structures which may disrupt chromatin, in a similar way to transcription, to allow recombinase mediated chromosome translocations to be brought about (Boehm et al, 1989; Lu et al, 1991).

1.10 AIMS OF THIS PROJECT

Despite the progress made in understanding the mechanisms which cause chromosome translocations which involve the juxtaposition of proto-oncogenes next to genes of the immune system, little is known about whether similar mechanisms cause other chromosome translocations involving gene fusion. The genes involved in gene fusion translocations have been characterised relatively recently. Because they are in various locations around the genome, the translocation breakpoints have been more difficult to clone. Of the fifty-six different translocations or inversions described in a review (Rabbitts, 1994), twenty-six could be caused by mechanisms involving immunoglobulin or T-cell receptor gene rearrangement. The mechanisms by which the other thirty arise remain unknown.

The first aim of this project was to clone chromosome translocation breakpoints from alveolar rhabdomyosarcoma cell lines which bear the t(2;13) translocation in an attempt to elucidate the mechanism by which they occur.

The activity of the PAX3 proto-oncogene appears to be very important in the aetiology of alveolar rhabdomyosarcoma. It would not be surprising, therefore, if it also plays a role in the development of embryonal rhabdomyosarcoma. The second aim of the project was to screen embryonal rhabdomyosarcoma cell lines and tumour samples for intragenic mutations in the eight PAX3 exons which might contribute to the genesis of embryonal rhabdomyosarcoma.

CHAPTER TWO MATERIALS AND METHODS

MATERIALS;

2.1 Chemical reagents

Chemicals were obtained from Sigma Chemical company except for the following;

Applied Biosystems: Oligonucleotide synthesis reagents

BRL Life technologies: Ultrapure Agarose

Stratagene: Perfectmatch

2.2 Enzymes

Boehringer Mannheim: All Restriction endonucleases (10U/ μ l)

Amersham: T4 DNA Ligase (10U/ μ l)

Klenow Polymerase (5U/ μ l)

Sequenase V2.0 T7 DNA polymerase(13U/ μ l)

Perkin-Elmer Cetus: Amplitaq Polymerase (5U/ μ l)

2.3 Nucleotides and Oligonucleotides

Amersham: [α -³⁵S] dATP aqueous solution

(370MBq/ml, 10mCi/ml)

[α -³²P] dCTP redivue solution

(370MBq/ml, 10mCi/ml)

[γ -³²P] dCTP redivue solution

(370MBq/ml, 10mCi/ml)

2',3'-dideoxynucleoside 5'-triphosphate

(ddATP, ddCTP, ddGTP, ddTTP)

Megaprime random oligomers

Pharmacia: 2'-deoxynucleoside 5'-triphosphate

Poly (dI.dC)(dI.dC)

2.4 DNA size markers

1 kb ladder from BRL Life technologies was used for all DNA size estimation.

2.5 Hybridisation membranes

Amersham Hybond-N+ was used for southern blotting and library screening.

2.6 X-ray film

Kodak Biomax or Fuji XR-100 medical X-ray film

2.7 Bacterial hosts

E.coli strain XL1-blue MR (Stratagene) genotype; $\Delta(mcrA)183$, $\Delta(mcrCB-hsdSMR-mrr)173$, endA1, supE44, thi-1, recA1, gyrA96, relA1, lac.

E.coli strain INV α F' (Invitrogen) genotype; F', endA1, recA1, hsdR17(r_k^- , m_k^+), supE44, thi-1, gyrA96, relA1, $\phi 80lacZ\Delta M15$, $\Delta(lacZYA-argF)U169$, deoR, λ^- .

2.8 Cell lines

RD; Cell line derived from a malignant embryonal rhabdomyosarcoma from a 7 year-old Caucasian female. European collection of animal cell cultures (ECACC) #85111502.

culture medium; Eagles minimum essential medium with Earles balanced salt solution, 2% non-essential amino acids, 2% Vitamins, 10% Foetal bovine serum.

sub-culture procedure; confluent cultures split 1:3 to 1:6 using 0.25% trypsin/EDTA.

RH30; Human alveolar rhabdomyosarcoma cell line. Gift from Dr Aidan McManus, Royal Marsden Hospital, Sutton, Surrey.

culture medium; Dulbecco's minimum essential medium/ Ham's F12 mix, 4% foetal bovine serum, glutamax(Gibco/BRL).

sub-culturing procedure; confluent cultures split 1:3 to 1:6 using 0.25% trypsin/EDTA.

A673; Human sarcoma line derived from a 15 year old female patient. From ECACC #85111504.

culture medium; Dulbecco's minimum essential medium, 10% foetal bovine serum.

sub-culturing procedure; confluent cultures split 1:3 to 1:10 using 0.25% trypsin/EDTA.

K562; Human Caucasian Chronic Myelogenous Leukemia. From ECACC #89121407.

culture medium; RPMI 1640 medium, 10% foetal bovine serum.

sub-culturing procedure; non-adherent cells maintained at between 10^5 - 10^6 cells/ml

A431; Human squamous carcinoma. ECACC #85090402.

culture medium; Eagles minimum essential medium with Earles balanced salt solution, bovine insulin (0.8mg/ml Bovine insulin in 100mM Oxaloacetate/50mM Sodium Pyruvate), 10% foetal bovine serum.

sub-culturing procedure; split 1:3 to 1:6 using trypsin/EDTA.

RPMI 8226; Human myeloma. ECACC #87012702

culture medium; RPMI 1640, 10% foetal bovine serum.

sub-culturing procedure; non-adherent cells maintained at between $3-9 \times 10^5$ cells/ml.

Hela; Human cervix carcinoma. ECACC #85060701

culture medium; Eagles minimum essential medium, 1% non-essential amino acids, 10% foetal bovine serum.

sub-culturing procedure; split 1:3 to 1:10 using trypsin/EDTA.

HL60; Human promyelocytic leukemia. ECACC #85011431

culture medium; RPMI 1640, 10% foetal bovine serum.

sub-culturing procedure; non-adherent cells maintained at $1-5 \times 10^5$ cells/ml.

GM03576; Fibroblast cell line. From the Coriell institute for medical research, USA.

culture medium; MEM Eagles medium-Earle Balanced salt solution, 10% foetal bovine serum.

sub-culturing procedure; Split 1:3 at confluency using trypsin/EDTA.

Karyotype;48,XY,+2,+21

GM10401; Fibroblast cell line. Coriell Institute for Medical Research, USA.

culture medium; MEM Eagles medium-Earles balanced salt solution, 10% Foetal bovine serum.

Karyotype;47,XX,+2

2.9 Cosmid libraries

A. Human genomic cosmid library from the placenta of a newborn caucasian male. Average insert size 35-41kb in vector pWE15 in XL1-blue MR host (Stratagene, #951202).

B. RHF14 genomic cosmid library from the RHF14 somatic cell hybrid cell line. Average insert size 35kb in vector lawrist4 in ED 8767 host (Mitchell and Cowell, 1989).

2.10 Media

LB (LB-agar):	1% w/v bactotryptone
	0.5% w/v bacto yeast extract
	1% w/v NaCl
	(1.5% agar)

SOC: 2% w/v bactotryptone
0.5% bacto yeast extract
10mM NaCl
2.5mM KCl
10mM MgCl₂
10mM MgSO₄
20mM Glucose

2.11 Solutions

TAE: 40mM Tris-acetate
1mM EDTA (pH8.0)

TBE: 89mM Tris-borate
1mM EDTA (pH8.0)

TE: 10mM Tris-HCl (pH 8.0)
0.1mM EDTA (pH8.0)

SSC: 150mM NaCl
15mM Sodium Citrate

SSPE: 150mM NaCl
10mM NaH₂PO₄
1mM EDTA (pH7.7)

Denaturing solution: 0.5M NaOH
1.5M NaCl

Neutralising solution: 0.5M Tris-HCl
0.3M Sodium Citrate
3M NaCl

pH adjusted to 5.5

Denhardts solution (100x): 2% Ficoll

2% Polyvinylpyrrolidone

2% Bovine serum albumin (fraction V)

Hybridisation buffer: 10x Denhardts solution

5x SSPE

0.5% w/v Sodium lauryl sulphate

100µg/ml denatured, sonicated salmon sperm DNA

Oligolabelling buffer: 50mM Tris-HCl (pH7.5)

5mM MgCl₂

0.1mM dATP, dGTP, dTTP

10mM β-mercaptoethanol

0.1mg/ml pd(N)₉

PCR buffer: 67mM Tris-HCl (pH8.8)

16.6mM (NH₄)₂SO₄

6.7mM MgCl₂

10mM β-mercaptoethanol

170µg/ml Bovine serum albumin

Restriction enzyme buffers: Buffers recommended by Boehringer Mannheim

were used in all reactions

Ligation buffer: 6mM Tris-HCl pH7.5

6mM MgCl₂

5mM NaCl

100µg/ml bovine serum albumin

7mM β -mercaptoethanol

0.1mM ATP

2mM dithiothreitol

1mM spermidine

SSCP gel loading buffer: 95% Formamide

10mM NaOH

0.25% Xylene cyanol

0.25% Bromophenol blue

Sequenase buffer: 40mM Tris-HCl pH7.5

20mM $MgCl_2$

50mM NaCl

Sequenase labelling mix(5x): 7.5 μ M dGTP

7.5 μ M dCTP

7.5 μ M dTTP

Sequenase termination mixes: ddG mix 80 μ M dGTP, dATP, dCTP, dTTP

8 μ M ddGTP

50mM NaCl

ddA mix 80 μ M dGTP, dATP, dCTP, dTTP

8 μ M ddATP

50mM NaCl

ddT mix 80 μ M dGTP, dATP, dCTP, dTTP

8 μ M ddTTP

50mM NaCl

ddC mix 80 μ M dGTP, dATP, dCTP, dTTP

	8 μ M ddCTP
	50mM NaCl
Sequenase stop solution:	95% Formamide
	20mM EDTA
	0.05% Bromophenol blue
	0.05% Xylene cyanol FF
Plasmid prep resuspension buffer:	50mM Tris-HCl pH7.5
	100g/ml RNase A
	10mM EDTA
Plasmid prep lysis solution:	200mM NaOH
	1% w/v Sodium lauryl sulphate
Plasmid prep neutralisation buffer:	1.32M Potassium acetate pH4.8
Plasmid wash solution:	50% Ethanol
	100mM NaCl
	10mM Tris-HCl pH7.5
	2.5mM EDTA
Genomic DNA prep reagent A:	10mM Tris-HCl (pH8.0)
	320mM Sucrose
	5mM MgCl ₂
	1% Triton X-100
Genomic DNA prep reagent B:	400mM Tris-HCl (pH8.0)
	60mM EDTA
	150mM NaCl
	1% SDS

METHODS;

2.12 Restriction endonuclease digestion of DNA;

2.12.1 Plasmid and Cosmid DNA

1. 0.5-1µg of plasmid or 2-3µg of cosmid DNA was digested for agarose gel analysis.

2. Incubations were;

xµl DNA

5µl 10x buffer

0.5µl restriction enzyme (10U/µl)

to 50µl with ddH₂O.

3. Digestions were incubated for 4-24 hours at 37⁰C (except TaqI at 65⁰C and ApaI at 30⁰C).

4. Digests were analysed by electrophoresis through 0.5-2% Agarose gels containing g/ml of Ethidium bromide in TAE buffer at 50-150V.

2.12.2 Preparation of genomic DNA

1a. From 5x10⁶ cultured cells, cells were scrapped into tubes and washed in cold PBS.

1b. From a 2mm block of solid soft tissue, tissue stored at -80⁰C was chopped as finely as possible.

2. Tissue was resuspended in 2ml of genomic DNA prep reagent B.

3. 500µl of 5M Sodium Perchlorate was added and the tissue left at room temperature on a rotary mixer for 15 minutes, followed by incubation at 65⁰C for 25 minutes with occasional mixing.

4. 2ml of Chloroform which had been stored at -20°C was added and the lysed, deproteinised tissue was left at room temperature for a further 10 minutes on a rotary mixer and then centrifuged at 800g for 1 minute.
5. 300 μl of Silica suspension (Nucleon) was added and the two phase mixture centrifuged at 1400g for 3 minutes.
6. The upper aqueous phase containing Genomic DNA was removed into another tube and briefly centrifuged to pellet residual Silica.
7. Two volumes of 4°C Ethanol was added and gently mixed to precipitate the DNA.
8. Following centrifugation pelleted DNA was washed in cold 70% Ethanol and air dried for 5 minutes.
9. 500 μl of TE buffer was added and the DNA left at 4°C overnight to allow dissolution.

2.13 Preparation of DNA from agarose gel slices

1. If DNA was to be isolated from an agarose gel care was taken to prevent contamination with DNA from other sections of the gel.
2. The gel slice weighing approximately 100mg was put in an eppendorf tube and incubated with 300 μl of QX1 buffer and 10 μl of Qiaex II silica (Qiagen, 20021) at 50°C for 10 minutes.
3. Following centrifugation the supernatant was removed and the silica-DNA pellet resuspended in 500 μl of QX1 buffer.
4. By serial centrifugation and resuspension the silica-DNA was washed twice in 500 μl of PE buffer (Qiagen).
5. After a final centrifugation the pellet was air dried for ten minutes.

6. The DNA was eluted from the silica by resuspending in 10-20 μ l of ddH₂O, incubating at room temperature for 5 minutes and centrifugation to pellet the silica.

7. DNA was stored at 4⁰C or -20⁰C.

2.14 Polymerase chain reaction (PCR) methodology;

2.14.1 Oligonucleotide synthesis

1. Oligonucleotide primers for PCR were designed using the following guidelines;

a. 18-25bp in length,

b. the 3' end being a G or C nucleotide,

c. preferably no more than a three base stretch of a single nucleotide,

d. no palindromic sequences within the primer,

e. no complimentary sequences between primer pairs, particularly at the 3' end.

2. Primers were synthesised on an Applied Biosystems 392 DNA/RNA oligonucleotide synthesizer, using phosphoramidite chemistry.

3. Primers were deprotected overnight at 55⁰C in 1ml of 30% Ammonium Hydroxide.

4. Primer in solution was dried in a vacuum centrifuge and resuspended in 300 μ l of dH₂O.

5. 4 μ l of primer was diluted in 96 μ l of H₂O and measured for optical density (OD₂₆₀).

Primer concentration (μ g/ml) was calculated as OD₂₆₀ multiplied by 500 (assuming 1 absorbance unit at OD₂₆₀ contains 20 μ g/ml of oligonucleotide).

2.14.2 Basic PCR protocol

1. Reactions were carried out in a final volume of 50 μ l including;

5 μ l 10X PCR buffer (Perkin-Elmer)

1µl dNTPs (0.2mM of each)

5µl primer A (100ng/µl)

5µl primer B (100ng/µl)

0.2µl Taq DNA polymerase (5U/µL)(Perkin-Elmer)

1µl DNA template (200ng genomic DNA or 10ng vector/insert DNA)

32.8µl ddH₂O.

Reactions were overlaid with a drop of mineral oil to prevent evaporation during PCR cycling.

2. PCR cycling was carried out in a Perkin-Elmer 480 Thermal cycler under the following conditions;

1 cycle	94 ⁰ C for 5 minutes
30 cycles	94 ⁰ C for 1 minute
	55-60 ⁰ C for 1 minute
	72 ⁰ C for 1-3 minutes
1 cycle	72 ⁰ C for 10 minutes.

TAQ DNA polymerase was added to the reaction mix during the first cycle of PCR to 'hot start' the reaction.

3. PCR products were analysed by electrophoresis on 0.5-2% agarose gels containing Xg/ml Ethidium bromide in TAE buffer at 50-150V.

2.14.3 Cloning of PCR products

1. PCR products to be cloned were prepared from agarose gel slices or, if a clear single PCR product was present, taken directly from the PCR reaction tube.

2. The cloning strategy relies on the presence of overhanging adenosine nucleotide residues normally found at the ends of PCR products. PCR products are ligated into a plasmid vector which has complimentary overhanging thymidine nucleotide residues. This method prevents vector self ligation and insertion of multiple PCR product copies.

3. Incubation constituents are as follows;

- 1µl Fresh PCR product (50-100ng)
- 1µl 10x ligation buffer
- 1µl vector (25ng) (pCRII Invitrogen)
- 7µl ddH₂O
- 1µl T4 DNA ligase (4U/µl).

Ligation reactions were incubated at 14⁰C overnight.

Cloned products were then used to transform bacterial hosts as follows;

4. A thin walled tube containing 50µl of competent *E.coli* INVαF' bacterial host cells, stored at -80⁰C, were allowed to thaw on ice.

5. 2µl of 0.5M β-Mercaptoethanol and 2µl of the ligation reaction were mixed very gently with the competent cells and incubated on ice for 30 minutes.

6. The cells were then 'heat shocked' at 42⁰C for 30 seconds and then returned to ice for a further 2 minutes.

7. 450µl of SOC medium was added and the tubes left in a shaking 37⁰C incubator.

8. 50-200µl aliquots of cells were then plated onto LB agar plates containing antibiotic and X-gal. Plates were left inverted overnight at 37⁰C.

9. Individual white colonies were picked into 5ml of LB medium containing antibiotic and grown overnight in a shaking 37⁰C incubator.

10. By addition of sterile glycerol to a final concentration of 20% stocks of bacteria containing cloned PCR products were stored at -20°C .

2.14.4 SSCP-PCR

SSCP-PCR incubations were the same as basic PCR with the following alterations;

1. Incubations were carried out in a final volume of $25\mu\text{l}$ so the volume of each constituent of the reaction was adjusted accordingly.
2. The dNTP mix was low in dCTP (0.05mM) to increase incorporation of radiolabelled $[\alpha^{32}\text{P}]$ dCTP.
3. $0.1\mu\text{l}$ of $[\alpha^{32}\text{P}]$ dCTP was included in the reaction.
4. Following cycling $25\mu\text{l}$ of SSCP loading buffer was added and reactions were stored at -20°C until loading onto an SSCP gel.
5. Before gel loading samples were heated at 95°C for ten minutes and immediately cooled on ice.
6. Samples were loaded onto a 6% polyacrylamide gel containing 5% glycerol in TBE buffer and run for 16 to 22 hours at 7-9 Watts at 4°C .
7. The gel was fixed with 10% acetic acid/10% methanol for 15 minutes, dried and autoradiographed overnight.

2.15 Radiolabelling probes

1. 20ng of DNA probe in $31\mu\text{l}$ of ddH_2O was mixed with $5\mu\text{l}$ of random nonamer (Amersham) and denatured in a boiling water bath for 5 minutes.
2. On cooling to room temperature $10\mu\text{l}$ of labelling buffer, $2\mu\text{l}$ of $[\alpha^{32}\text{P}]$ dCTP and finally $2\mu\text{l}$ of Klenow polymerase were added and incubated at 37°C for 10 minutes.

3. The reaction was stopped by addition of 5µl of 0.2M EDTA.
4. Labelled DNA probe was separated from unincorporated nucleotide in a final volume of 400µl by filtration through a Nick column (Pharmacia, 17-0855). The level of incorporation was estimated by comparing the geiger count of the column (unincorporated nucleotides) with the eluate. A greater than 50% incorporation was deemed acceptable.
5. Before use in hybridisation labelled DNA probe was denatured by boiling for 5 minutes, then chilled on ice.

2.16 Cosmid library plating and filter lifts

1. The titre of cosmid libraries was determined using serial dilutions down to 10⁻⁹. 100µl of each dilution was plated on L-agar containing antibiotic and incubated overnight at 37⁰C. The plate with at least 100 colonies was counted. The titre was calculated using the following;

$$\frac{\text{colony number} \times \text{dilution volume}}{\text{volume plated} \times \text{dilution}}$$

2. Five L-agar + antibiotic 25x25cm plates were prepared containing 100,000 colonies on each. Colonies were allowed to grow overnight to approximately 1mm diameter. The plates were then cooled at 4⁰C for 30 minutes.
3. 22x22cm Hybond-N filters were carefully laid over the cold plates, 'keyed' using a hot sterile needle and left for 10 minutes to maximise the number of colonies adhering to the filter.
4. Filters were carefully removed from the plates and carefully laid, colony side uppermost, onto denaturing solution soaked 3MM Whatman paper for 2x5 minutes.

5. Filters were transferred to 3MM Whatman paper presoaked with neutralising solution for a further 2x5 minutes.
6. Finally filters were briefly washed in 2xSSC and then allowed to air dry for 1 hour.
7. Filters were wrapped in SaranWrap and placed DNA-side down on a UV-transilluminator for 2 minutes (312nm wavelength) to fix the DNA to the nylon support.

2.17 Southern blotting

Following electrophoresis, DNA in agarose gels was treated in the following way prior to transfer to a Hybond-N nylon membrane;

1. DNA was nicked by treatment with 0.25M HCl for 15 minutes.
2. The gel was rinsed briefly and then denatured in Denaturing solution for 2x30 minutes followed by neutralisation for 2x30 minutes in Neutralising solution.
3. Blotting was carried out overnight in 20xSSC.
4. The filter was briefly rinsed in 2xSSC, air dried and the DNA fixed to the filter by crosslinking using a 310nm UV-light source for 2-3 minutes

2.18 Screening of DNA-bound filters by hybridisation

1. Filters were rolled together with mesh into Hybaid bottles and presoaked with 30ml of 20mM Sodium Phosphate.
2. Sodium phosphate solution was removed and replaced with 13ml of preheated prehybridisation solution. Bottles were placed in a 65⁰C rotisserie to prehybridise for 3 hours.
3. Prehybridisation solution was replaced with hybridisation solution containing labelled denatured probe. Bottles were left overnight.

4. Following hybridisation filters were washed twice for 10 minutes in 2 x SSC, 0.1% (w/v) SDS at room temperature.
5. Filters were further washed at varying stringency. Stringency was varied using salt concentration (SSC, 2x to 0.1x) and/or temperature (65⁰C to room temperature).
6. Washed filters were wrapped in SaranWrap and autoradiographed using markers (Stratagene) for alignment of filters and plates to developed film.
7. The area on the plate corresponding to a positive signal on the film was scraped into LB medium containing antibiotic and retitred.
8. Secondary screening of 100-300 colonies on 150mm diameter LB-agar plates was carried out as above.

2.19 Isolation of plasmid DNA

1. Cells from 3ml overnight cultures grown in LB+antibiotic were pelleted by centrifugation and resuspended in 180µl of plasmid prep resuspension solution.
2. The suspension was mixed gently with 180µl of lysis solution until clear.
3. The lysate was mixed gently with 180µl of neutralisation buffer to precipitate cellular debris.
4. Following centrifugation the supernatant or 'cleared lysate' was removed for further purification.
5. Cleared lysate was mixed with 1ml of plasmid miniprep purification resin (Promega). The resin bearing the plasmid DNA was collected on a support column (Promega) and washed with 2ml of plasmid wash solution then dried by centrifugation.
6. Plasmid DNA was eluted from the resin with 50µl of 65-70⁰C TE by centrifugation.
7. Plasmids were stored at -20 or 4⁰C.

2.20 Isolation of cosmid DNA

2.20.1 Small-scale;

1. The first four steps of the plasmid isolation method described above were followed for small-scale cosmid DNA isolation.
2. Cleared lysates were twice extracted with an equal volume of phenol/chloroform to remove protein, followed by extraction with an equal volume of chloroform to remove traces of phenol.
3. One tenth volume of 3M Sodium acetate (pH5.2) and two volumes of ethanol were added and left at -20°C for 30 minutes to precipitate cosmid DNA.
4. DNA pellet was washed twice in cold 70% ethanol, air dried for 5 minutes and resuspended in 50 μl of TE buffer.
5. Cosmids were stored at 4°C or -20°C .

2.20.2 Large-scale;

1. Cells from 500ml overnight cultures grown in LB+antibiotic were pelleted by centrifugation and resuspended in 15ml of plasmid prep resuspension solution.
2. Resuspended cells were mixed gently with 15ml of plasmid prep lysis solution until clear
3. The lysate was mixed gently with 15ml of plasmid prep neutralisation solution to precipitate cellular debris.
4. Following centrifugation the cleared lysate supernatant was removed for further purification.
5. Cosmid DNA was precipitated with 0.6 volumes of isopropanol.
6. Following centrifugation pellets were resuspended in 2ml of TE buffer.

7. Exactly 2g of Caesium Chloride was mixed with 2ml of Cosmid DNA solution.
8. 200 μ l of Ethidium Bromide (10mg/ml) was added to exactly 2.5ml of the Caesium Chloride/DNA mix.
9. The mix was loaded into 3ml polycarbonate tubes (Beckman) and centrifuged at 80000 rpm for 18 hours at 20⁰C in a TL100.3 fixed angle rotor (Beckman).
10. The cosmid DNA band was removed from the tube using a needle and syringe in a volume of approximately 100 μ l. The volume was increased to 300 μ l with TE buffer.
11. An equal volume of butanol was added, vortexed and centrifuged to extract Ethidium Bromide. The extraction was repeated (at least four times) until red colour could not be seen in either the aqueous or the butanol phase. The volume was increased to 1ml with TE buffer.
12. Caesium Chloride was removed with three rounds of centrifugation through Centricon-30 filters. Volume was readjusted to 1ml with TE buffer before each round.
13. Cosmid DNA concentration was determined at OD₂₆₀ assuming one absorbance unit at OD₂₆₀ contains 50 μ g/ml of double stranded DNA.
14. Cosmid DNA was stored at 4⁰C or -20⁰C.

2.21 Sequencing

All sequencing was carried out on plasmid DNA templates containing inserts of interest, using suitable vector primers flanking the insert on both sides eg. T7 or SP6 in vector pCRII (Invitrogen).

1. 3 μ g of plasmid DNA in 50 μ l of dH₂O was denatured by incubation at 37⁰C for 30 minutes in 0.2M NaOH/ 0.2mM EDTA.

2. The mixture was neutralised by addition of 0.1 volumes of 3M Sodium acetate (pH5.2) and the DNA precipitated at -70°C for 15 minutes with 3 volumes of Ethanol.
3. Following centrifugation, the pelleted DNA was washed with cold 70% Ethanol, air dried for 5 minutes and resuspended in $7\mu\text{l}$ of dH_2O .
4. $2\mu\text{l}$ of sequenase reaction buffer and $1\mu\text{l}$ of primer (XX) were added, the mixture was then heated to 65°C for 2 minutes and allowed to cool slowly (30 minutes) to room temperature to anneal primer to template.
5. The following were then added to the mixture in order;

Dithiothreitol (0.1M)	$1\mu\text{l}$
Diluted labelling mix (1x)	$2\mu\text{l}$
$[\text{}^{35}\text{S}]$ dATP	$0.5\mu\text{l}$
Diluted sequenase (1.63U/ μl)	$2\mu\text{l}$
6. After a 2 minute incubation at room temperature $3.5\mu\text{l}$ of this mixture was transferred to four tubes each containing $2.5\mu\text{l}$ of a termination mix.
7. After a further 2 minute incubation $4\mu\text{l}$ of stop solution was added to all four tubes.
8. The tubes were heated to 95°C for 2 minutes and $4\mu\text{l}$ of each mix was then loaded onto a 6% polyacrylamide gel in four adjacent lanes.
9. Electrophoresis was carried out in TBE buffer at 800Volts for 3-6 hours.
10. Gels were fixed in 10% acetic acid/ 10% methanol for 15 minutes, dried and autoradiographed overnight.

2.22 Whole cell protein extract preparation

(Adapted from a method by Andrews and Faller 1988, Nucleic Acids Research, Vol. 19, p2499.)

1. Whole cell extracts, prepared from between 5×10^5 and 10^7 cells, were washed in 10ml of ice cold PBS and either scrapped or pelleted into eppendorf tubes.
2. Cell pellets were resuspended in 400 μ l of ice cold hypotonic buffer (10mM HEPES-KOH pH7.9, 1.5mM MgCl₂, 10mM KCl, 0.5mM DTT, 0.2mM PMSF) and incubated on ice for 10 minutes.
3. Cells were centrifuged for 10 seconds at 13000rpm then resuspended in 100 μ l ice cold hypertonic buffer (20mM HEPES-KOH pH7.9, 25% glycerol, 420mM NaCl, 1.5mM MgCl₂, 0.2mM EDTA, 0.5mM DTT, 0.2mM PMSF) and incubated for a further 20 minutes.
4. Cellular debris was removed by centrifugation for 2 minutes at 4⁰C, the supernatant was aliquoted and stored at -70⁰C.
5. A 10 μ l sample was taken to measure protein concentration.

2.23 Protein concentration assay (Sigma P5656)

1. 10 μ l of extract was diluted into 400 μ l of dH₂O, then mixed with 400 μ l of Lowry reagent and left at room temperature for 20 minutes.
2. Protein standard solutions (Bovine serum albumin) ranging in concentration from 25 to 300 μ g/ml in a volume of 400 μ l were also mixed with Lowry reagent and left at room temperature for 20 minutes.
3. 200 μ l of Folin & Ciocalteu's phenol reagent was added to each mix and left to allow colour to develop for 30 minutes.

- Absorbance at 750nm wavelength was measured.
- Extract protein concentrations were calculated by extrapolation from a calibration curve which was based on the absorbance of the protein standard solutions.

2.24 End-labelling of oligonucleotides

- 250ng of oligonucleotide was end-labelled with T4-polynucleotide kinase and [γ - 32 P]ATP in buffer (50mM Tris-HCl pH7.6, 10mM MgCl₂, 5mM Dithiothreitol, 0.1mM Spermidine, 0.1mM EDTA), as follows;
x μ l Oligonucleotide
1 μ l [γ - 32 P]ATP
5 μ l Buffer
1 μ l T4 polynucleotide kinase (3U/ μ l)
to 50 μ l with dH₂O.
- Reactions were incubated at 37⁰C for 30 minutes and terminated at 65⁰C for 5 minutes.
- 5 μ l of Sodium acetate and 100 μ l of Ethanol were added, mixed and left at -20⁰C for 30 minutes
- Precipitated oligonucleotides were pelleted by centrifugation at 13000rpm for 5 minutes, washed twice with cold 70% ethanol and air-dried.
- Pellets were resuspended in 250 μ l of TE.

2.25 Electrophoretic mobility shift assay (EMSA)

1. 7 μ g of whole cell extract was mixed with 4 μ g of poly(dI.dC)(dI.dC) in binding buffer (10mM Tris-HCl (pH7.5), 50mM NaCl, 1mM EDTA, 1mM DTT, 5% glycerol) and preincubated at room temperature for 15 minutes to adsorb non specific DNA-binding proteins which are present in protein extracts.

2. 10ng of labelled oligonucleotide (1000cpm/ng) was then added to make the final

reaction mixture volume 20 μ l, as follows; 7 μ g Protein extract

2 μ l 10X binding buffer

4 μ l Poly(dI.dC)(dI.dC)

1 μ l 10ng/ μ l Labelled oligonucleotide

to 20 μ l with dH₂O.

3. Reactions were incubated at room temperature for 20 minutes.

4. 4 μ l of loading buffer was added and DNA-protein complexes were separated on 5% non-denaturing polyacrylamide gels in TBE at 100 volts for 2 hours.

5. Gels were fixed in 10% acetic acid, 10% methanol prior to drying and autoradiography.

CHAPTER 3. MUTATION ANALYSIS IN THE PAX3 GENE IN EMBRYONAL AND ALVEOLAR RHABDOMYOSARCOMA

The introductory chapter above, describes the evidence which implicates the PAX3 gene in the genesis of alveolar rhabdomyosarcoma. It is the DNA binding elements of this gene, rather than those of the FKHR gene, which remain intact following the creation of the fusion gene product by chromosome translocation. Therefore, it would not be surprising if an aberration in PAX3 gene function also played a role in the genesis of embryonal rhabdomyosarcoma. The most common cytogenetic feature of embryonal rhabdomyosarcoma, which supports this hypothesis, is the presence of trisomic chromosome 2, the chromosome on which the PAX3 gene resides (Beigel et al, 1995).

Point mutation in an oncogene as a possible cause of transformation has been demonstrated previously in studies of the receptor tyrosine kinase gene, RET. RET gene mutations have been found in both familial and sporadic forms of medullary thyroid carcinoma (MTC), papillary thyroid carcinoma (PTC) and multiple endocrine neoplasia types 2A and 2B (MEN 2A or MEN 2B) (Mulligan et al, 1993; Donis-keller et al, 1993; Hofstra et al, 1994). The RET gene, like the PAX3 gene, was first defined as a dominant oncogene by DNA transformation assay (Grieco et al, 1990). Oncogenic activation, in patients with PTC, was shown to be due to chromosomal rearrangement in which the extracellular domain of RET is replaced by protein fragments from various other genes. Patients with MTC or MEN 2A carry point mutations which involve four conserved cysteine residues in a cysteine-rich domain of the RET gene. These mutations are thought to interfere with ligand binding (Mulligan et al, 1993; Donis-

Keller et al,1993). In patients with sporadic forms of MTC and MEN 2B, point mutation causes a methionine-to-threonine change in the tyrosine kinase domain of RET which may cause altered target specificity (Hofstra et al, 1994).

To test the hypothesis that alterations in the PAX3 gene locus are involved in embryonal as well as alveolar rhabdomyosarcoma, tumour samples and cell lines were analysed for intragenic mutations.

3.1 Mutation detection methods.

Several methods are available for analysing gene sequences for possible mutations. The most thorough but also most inefficient method would be to clone and sequence the whole gene in each tumour sample. Other more rapid methods rely on the ability to differentiate between wild type and mutant alleles to identify samples bearing mutations, followed by cloning and sequencing of the mutant allele (see below). Because of their structure in eukaryotes, genes have to be analysed exon by exon.

The chemical cleavage of mismatch (CCM) method results in the base-specific chemical modification and cleavage of mismatched basepairs following the cross-hybridisation of patient to wildtype DNA (Cotton et al, 1988). The chemicals, hydroxamine and osmium tetroxide are used to modify specifically mismatched cytosine and thymine nucleotides respectively. The chemically altered mismatch sites are then cleaved with piperidine to produce two DNA fragments. The sizes of the fragments will depend on the location of the mismatch.

In the enzyme cleavage of mismatch (ECM) method, mutations are detected by a method which is similar to CCM. Rather than requiring multiple incubation steps which involve toxic chemicals, mismatch sites in patient-wildtype heteroduplex DNA are recognised and cleaved by the bacteriophage T4 endonuclease VII enzyme (Youil et al, 1993). However, this enzyme has residual double-stranded DNA activity which results in high background cleavage, so the ECM method is not as efficient as other methods.

The heteroduplex analysis technique is based on differences in the migration rates of homoduplex and heteroduplex DNA molecules through polyacrylamide gels (White et al, 1992). Under appropriate conditions, an extra band will appear in samples containing both wildtype and mutant alleles. Heteroduplexes can be created by denaturing double-stranded DNA at 95⁰C and reforming the double-strands by slow cooling prior to loading on a gel.

The denaturing gradient gel electrophoresis (DGGE) method exploits the differences in DNA melting temperatures which are caused by altered base composition (Sheffield et al, 1989). DNA samples are PCR amplified using a primer with a GC-rich 5' end. Samples are electrophoresed through a gradient of denaturant. Depending on the base composition, samples will melt but the two DNA strands will be held together by the 5' 'GC-clamp'. The denatured and clamped sample molecules have much lower mobility. Mutant and wildtype alleles will melt at different positions in the gel and thus can be resolved.

Finally, the single strand conformation polymorphism (SSCP) method exploits the fact that single-stranded DNA fragments will adopt a particular conformation, which is dependent on their base composition, when run under appropriate conditions on a non-

denaturing polyacrylamide gel (Orita et al, 1989). The SSCP method was developed to look for subtle mutations such as point mutations or small deletions/insertions in sequences up to 400bp in length. The technique can be used to detect such changes whether they are present in both alleles of a gene or within one allele. The method involves radiolabelling PCR products, denaturing the double-stranded helix and then slowly separating the two single stranded DNA molecules by electrophoresis on a non-denaturing polyacrylamide gel. Slow electrophoresis allows the single-stranded DNA molecules to form secondary structures which will affect gel mobility. Even a single point mutation within an allele can alter this secondary structure conformation and thus alter mobility.

For this study, genomic DNA from both embryonal and alveolar rhabdomyosarcoma tumours were analysed for intragenic mutation using the Single-Strand Conformation Polymorphism (SSCP) method (Orita et al, 1989). The same method has been used previously to look for mutations in the PAX3 gene in patients with Waardenburg syndrome in which inactivating mutations in the PAX3 gene have been detected (Tassabehji et al, 1992). It is feasible that activating/enhancing mutations alter the DNA binding kinetics or the messenger RNA/protein stability of PAX3 in Embryonal Rhabdomyosarcoma, thus contributing to neoplasia. However, if disease associated mutations are not present within the PAX3 gene, this finding would not necessarily exclude its involvement. Promoter element or position effect mutations might be involved, and such mutations are more difficult to detect.

The PAX3 gene is comprised of eight exons. Exons 2-6, which code for the DNA-binding elements within the gene, have been studied previously using the SSCP method in patients with Waardenburg syndrome (Tassabehji et al, 1992). Primer sequences and expected PCR product sizes are indicated in table 3.1 below;

Table 3.1 Primers used in SSCP analysis of the PAX3 gene.

EXON	PRIMER	PRIMER PAIRS	PRODUCT SIZE
1	E1F	5'-GATATAATTTCCGAGCGAAGTGC	160bp
	E1R	5'-GTCCCAGGCCCTGGGATCCAGG	
2	E2F	5'-GAAGACTGCGAAATTACGTGCTGC	368bp
	E2R	5'-GACCACAGTCTGGGAGCCAGGAGG	
3	E3F	5'-AGGGTACCGGGTACCAACGCCTGC	185bp
	E3R	5'-GTAATAGCGACTGACTGTCGCGCC	
4	E4F	5'-AGCCCTGCTTGTCTCAACCATGTG	243bp
	E4R	5'-TGCCCTCCAAGTCACCCAGCAAGT	
5	E5F	5'-ACTGTAATGGTGTCTTGCAGTCGG	350bp
	E5R	5'-CTGTCTGGACTGAAGTAGGACACG	
6	E6F	5'-AACTATTATTTTCATCAGTGAAATC	291bp
	E6R	5'-ATCACTTGTATAAAATATCCACC	
7	E7F	5'-GCTGCAATCTGGATCTCTCTG	361bp
	E7R	5'-GAGTTGAGTTTATCTCCCTTC	
8	E8F	5'-GTAAGCCTTGGACTTTTTAG	363bp
	E8R	5'-GACAAGTTTGACACACAACCTTG	

Unlabelled PCR amplification products for each exon were analysed by electrophoresis through 2% agarose gels to ensure that clean single products of the expected size could be obtained from each DNA sample.

Despite using a range of primer annealing temperatures (55-65⁰C) and a range of concentrations of Magnesium chloride (1-4mM) during PCR amplification it was not possible to amplify a single clean product from exon 1 of the PAX3 gene. Similar difficulties had been encountered previously in amplifying this product from genomic DNA samples of patients with Waardenburg syndrome (Tassebehji et al, 1994). Therefore, it was not possible to screen exon 1 for mutations by SSCP analysis.

The remaining seven exons could be amplified cleanly from genomic DNA samples, so they were analysed by SSCP. The seven exons were each analysed in of total of twenty genomic DNA samples. The samples consisted of nine embryonal rhabdomyosarcoma samples (one cell line and eight primary tumour samples), seven alveolar rhabdomyosarcoma samples (three cell lines and four primary tumour samples), and four control samples. The karyotypes of the eight primary embryonal rhabdomyosarcoma samples and four primary alveolar rhabdomyosarcoma samples are not known. The four control samples comprised of a human placental DNA, DNA from the cell line A673 which was derived from a neuroepithelioma tumour, and two fibroblast derived cell lines (GM10401 and GM03576, see materials and methods) which had been cytogenetically characterised and shown to be trisomic for chromosome 2. If there was an indication of mutation in any of the test samples in any of the exons, a similar analysis was performed on a further twenty human control genomic DNA samples to determine whether these mutations were specifically associated with rhabdomyosarcoma.

Table 3.2 summarises the SSCP analysis;

DNA source	Exon2	Exon3	Exon4	Exon5	Exon6	Exon7	Exon8
Embryonal RMS (9)	1/9	0/9	0/9	0/9	0/9	0/9	1/9
Alveolar RMS (7)	2/7	0/7	0/7	0/7	0/7	0/7	0/7
Control (24)	2/24	0/24	0/24	0/24	0/24	0/24	2/24

The SSCP banding pattern in three out of sixteen rhabdomyosarcoma samples indicated that there could be a polymorphism in exon2 of the PAX3 gene. However, three out of twenty-four control samples also had a polymorphism which migrated at the same rate as the rhabdomyosarcoma samples (see figure 3.1). Similarly, one of sixteen rhabdomyosarcoma samples had a polymorphism in exon8 of the PAX3 gene, but two out of twenty four control samples also had the same polymorphism (see figure 3.2). Since the observed polymorphisms are present in both test and control samples it is unlikely that they contribute towards the malignant phenotype.

In an attempt to confirm that the abnormalities detected were merely polymorphic variants, PCR amplification products from exons two and eight were cloned and sequenced. The polymorphism in exon2 is a point mutation at base 44 which does not alter the subsequent amino acid structure of the PAX3 protein (see figure 3.3). The polymorphism in exon8 is a point mutation in the 15bp 'Pyrimidine-rich region' of the 3' splice site (see figure 3.4). The Pyrimidine-rich region is commonly, but not universally, found at the 3' splice site so it is unlikely that this polymorphism will affect splicing of the PAX3 gene (Padgett et al, 1986). The effect of the splice site

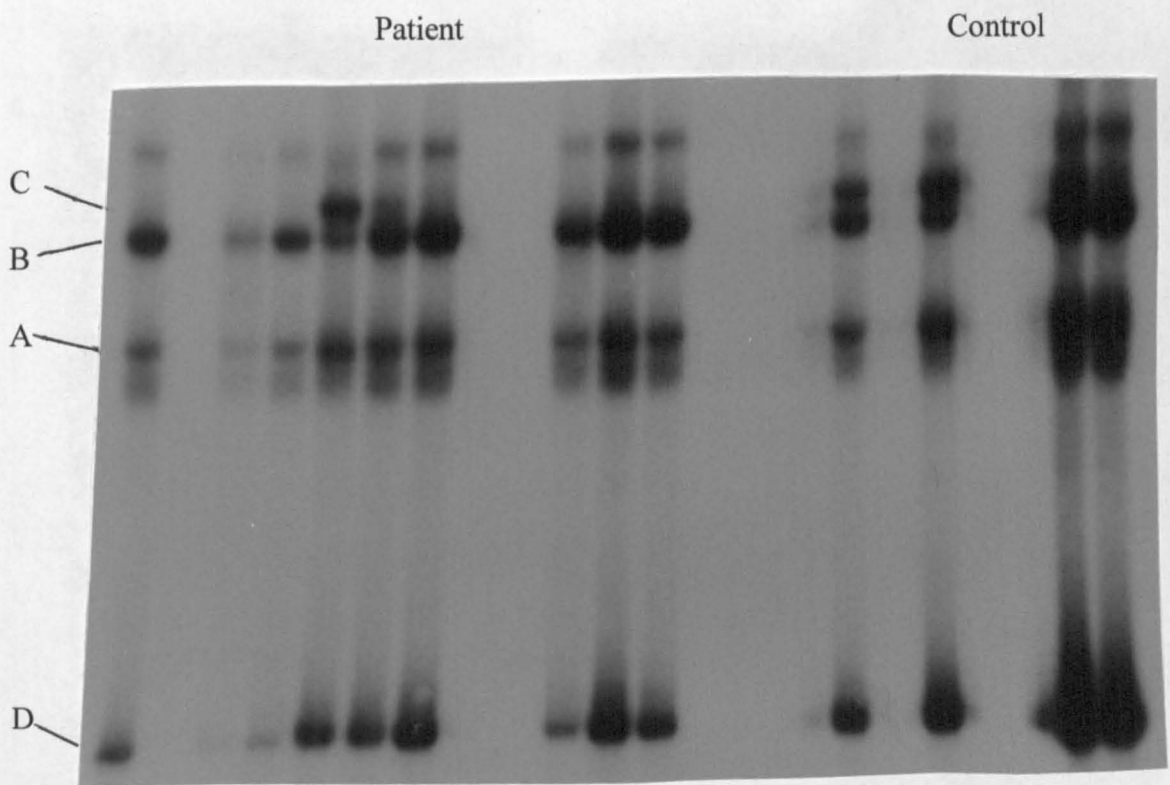


Figure 3.1 Analysis of the PAX3 exon2 polymorphism by single strand conformation polymorphism (SSCP). Labelled PCR amplification products using primers E2F and E2R (table 3.1) from rhabdomyosarcoma cell lines and patient samples were denatured and the two single stranded DNA molecules (bands A and B) separated by electrophoresis on a non-denaturing polyacrylamide gel. Bands from samples bearing a sequence polymorphism are slightly shifted (band C). Similar shifted bands from normal DNA samples are also present indicating that the polymorphisms are not disease associated. Band D is non-denatured sample. Gaps between sample lanes are due to poor well formation.

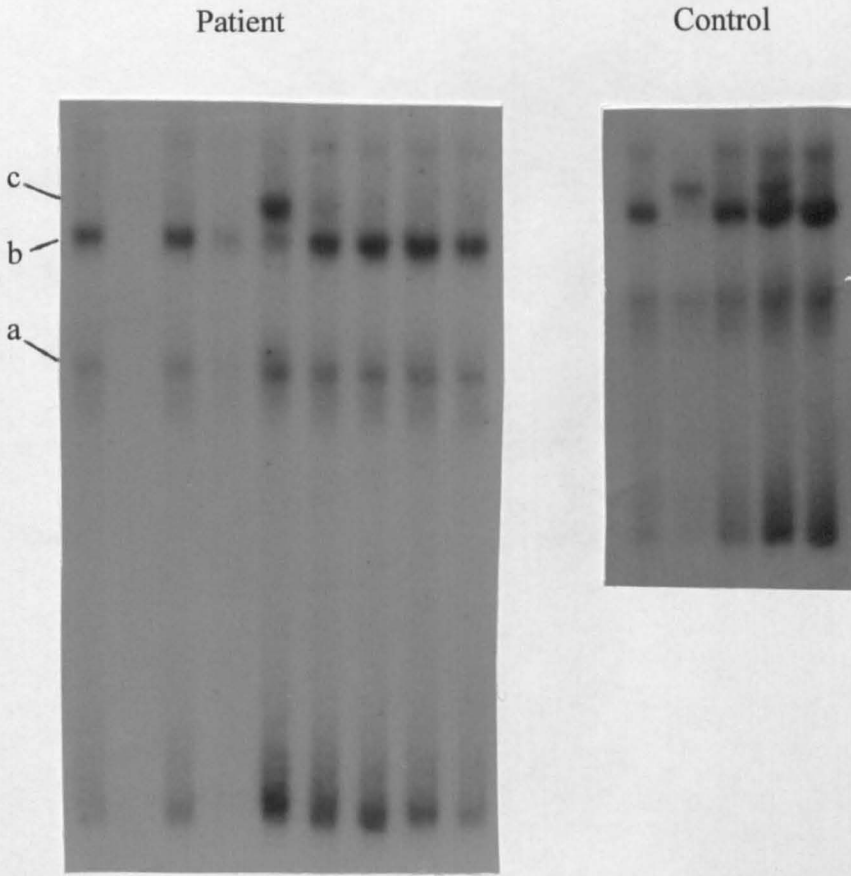


Figure 3.2 Analysis of polymorphisms in PAX3 exon8 observed by SSCP. Labelled PCR amplification products using primers E8F and E8R (table 6.1) from rhabdomyosarcoma cell lines and patient samples were denatured and the two single stranded DNA molecules (bands A and B) separated by electrophoresis on a non-denaturing polyacrylamide gel. A band from a patient sample bearing a sequence polymorphism is slightly shifted (band C). A similar shifted band is also present in control samples indicating that there is no disease association.

```

control  GAAGACTGCGAAATTACGTGCTGCTGTTCTTTGCTTTTTATTTTCCTCCAGTGACTTTTCCCTTGCTTCTCTTTT
rms      GAAGACTGCGAAATTACGTGCTGCTGTTCTTTGCTTTTTATTTTCCTCCAGTGACTTTTCCCTTGCTTCTCTTTT

          INTRON 1 | EXON 2
                  | S T P L G Q G R V N Q L G G V F I N G R
control  CACCTTCCCACAGTGTCCACTCCCCTCGGCCAGGGCCGCGTCAACCAGCTCGGCGGGGTTTTTATCAACGGCAGG
          *
rms      CACCTTCCCACAGTGTCCACTCCCCTCGGCCAGGGCCGCGTCAACCAGCTCGGCGGGGTTTTTATCAACGGCAGG

```

Figure 3.3 The PAX3 exon2 sequence. The sequence of four control samples of the 3'-end of intron1 and the 5'-end of exon2 of the PAX3 gene is identical to published sequences. Sequencing of four normal and rms polymorphic sample clones indicates that a single base substitution has occurred at the 44th base in exon2 (*). The substituted base is the third in a codon and, due to the degeneracy of the code, does not alter the resulting encoded amino acid structure (glycine). The E2F primer is underlined.

polymorphism on the splicing of the PAX3 gene could be analysed by reverse transcription-PCR of PAX3 mRNA. The tumour samples used in this study were used to allow the isolation of PAX3 mRNA.

Although direct mutations within the promoter region of the PAX3 gene could not be found in embryonal rhabdomyosarcoma in this study, this finding does not preclude the involvement of the PAX3 gene in embryonal forms of the disease. Since the PAX3 gene is only normally expressed during fetal development, any mutation which allows deregulation of expression in the adult could contribute to embryonal rhabdomyosarcoma. Promoter elements or protein-coding mutations which could cause such deregulation of PAX3 gene expression are discussed in the final section of this thesis.

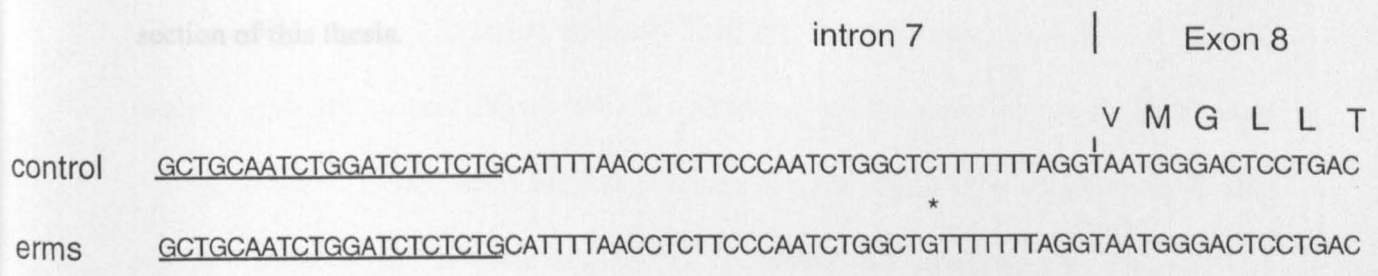


Figure 3.4 The PAX3 exon8 sequence. The sequence of four control samples of the 3'-end of intron7 and the 5'-end of exon8 of the PAX3 gene is identical to published sequences. Sequencing of the three polymorphic samples indicate that a point mutation has occurred in the pyrimidine-rich region of the 3' splice site of intron7.

polymorphism on the splicing of the PAX3 gene could be analysed by reverse transcriptase-PCR (RT-PCR) of PAX3 mRNA. However, the tumour samples used in this study were too small to allow the isolation of RNA.

Although disease associated mutations within the coding region of the PAX3 gene could not be found in embryonal or alveolar rhabdomyosarcoma in this study, this finding does not preclude the involvement of the PAX3 gene in the embryonal form of the disease. Since the PAX3 gene is only normally expressed during foetal development, any mutation which allows deregulation of expression in the adult could contribute to embryonal rhabdomyosarcoma. Promotor element or position effect mutations which could cause such deregulation of PAX3 gene expression are discussed in the final section of this thesis.

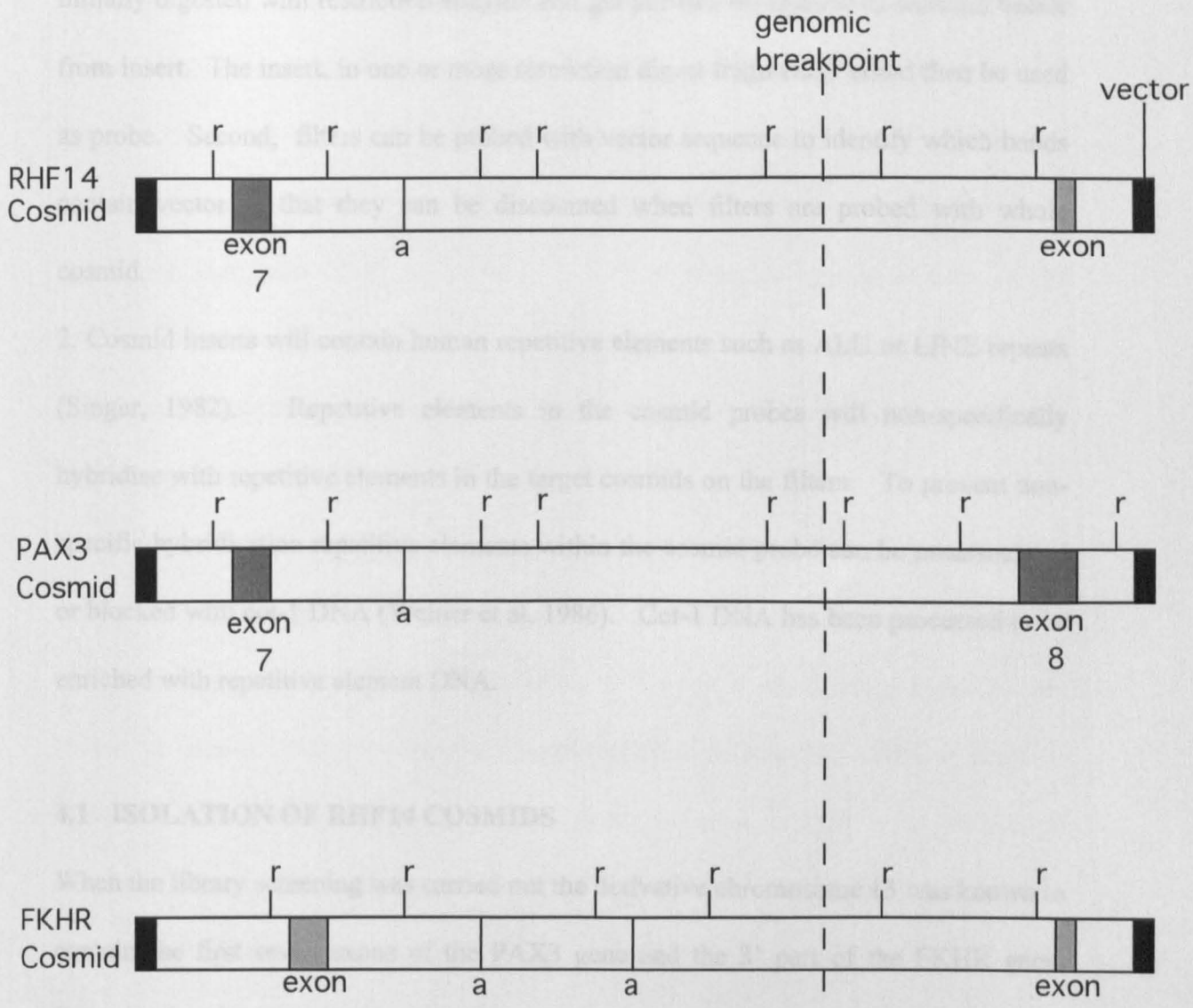
CHAPTER 4 CLONING OF CHROMOSOME TRANSLOCATION BREAKPOINTS USING COSMIDS

The experimental data presented in this chapter was gathered before the sequencing of the PAX3 gene intron in which chromosome translocation breakpoints are grouped and before the FKHR gene structure had been investigated (Macina et al, 1994; Davis et al, 1995). A method was developed to isolate cosmid derived clones which span the translocation breakpoint in the cell line RH28. A cosmid library had previously been constructed from genomic DNA of the somatic cell hybrid RHF14. The RHF14 hybrid contains the derivative chromosome 13 of the RH28 cell line but not the derivative chromosome 2 or the two corresponding normal chromosomes (Mitchell et al, 1991). The method involved isolating cosmids from the RHF14 library and from a normal human genomic cosmid library which hybridised with probes based on PAX3 and FKHR exons. It was assumed that cosmids isolated using these probes would also contain portions of the relevant intron in which the chromosome translocation breakpoint lay. By probing RHF14 cosmids, digested with various restriction enzymes, with appropriate cosmids from both normal chromosomes it ought to be possible to isolate a single restriction fragment which hybridises with both probes. A band which has sequences derived from both normal chromosomes presumably contained the translocation breakpoint (see figure 4.1). Two technical problems had to be overcome;

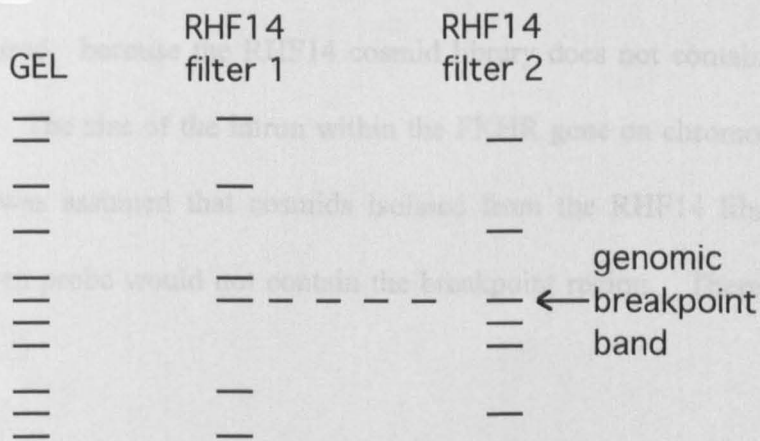
1. Using a whole cosmid as a probe to search for common genomic sequences in another cosmid would cause hybridisation between common vector sequences. There are two ways of overcoming this problem. First, cosmids to be used as probe are

Figure 4.1 Cosmid/ cosmid hybridisation method to isolate breakpoint spanning sequences.

Cosmids were isolated from the RHF14 library and from a normal human genomic cosmid library which hybridised with probes based on PAX3 and FKHR exons. By probing RHF14 cosmids, digested with various restriction enzymes, with cosmids from both normal chromosomes it should have been possible to isolate a single restriction fragment which hybridises with both probes. A band which has sequences derived from both normal chromosomes was presumed to contained the translocation breakpoint.



r - restriction site
 a - repetitive sequence



initially digested with restriction enzyme and gel purified on agarose to separate vector from insert. The insert, in one or more restriction digest fragments, could then be used as probe. Second, filters can be probed with vector sequence to identify which bands contain vector so that they can be discounted when filters are probed with whole cosmid.

2. Cosmid inserts will contain human repetitive elements such as ALU or LINE repeats (Singer, 1982). Repetitive elements in the cosmid probes will non-specifically hybridise with repetitive elements in the target cosmids on the filters. To prevent non-specific hybridisation repetitive elements within the cosmid probe can be preassociated or blocked with cot-1 DNA (Weiner et al, 1986). Cot-1 DNA has been processed to be enriched with repetitive element DNA.

4.1 ISOLATION OF RHF14 COSMIDS

When the library screening was carried out the derivative chromosome 13 was known to contain the first seven exons of the PAX3 gene and the 3' part of the FKHR gene. Because the chromosome translocation breakpoint region in the RH28 cell line had previously been mapped to a region approximately 2 kilobases from exon eight of the PAX3 gene on chromosome 2 and approximately 14 kilobases from exon seven (Barr et al, 1993). Only probes homologous to PAX3 exon seven and the 3' part of the FKHR gene could be used, because the RHF14 cosmid library does not contain PAX3 exon eight sequence. The size of the intron within the FKHR gene on chromosome 13 was unknown. It was assumed that cosmids isolated from the RHF14 library with the PAX3 exon seven probe would not contain the breakpoint region. Therefore cosmids

were also isolated using a 3' FKHR exon probe assuming that the breakpoint region on chromosome 13 would be less than 14 kilobases from 3' FKHR exonic sequences.

Primary screening with the FKHR probe produced nine possible positive colonies. Secondary screening confirmed that four of the nine primary positive colonies hybridised specifically with the probe. Individual colonies containing each of the positive cosmids were picked and grown overnight in L-broth containing antibiotic. Small scale preparations of cosmids were digested and the resulting fragments separated by electrophoresis on 0.8% agarose gels (see figure 4.2).

Since the cosmids were isolated from a cosmid library from a somatic cell hybrid (human DNA in a mouse background) it was feasible that they originated from the murine homologue of the FKHR gene locus. To test if the cosmids isolated were human or murine, PCR products were amplified from murine 3T3 cell line genomic DNA and from RH28 genomic DNA. The primers chosen for PCR were designed to allow amplification of sequences from the 3' FKHR exon which had been used as probe in the cosmid library screening. The PCR products were cloned and sequenced. The sequence of the RH28 derived PCR product clones exactly matched the published human FKHR gene sequence (Galili et al, 1993). Ten conservative base changes were apparent in the sequence of the 3T3 cell line derived PCR product clones (see figure 4.3). Thus the absence of base changes in the cosmid clones strongly suggests that they are of human origin.

Published

sequence

RH28

3T3

```
5'-AATTCGTCAGAAATCTLRCCCTACACAGCAAGTTCATTGCTGTCAG  
5'-AATTCGTCAGAAATCTLRCCCTACACAGCAAGTTCATTGCTGTCAG  
5'-AATTCGTCAGAAATCTLRCCCTACACAGCAAGTTCATTGCTGTCAG
```

amino acids

```
I R H N L S L H S K P L R V Q
```

```
AATGAA GGA ACT GGA AAA TTT TTT TTT TTT TTT TTT TTT TTT  
AATGAA GGA ACT GGA AAA TTT TTT TTT TTT TTT TTT TTT TTT  
AATGAA GGA ACT GGA AAA TTT TTT TTT TTT TTT TTT TTT TTT
```

```
N E G T A M L N P E G G K
```

```
AGC GGG AAA TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT  
AGC GGG AAA TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT  
AGC GGG AAA TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT
```

```
S G K S P R A R M D N N S K F
```

GCT AAG AAC CGA AGC-3'

GCT AAG AAC CGA AGC-3'

GCT AAG AAC CGA AGC-3'

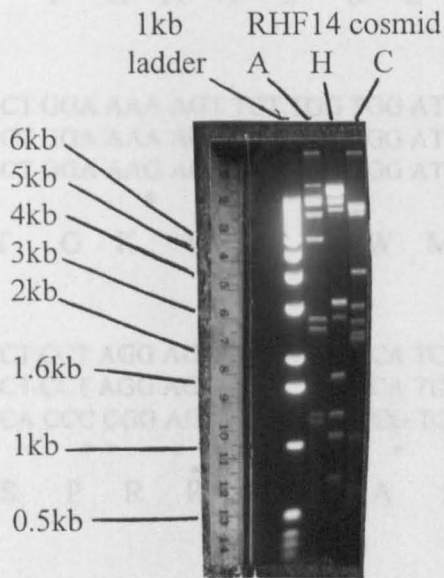


Figure 4.2 RHF14 cosmid restriction digest patterns. Restriction digestion of RHF14 cosmid with ApaI (A), HindII (H), ClaI (C) indicating a size of 30-35kb.

The PCR products (primers underlined), of the 3' FKHR exon which has been used as probe in the cosmid library screening, were amplified from murine 3T3 cell line genomic DNA and from RH28 genomic DNA. The PCR products were cloned and sequenced. The sequence of the RH28 derived PCR product clones exactly matched the published human FKHR gene sequence (Kishikawa et al., 1991). Two conservative base changes (*) were apparent in the sequence of the 3T3 cell line derived PCR product clones. Thus the absence of base changes in the cosmid clones strongly suggests that they are of human origin.

Published
sequence 5'-A ATT CGT CAT AAT CTG TCC CTA CAC AGC AAG TTC ATT CGT GTG CAG
Rh28 5'-A ATT CGT CAT AAT CTG TCC CTA CAC AGC AAG TTC ATT CGT GTG CAG
3T3 5'-A ATT CGT CAT AAT CTG TCC CTT CAC AGC AAG TTT ATT CGA GTG CAG

amino acids I R H N L S L H S K F I R V O

AAT GAA GGA ACT GGA AAA AGT TCT TGG TGG ATG CTC AAT CCA GAG GGT GGC AAG
AAT GAA GGA ACT GGA AAA AGT TCT TGG TGG ATG CTC AAT CCA GAG GGT GGC AAG
AAT GAA GGA ACT GGA AAG AGT TCT TGG TGG ATG CTC AAT CCA GAG GGA GGC AAG

N E G T G K S S W W M L N P E G G K

AGC GGG AAA TCT CCT AGG AGA AGA GCT GCA TCC ATG GAC AAC AAC AGT AAA TTT
AGC GGG AAA TCT CCT AGG AGA AGA GCT GCA TCC ATG GAC AAC AAC AGT AAA TTT
AGC GGA AAA TCA CCC CGG AGA AGA GCT GCG TCC ATG GAC AAC AAC AGT AAA TTT

S G K S P R R R A A S M D N N S K F

GCT AAG AGC CGA AGC-3'
GCT AAG AGC CGA AGC-3'
GCT AAG AGC CGA AGC-3'

Figure 4.3 Sequence differences between mouse and man in the FKHR gene. PCR products (primers underlined), of the 3' FKHR exon which had been used as probe in the cosmid library screening, were amplified from murine 3T3 cell line genomic DNA and from RH28 genomic DNA. The PCR products were cloned and sequenced. The sequence of the RH28 derived PCR product clones exactly matched the published human FKHR gene sequence (Galili et al, 1993). Ten conservative base changes (*) were apparent in the sequence of the 3T3 cell line derived PCR product clones. Thus the absence of base changes in the cosmid clones strongly suggests that they are of human origin.

4.2 ISOLATION OF NORMAL CHROMOSOME 2 AND 13 COSMIDS

To screen for chromosome 13 cosmid clones from a normal cosmid library, the same 3' FKHR exonic sequence probe was used. Chromosome 2 cosmid clones were sought by screening with a probe based on the published PAX3 exon 8 sequence (Barr et al, 1993). Two cosmid clones were isolated with each probe, which when digested contained similar restriction fragments, again suggesting that they originated from the same overlapping genomic region.

4.3 COSMID TO COSMID HYBRIDISATION

The two RHF14 cosmid clones which had been isolated as described above were digested with a number of different restriction enzymes to find out which would produce an appropriate series of fragments on an agarose gel. Restriction with TaqI resulted in 10 fragments ranging in size from 5kb to 500bp. DNA was transferred to Hybond-N+ filters by Southern blotting. Filters were then probed with vector sequence to determine which bands contained vector. The same filters were then stripped to remove the vector probe and probed with each normal cosmid probe which had been preassociated with Cot-1 DNA. Both cosmid probes hybridised to a 2kb band (see figure 4.4). The band was isolated from an agarose gel and used to probe filters containing digests of the normal PAX3 and FKHR cosmids (see figure 4.5). The 2kb probe was preassociated with Cot-1 DNA as it might have contained repetitive elements. It hybridised to specific bands from both sets of normal cosmids from chromosomes 2 and 13, suggesting that it contained sequences derived from both chromosomes. The 2kb band was cloned and sequenced. It contains a sequence of approximately 200bp with homology to a long interspersed repeat (LINE).

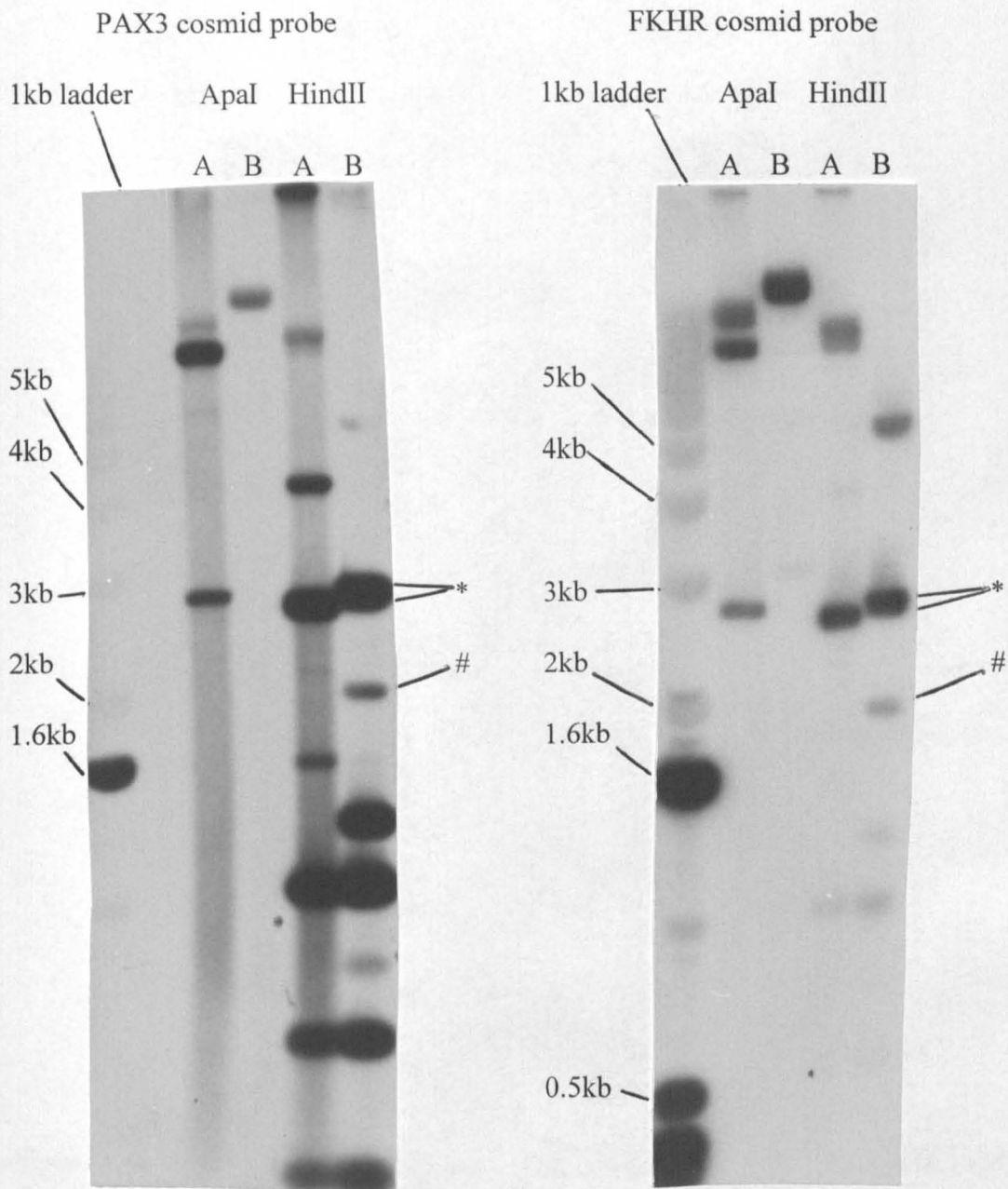


Figure 4.4 Cosmid/ cosmid hybridisation reveals a possible 2kb cross-hybridising restriction fragment. The two RHF14 cosmid clones were digested with *TaqI* or *HindII* and the DNA transferred to Hybond-N+ filters by Southern blotting. Filters were then probed with vector sequence to determine which bands contained vector (bands *). The same filters were then stripped to remove the vector probe and probed with each normal cosmid probe which had been preassociated with Cot-1 DNA. A 2kb band hybridised to both cosmid probes(#).

PAX3 cosmid

FKHR cosmid

HindII ApaI

HindII ApaI

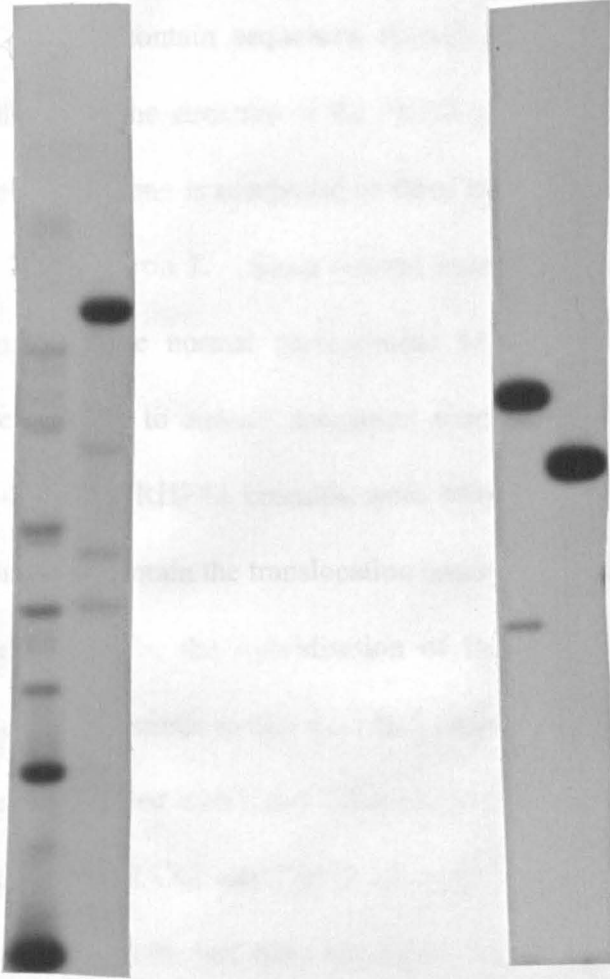


Figure 4.5 Hybridisation of the 2kb restriction fragment to PAX3 and FKHR cosmids.

The 2kb band was isolated from an agarose gel and used to probe filters containing HindII or ApaI digests of the normal PAX3 and FKHR cosmids. The 2kb probe was preassociated with Cot-1 DNA as it might have contained repetitive elements. It hybridised to specific bands from both normal cosmids, suggesting that it contained sequences derived from both chromosomes. The 2kb band was cloned and sequenced. It contains a sequence of approximately 200bp with homology to a long interspersed repeat (LINE).

At this point the sequence of the PAX3 intron 7 was published and submitted to the Genbank database. Comparison of the 2kb clone and intron 7 sequences indicated that the 2kb clone did not contain sequences derived from intron 7. Following this disappointing discovery the structure of the FKHR gene locus was described (Davis et al, 1995). The FKHR gene is composed of three exons and the t(2;13) translocations occur within a 130kb intron 1. Since cosmid inserts are on average 40kb in size it became apparent that the normal chromosome 13 cosmids that had been isolated previously were unlikely to contain sequences spanning the translocation breakpoint. Furthermore, since the RHF14 cosmids were selected using an FKHR gene based probe, they would not contain the translocation breakpoint either.

The likely explanation for the hybridisation of the 2kb probe to both the normal chromosome 2 and 13 cosmids is that the LINE repeat present within the 2kb sequence was not sufficiently blocked with Cot-1 DNA and could therefore hybridise to repetitive elements in the normal PAX3 and FKHR cosmids. This problem would have been highlighted if the 2kb clone had been hybridised to filters containing genomic DNA from normal tissue and from the parent RMS cell line which contains the chromosome translocation, Rh28.

On reflection, a better way of cloning the Rh28 translocation breakpoint would have been to sequence the PAX3 intron 7 starting from exon 8 through the breakpoint region of Rh28. It would be necessary to sequence approximately 2-3kb of the intron to cover the breakpoint region. The 2-3kb region of the intron immediately 5' of PAX3 exon eight also contains the breakpoint regions for two other t(2;13) bearing cell lines (personal communication, FG Barr). The sequence data could then be used to design a

probe to screen the RHF14 genomic cosmid library. By sequencing the relevant part of the PAX3 intron a probe could be designed which did not contain any repetitive elements. Cosmids which had previously been isolated could then be screened using the same probe for fragments containing breakpoint sequences.

CHAPTER 5 CLONING TRANSLOCATION BREAKPOINTS BY VECTORETTE PCR

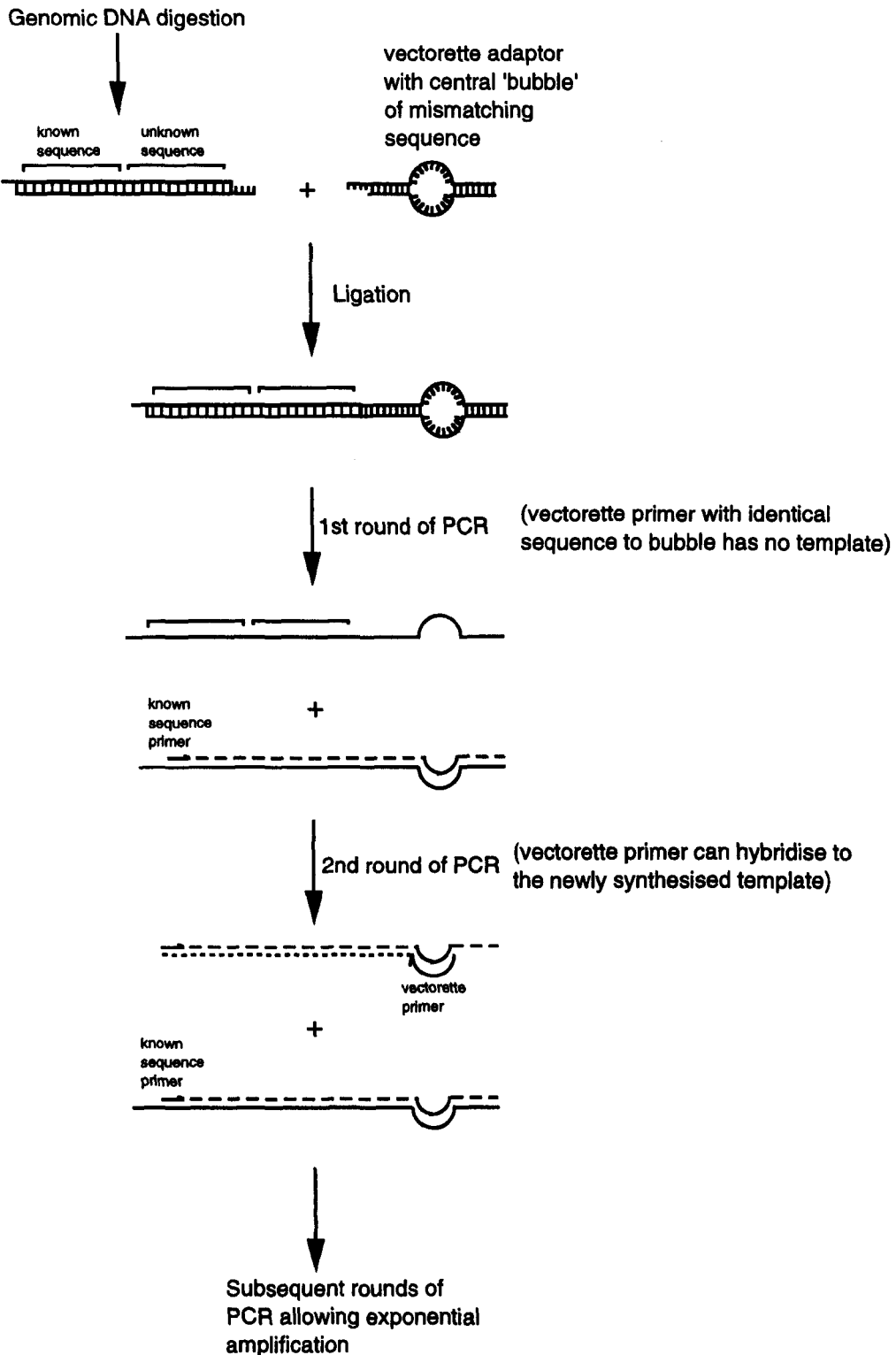
A vectorette PCR based approach was used to clone novel unknown chromosome 13 sequences that had been juxtaposed to known, chromosome 2, PAX3 intronic sequences as a result of the t(2;13) chromosome translocation. The method involves digestion of genomic DNA with an appropriate restriction enzyme, followed by ligation of compatible vectorette 'adaptors' to these ends to produce a vectorette 'library' (see figure 5.1; Riley et al,1990). Using an appropriately sited PAX3 intron primer together with a vectorette adaptor primer it should be possible to amplify a product spanning the translocation breakpoint by PCR. Non-specific amplification is reduced because of the design of the vectorette adaptors which allows only templates generated from the specific PAX3 primer to bind vectorette primer (see figure5.1). To confirm that any novel sequence cloned in this way is derived from chromosome 13, primers based on this sequence may be used to amplify a PCR product from a somatic cell hybrid which contains human chromosome 13 as its only human component (PGME1 or 289; Mitchell and Cowell, 1989; Zhong et al, 1992).

5.1 Cloning of the derivative chromosome 13 translocation breakpoint in Rh28.

The cell line, Rh28, contains the t(2;13)(q35;q14) chromosome translocation associated with alveolar rhabdomyosarcoma. Cytogenetic analysis indicates that Rh28 bears both the derivative and the normal chromosome 2, the derivative chromosome 13, but not the normal chromosome 13 (Douglass et al, 1987). A somatic cell hybrid, RHF14, has been generated, based on the RH28 cell line, which bears the derivative chromosome 13, but not the derivative or normal chromosome 2 (Mitchell et al, 1991). All the PCR-based experiments described in this section have been carried out on both

Figure 5.1 Vectorette PCR method. Genomic DNA from the translocation bearing cell line is digested with a restriction enzyme. Vectorette adaptors with compatible ends are ligated to the digested fragments to produce a vectorette 'library'. In the first round of PCR the primer based on the known sequence will bind to its template allowing synthesis of a DNA strand which spans the adjacent unknown sequence and the vectorette adaptor. The vectorette adaptor primer will not have a template to bind to, thus preventing synthesis of the second strand.

In the second round of PCR the known sequence primer will bind and allow strand synthesis as in the first round of PCR. The vectorette adaptor primer will bind to the strand that was synthesised in the first round of PCR, thus allowing synthesis of a second DNA strand which will span the adjacent unknown sequence and the known sequence. Exponential amplification can then proceed in further rounds of PCR.



Rh28 and RHF14 genomic DNA to ensure that amplification products come from the derivative chromosome 13 and not the normal chromosomes 2.

The genomic translocation breakpoint of the Rh28 cell line has been mapped previously to a 0.6kb region between a SacI and a ApaI site in the 16kb intron 7 of the PAX3 gene (Barr et al, 1993). The presence of these two restriction sites was confirmed in the complete sequence of the intron in the Genbank database (Macina et al, 1995; accession no. U12259). To verify that intron sequences 5' of the SacI site were still present in the derivative chromosome 13, genomic DNA from Rh28 and the associated RHF14 somatic cell hybrid were used as template to amplify a predicted 484 base-pair fragment from positions 12794 to 13278 in the intron, which should include a PstI site at position 12845 (see figure 5.2b). The predicted amplification product was identified and digestion with PstI restriction enzyme resulted in two bands of the predicted size, 51 and 433 base-pairs, confirming the presence in both RH28 and RHF14 of the relevant intron sequences.

Four vectorette libraries in total were then constructed from genomic DNA from Rh28 or RHF14 cell lines, one pair using HaeIII restriction enzyme and the other pair using Sau3A restriction enzyme. Nested primers, based on sequences immediately 5' of the SacI site, were designed to be used in conjunction with vectorette adaptor primers. The HaeIII and Sau3A restriction enzymes were chosen because their recognition sites are not present within the breakpoint region nor between the primer binding and SacI sites. Primers 12800F and UVP were used in a first round of amplification, and nested primers 13300F and USP were used in a second round of amplification. A 200bp band was amplified from the Rh28/HaeIII and the RHF14/HaeIII vectorette libraries and a 400bp band was amplified from the Rh28/Sau3A and the RHF14/Sau3A vectorette libraries (see figure 5.2c).

Figure 5.2 Schematic representation of the two normal and two derivative chromosome breakpoint regions in the Rh28 cell line.

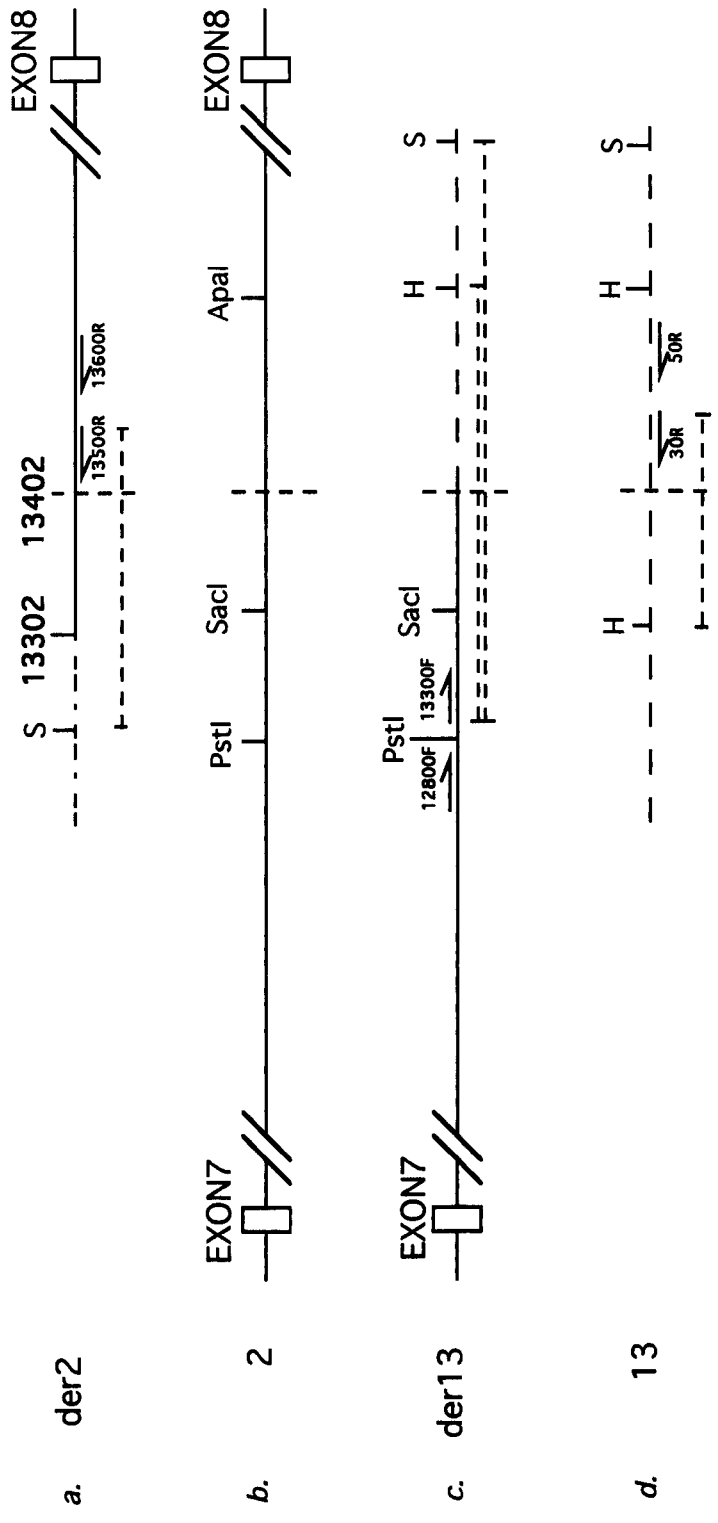
a, Derivative chromosome 2. The vectorette PCR product extends from primer 13500R to a Sau3A restriction site (S). The derivative 2 breakpoint is at position 13302, 101 bp from the derivative 13 breakpoint at position 13402.

b, Normal chromosome 2. The breakpoint region in Rh28 has been mapped to a 0.6kb region between a SacI and a ApaI site within the 17.5kb PAX3 intron 7.

c, Derivative chromosome 13. PCR was used to verify the presence of sequences immediately 5' of the SacI site, including a PstI site at position 12845, in the derivative chromosome 13. Vectorette PCR products span the breakpoint from primer 13300F to either the HaeIII (H) or the Sau3A(S) site. The breakpoint is at position 13402 in the PAX3 intron.

d, Normal chromosome 13. The vectorette PCR product from primer 30R to a HaeIII site (H) spans the breakpoint region on normal chromosome 13.

Figure 5.2



All four products were cloned and sequenced and in all four cases the sequence obtained indicates that the genomic breakpoint on the derivative chromosome 13 occurs at position 13402/3 of the PAX3 intron. Novel sequence, presumed to represent intron sequence within the FKHR gene on chromosome 13, was identified lying 3' of the breakpoint in both Sau3A and HaeIII library clones. Two hundred and sixty nine base-pairs of novel sequence were identified in the Sau3A clones, of which the first 66 base-pairs exactly matched those present in the HaeIII clones (see figure 5.3).

5.2 Cloning of chromosome 13 sequences which span the translocation breakpoint in Rh28.

As the cytogenetic analysis of Rh28 indicates that there is no normal chromosome 13 present, the vectorette libraries used to clone derivative chromosome 13 breakpoints described above would not be appropriate for cloning normal chromosome 13 products. However, reverse transcriptase PCR (RT-PCR) indicates that the normal FKHR gene product is expressed in Rh28 (Galili et al, 1993). This finding suggests that either a complete or a partial chromosome 13 is present in the cell line. Therefore, the same vectorette libraries were used to attempt to clone normal chromosome 13 sequences spanning the breakpoint in Rh28. The novel sequences found 3' of the derivative 13 translocation breakpoint were used to design a second pair of nested primers. Primers 50R and UVP were used in a first round of amplification as above. Nested primers 30R and USP were used in second round amplification. A 150bp and a 800bp product were amplified from the Rh28/HaeIII vectorette library and a similar 800bp product was amplified from the RHF14/HaeIII vectorette library (see figure 5.2d).

The 800bp products are of the size predicted if amplification was from the derivative chromosome 13 template. The 150bp product, present in Rh28 only, ought to represent normal chromosome 13 sequence. No PCR product could be identified from the

Figure 5.3 Rh28 derivative chromosome 13 breakpoint sequence from PAX3 intron primer 13300F to Sau3A restriction site (both underlined). The breakpoint occurs at position 13402 of the PAX3 intron (v). 269bp of novel chromosome 13 sequence follows the breakpoint. Putative Translin recognition sequences occur immediately 5' and 3' of the breakpoint (double underlined).

1

TACATATAGCTCATGGGAAAGTGTGCATGTTTTGAATAAAAATTCTGCTCTGAGCTCATCA
-----+-----+-----+-----+-----+-----+-----+
ATGTATATCGAGTACCCTTTCACACGTACAAACTTATTTTAAGACGAGACTCGAGTAGT

AGTATAACTGTACATACAGCATTCCATTTCTCCCTTAAAGTGACATAACCATTGACTGAAT
-----+-----+-----+-----+-----+-----+-----+
TCATATTGACATGTATGTCGTAAGGTAAAGAGGGAATTTCACTGTATGGTAACTGACTTA

U

TTCAAGTAAGTATCTTCATGAAGGACAATATACATTTTAAGACCTGCTCTAGGGGTTGT
-----+-----+-----+-----+-----+-----+-----+
AAGTTCATTCATAGAAGTACTTCCTGTTATAATGTAAAATTCTGGACGAGATCCCCAACA

AAAATAGTACTGCTGTTATTGGGAGGCGTTGGGTAGGGGCCACAGCAGCAGGGAGAGAGA
-----+-----+-----+-----+-----+-----+-----+
TTTTATCATGACGACAATAACCCTCCGCAACCCATCCCCGGTGTGTCGTCCCTCTCTCT

GGCTCCCATCGTACAACTTGCAGACTTTTATTAAAACGTATTTATTGATTGCAAATTAT
-----+-----+-----+-----+-----+-----+-----+
CCGAGGGTAGCATGTTTGAACGTCTGAAAATAATTTTGCATAAATAACTAACGTTTAATA

GTCAACAGAGTGAGGCTTTGGTAACTCTTCAAGTTACTAGTGTTTCCTTCTTTGCATAGT
-----+-----+-----+-----+-----+-----+-----+
CAGTTGTCTCACTCCGAAACCATTGAGAAGTTCAATGATCACAAAGGAAGAAACGTATCA

TGTTTCCCATCACCTTTAAACATCTGCTGCTTCTGGTCTTGTCCAGTCATGCTTGAGTG
-----+-----+-----+-----+-----+-----+-----+
ACAAAGGGTAGTGGGAAATTTGTAGACGACGAAGACCAGAACAGGTCAGTACGAACTCAC

ATC

TAG

experiments using the Rh28/Sau3A or the RHF14/Sau3A vectorette libraries, possibly because the nearest Sau3A site is too far from the breakpoint to allow efficient amplification. The 150bp amplification product was then cloned and sequenced. The 3' sequence immediately adjacent to the 30R primer matches that of the derivative 13 clones up to the breakpoint. The clone also contained 111bp of unique sequence 5' of the breakpoint (see figure 5.4). Together with the sequences 3' of the breakpoint on the derivative chromosome 13, a total of 380bp of novel sequence have been cloned. To confirm that this represents a contiguous sequence on chromosome 13, a primer pair close to the two ends of this unique sequence were designed. The expected 313bp product could be amplified, cloned and sequenced from Rh28, Placenta, PGME1 and 289 genomic DNA. Figure 5.4 includes the entire sequence obtained from chromosome 13.

5.3 Cloning of the derivative chromosome 2 translocation breakpoint in Rh28.

To determine whether there have been any insertions and/or deletions associated with the chromosome translocation a similar approach to that above was used to clone the reciprocal translocation breakpoint on derivative chromosome 2. Nested primers, 13600R and 13500R, were designed, based on sequences in the PAX3 intron immediately 3' of position 13402/3, now established as the translocation breakpoint position on the derivative chromosome 13. It is possible that these sequences are deleted from derivative chromosome 2 but there was no straightforward way of testing this hypothesis as a somatic cell hybrid cell line containing only derivative chromosome 2 material was not available.

Amplification of the Rh28/HaeIII and Rh28/Sau3A vectorette libraries was carried out with primers 13600R and UVP, followed by nested 13500R and USP primers. A 300bp product was amplified from the Rh28/Sau3A vectorette library and a 900bp

Figure 5.4 Normal chromosome 13 sequence which spans the Rh28 derivative chromosome 13 breakpoint. Chromosome 13 based primer 30R and HaeIII restriction site are underlined. 111bp of novel sequence 5' of the derivative 13 breakpoint have been cloned. This together with the novel sequence cloned above represents a contiguous 380bp of chromosome 13.

2

GGCCTTGTTCTTGAGAATAGGTTTCTTGTTTATCTGAATGGCTTTTCTTTGCGTCACTA
-----+-----+-----+-----+-----+-----+-----+
CCGGAACAAGAACCTCTTATCCAAGAACAATAGACTTACCGAAAAGAAACGCAGTGAT

TGCCAGATAGTCACTCTTGAGCAGAAATGATGGTTTTAAAATCTCTCTTACATTTTAAG
-----+-----+-----+-----+-----+-----+-----+
ACGGTCTATCAGTGAGAACCTCGTCTTTACTACCAAATTTTAGAGAGAATGTAAAATTC

ACCTGCTCTAGGGGTTGTAAAATAGTACTGCTGTTATTGGGAGGCGTTGGGTAGGGGCCA
-----+-----+-----+-----+-----+-----+-----+
TGGACGAGATCCCCAACATTTTATCATGACGACAATAACCTCCGCAACCCATCCCCGGT

CAGCAGCAGGGAGAGAGAGGGCTCCCATCGTACAACTTGCAGACTTTTATTAAAACGTAT
-----+-----+-----+-----+-----+-----+-----+
GTCGTCGTCCCTCTCTCTCCGAGGGTAGCATGTTTGAACGTCTGAAAATAATTTTGCATA

TTATTGATTGCAAATTATGTCAACAGAGTGAGGCTTTGGTAACTCTTCAAGTTACTAGTG
-----+-----+-----+-----+-----+-----+-----+
AATAACTAACGTTTAATACAGTTGTCTCACTCCGAAACCATTGAGAAGTTCAATGATCAC

TTTCCTTCTTTGCATAGTTGTTTCCCATCACCTTTAAACATCTGCTGCTTCTGGTCTTG
-----+-----+-----+-----+-----+-----+-----+
AAAGGAAGAAACGTATCAACAAAGGGTAGTGGGAAATTTGTAGACGACGAAGACCAGAAC

TCCAGTCATGCTTGAGTGATC
-----+-----+-----+-----+-----+-----+-----+
AGGTCAGTACGAACTCACTAG

product amplified from the Rh28/HaeIII vectorette library (see figure 5.2a). The 900bp product was of the size predicted for normal chromosome 2. The 300bp product, presumed to represent derivative chromosome 2, was cloned and sequenced (see figure 5.5). The first 164bp of sequence extending from the 13500R primer matched the original chromosome 2 sequence, indicating that the translocation breakpoint in the derivative 2 chromosome is at position 13301/2 of the PAX3 intron, which therefore differs from the breakpoint in the derivative 13 chromosome at position 13402/3. As 101bp of chromosome 2 sequence is present in both derivative chromosomes, there appears to have been a duplication in this region. The final 82bp sequence of this clone was again novel and did not match the newly-derived chromosome 13 sequence described above. The 82bp sequence could be amplified from Rh28 and placental genomic DNA, but not from chromosome 2 or chromosome 13 monochromosomal somatic cell hybrids (GM10826B, 289 and PGME1). That the 82bp sequence is not derived from either chromosome 2 or chromosome 13 suggests that an insertion of material from another chromosome may have occurred.

5.4 Cloning the derivative 13 breakpoints in other t(2;13) bearing cell lines

The breakpoint regions from five other t(2;13) bearing cell lines have been mapped to distinct regions within intron 7 of the PAX3 gene (see figure 5.6, personal communication from Dr Fred Barr). The breakpoint regions for the cell lines CW12 and TTC487 have been mapped to overlapping regions of intron 7 approximately 9kb 5' of the Rh28 breakpoint. The Rh5 breakpoint maps to a region 2kb 5' of the Rh28 breakpoint. The Rh30 and Rh18 breakpoints have been mapped to regions approximately 1kb and 2kb 3' respectively of the Rh28 breakpoint (see figure 5.6). Vectorette libraries were constructed from genomic DNA of each of these cell lines using appropriate restriction enzymes and vectorette adaptors (see table 5.1). Somatic cell hybrids containing the derivative 13 chromosomes of these cell lines are not

GATCCTCAGAGAGTTGTGGCGATAGTTACTAGAACATAGCAAACTAATTGGGCAAAGG
 -----+-----+-----+-----+-----+-----+-----+
 CTAGGAGTCTCTCAACACCGCTATCAATGATCTTGTATCGTTTTGATTAACCCGTTTTCC

v

GTTCTGCTCGCTGTGCTCTGTGGCTCATCAAGTATAACTGTACATACAGCATTCCATTTC
 -----+-----+-----+-----+-----+-----+-----+
CAAGACGAGCGACACGAGACACCGAGTAGTTCATATTGACATGTATGTCGTAAGGTAAAG

TCCCTTAAAGTGACATACCATTGACTGAATTTCAAGTAAGTATCTTCATGAAGGACAATA
 -----+-----+-----+-----+-----+-----+-----+
 AGGGAATTTCACTGTATGGTAACTGACTTAAAGTTCATTCATAGAAGTACTTCCTGTTAT

TTAATTTAGAGTTTATTTTCCTTTTATTTAGTGCCACTGGATATCTTTCCTCTCTGGACT
 -----+-----+-----+-----+-----+-----+-----+
AATTAATCTCAAATAAAAGGAAAATAAATCACGGTGACCTATAGAAAGGAGAGACCTGA

GATATAG

CTATATC

Figure 5.5 Rh28 derivative chromosome 2 breakpoint sequence from PAX3 intron primer 13500R to Sau3A restriction site (both underlined). This breakpoint occurs at position 13301 of the PAX3 intron (v) indicating that 101bp of PAX3 intron sequence has been duplicated. Somatic cell hybrid PCR suggests the remaining 82bp of sequence (bold) is not derived from either chromosome 2 or chromosome 13.

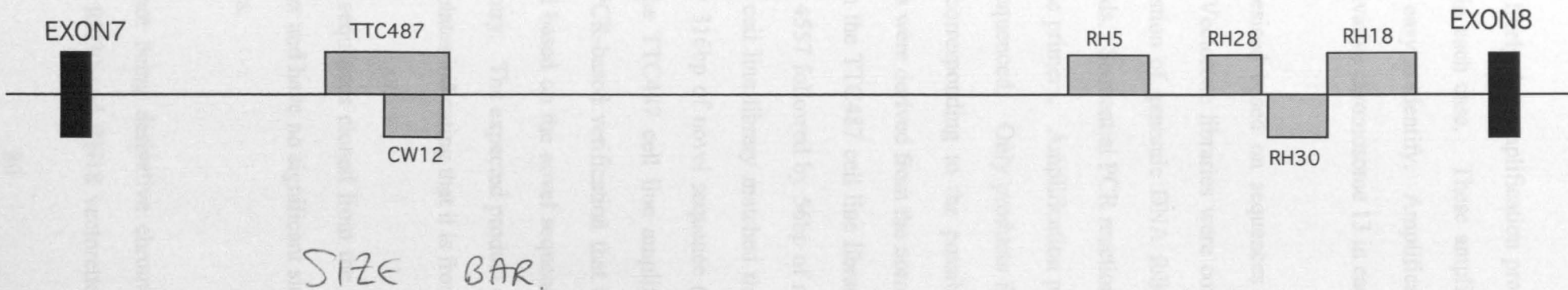


Figure 5.6 The breakpoint regions for the six t(2;13) cell lines used in this study have been mapped to the positions indicated within the PAX3 intron seven. The breakpoint regions for cell lines CW12, Rh28 and Rh18 were described in Barr et al, 1993. The TTC487, Rh5 and Rh30 breakpoint regions were determined by Dr. F.G.Barr (personal communication).

available. It was therefore likely that amplification products would be produced from the normal chromosome 2 in each case. These amplification products should be of known size and will thus be easy to identify. Amplification products of different sizes could originate from the derivative chromosome 13 in each case.

Nested primers were synthesised based on sequences immediately 5' of breakpoint regions in each cell line. Vectorette libraries were constructed using the appropriate restriction enzyme for digestion of genomic DNA followed by ligation of vectorette adaptors with compatible ends. Sequential PCR reactions were performed with pairs of intron specific and vectorette primers. Amplification products from all five vectorette libraries were cloned and sequenced. Only products from the TTC487 and RH5 cell lines contained sequences corresponding to the possible derivative chromosome 13. Products from other libraries were derived from the normal chromosome 2 or were non-specific. The product from the TTC487 cell line library matched the sequence of the PAX3 intron up to position 4557 followed by 56bp of novel sequence (see figure 5.7). The product from the RH5 cell line library matched the PAX3 intron sequence up to position 11099 followed by 316bp of novel sequence (see figure 5.8). The 56bp of novel sequence found in the TTC487 cell line amplification product is not a large enough template to allow PCR-based verification that it is derived from chromosome 13. Primers were designed based on the novel sequence found in the clone amplified from the RH5 vectorette library. The expected product could be amplified from PGME and 289 genomic DNA templates indicating that it is from chromosome 13.

The novel chromosome 13 sequences cloned from the RH28, RH5 and TTC487 cell lines do not match each other and have no significant similarity to any sequences in the Genbank or EMBL databases.

An explanation for there not being derivative chromosome 13 based amplification products from the CW12, RH30 and RH18 vectorette libraries may be that the gap

U

CTGGGTCTCTCTCCACAGGACTTGTTTGCTTCCTCTCTGATTGTTTTGTCAAAGCTTTAG
 -----+-----+-----+-----+-----+
 GACCCAGAGAGAGGTGTCCTGAAACAAACGAAGGAGAGACTAACAAAACAGTTTCGAAATC

TGTTTTCTTATGAGGTGCTTTTGATTGAAACCTCATCAATTTAAAACCTTATCAG
 -----+-----+-----+-----+-----+
 ACAAAGAATACTCCACGAAAATACTTTGGAGTAGTTAAATTTTGAATAGTC

Figure 5.7 TTC487 derivative chromosome 13 sequence. The breakpoint occurs at position 4557 of the PAX3 intron (U). Putative Translin recognition sequences are double underlined.

Figure 5.8 Rh5 derivative chromosome 13 sequence. The breakpoint occurs at position 11099 of the PAX3 intron (v). Putative Translin recognition sequences are double underlined.

GTTATGCAACTCATTCTGCGCTAATTCTCTGTCTAGATTTATATTTTAGTCATCTTAGG
-----+-----+-----+-----+-----+-----+-----+
CAATACGTTGAGTAAAGACGCGATTAAGAGACAGATCTAAATATAAAATCAGTAGAATCC

GTAGTCTCCTTGATTACTTTTTTTTTTTTTCTTGAGAATGGGTCAGGTTTTCCATTTT
-----+-----+-----+-----+-----+-----+-----+
CATCAGAGGAACTAATGAAAAAAAAAAAAAAAAAGAACTCTTACCCAGTCCAAAAAGGTAAAG

TTCTTATAGGGAGGTAATTTTGGATTATATTCTGGATATTACAAATGATAAATGATAATT
-----+-----+-----+-----+-----+-----+-----+
AAGAATATCCCTCCATTAACCTAATAAGACCTATAATGTTTACTATTTACTATTAA

TATATATATAAATTTTCGTTATGCCCGTCTTAAAAGCATTGATTTTCTTTTCTAAAGCAG
-----+-----+-----+-----+-----+-----+-----+
ATATATATATTTAAAAGCAATACGGGCAGAATTTTCGTAACATAAAGAAAAGATTTTCGTC

GCAATTGATTAACCTTACCTGAATTCAAGCTGCAAACTATCCCCCATAGTAGGTGGCAG
-----+-----+-----+-----+-----+-----+-----+
CGTAACTAATTGAATGGACTTAAGTTCGACGTTTTGATAGGGGGGTATCATCCACCGTC

TGTACATTTAGTTTCAAGTCTCTTAGCTGGCCTACTTGGAGCCTGCTCTGTGCATGTGTG
-----+-----+-----+-----+-----+-----+-----+
ACATGTAAAGTCAAGTCAAGAGAATCGACCGGATGAACCTCGGACGAGACACGTACACAC

TTTCACGGATC
-----+-----
AAAGTGCCTAG

between intron specific- and vectorette primers is too large to allow efficient amplification. In anticipation of this problem the vectorette libraries were constructed using frequent cutting restriction enzymes, for example Sau3AI or TaqI, which have four base recognition sites. However, technical problems such as inefficient primer binding or vectorette library construction may have been the cause. Because it is necessary to clone PCR products and sequence clones following amplification it was not feasible to test a number of vectorette PCR conditions.

Table 5.1 Restriction enzymes used in vectorette library construction for t(2;13) bearing cell lines.

CELL LINE	Restriction enzyme	breakpoint region	region size
CW12	Sau3AI	4552-4741	189bp
TTC487	TaqI	4221-4741	520bp
RH5	Sau3AI	11078-11947	869bp
RH30	TaqI	13935-14478	543bp
RH18	TaqI	14478-15444	966bp

5.5 Cloning of normal chromosome 13 sequences spanning the breakpoints in cell lines RH5 and TTC487

RT-PCR analysis indicates that the normal FKHR transcript is expressed in RH5, suggesting that the normal intact locus is present (Galili et al, 1993). RT-PCR analysis has not been carried out on TTC487. The same RH5 and TTC487 vectorette libraries have been used to clone sequences from normal chromosome 13 which span the translocation breakpoint. Since it is not known whether the TTC487 cell line has an intact FKHR gene locus, another vectorette library was constructed from human placental genomic DNA using the same restriction enzyme, TaqI, as in the TTC487 vectorette library. Several amplification products were cloned and sequenced from

each of these vectorette libraries but in every case were either from the derivative 13 chromosome or were non-specific.

5.6 Analysis of sequences around the translocation breakpoints

The whole 16.6kb intron7 of the PAX3 gene on chromosome 2 has been sequenced (Macina et al, 1995). Three microsatellite and two Alu repeat sequences are present within the intron. These common sequences are not close to any of the breakpoints in the cell lines TTC487, RH5 or RH28 which suggests that they are unlikely to be involved in the mechanism by which the t(2;13) chromosome translocations have occurred. Recombinase protein signal sequences and oligopurine/oligopyrimidine motifs have been implicated in chromosome translocation mechanisms in lymphoid malignancies; these motifs could not be found in a search of the sequences around the t(2;13) translocation breakpoints.

Recently a novel protein, now called Translin, has been identified which specifically binds to sequences found adjacent to breakpoint junctions of chromosome translocations in many cases of lymphoid malignancy (Aoki et al, 1995). The authors compared nucleotide sequences at the 5' flanking site of chromosome breakpoint junctions in 91 cases of human lymphoid neoplasms. These included translocations involving immunoglobulin gene loci, T-cell receptor gene loci and others not involving loci which normally undergo gene rearrangement. The consensus sequence derived from the 91 cases is;

A(65%), T(88%), G(85%), C(75%), A(96%), G(78%), and

G(86%), C(76%), C(63%), A/T(40%/56%), C(68%), C(64%), T(45%).

The consensus is in two parts which maybe separated by up to four bases. The same Translin protein can also bind to a dimer of the second part of the consensus sequence. The sequences flanking breakpoint junctions usually match the consensus sequence at 7 to 10 of the 13 bases (54-77% homology).

Analysis of the sequences in the derivative 13 chromosomes in the three cell lines above indicate the presence of potential Translin binding sites both 5' and 3' of the breakpoints (see table 5.2)

Table 5.2 Translin binding sites which flank t(2;13) breakpoints in alveolar rhabdomyosarcoma with published translin sites which flank other breakpoints showing similar homology to the consensus translin binding site.

	breakpoint position	translin site	% homology	reference
consensus	-	ATGCAGGCC (A/T)CCT	-	Aoki <i>et al</i> 1995
bcr/abl	-	TTGCAGTGA G CCG	62	Chen <i>et al</i> 1989
TCR/TTG2	-	GTGAAGGCT C CGC	62	Boeh <i>et al</i> 1988
IgH/bcl6	-	CTGCACGTG C GAT	43	Baron <i>et al</i> 1993
Rh28	13402	ATGAAGGAC A ATA	62	
		TTTAAGACC T GCT	62	
Rh5	11099	ATGCAACTC A TTT	62	
		AGGGTAGTC T CCT	62	
TTC487	4557	TTGTTTGCT T CCT	62	
		ATGCATATA T GTA	43	

Analysis of sequences around breakpoints in both derivative chromosomes of the Rh28 cell line suggests that inter-chromosomal translocations are more complex at the sequence level than might have been expected. There has been both duplication of 101bp of PAX3 intron 7 and insertion of DNA of undetermined size into the derivative

chromosome 2. The inserted DNA does not appear to originate from chromosomes 2 or 13.

The striking feature of the sequences flanking the derivative chromosome 13 in the RH28 cell line and also in the two other cell lines analysed is the presence of potential Translin protein DNA-binding sites. Similar sites have been found adjacent to chromosome translocation breakpoints in a wide range of lymphoid neoplasms. To determine whether these newly identified sequences can act as Translin binding sites, a series of studies based on *in vitro* gel shift analysis were then carried out (see next chapter).

CHAPTER 6 PROTEIN BINDING STUDIES

A number of techniques are available to characterise DNA-Protein interactions which can also serve as a basis for isolating DNA-binding proteins. The preparation of cell extracts from the cytoplasm or nucleus is usually the first step in any of these techniques. In DNaseI protection mapping (footprinting) the region or regions to which a protein binds, either in co-operation with other proteins or on its own, can be elucidated (Galas and Schmitz, 1978). For example, footprinting can be used to find the binding region of a regulatory protein in several hundred base pairs of a gene promoter sequence. However, a more rapid and simple method for studying DNA-protein interactions is the electrophoretic mobility shift assay (EMSA). The basic protocol involves allowing binding of protein to radiolabelled DNA oligonucleotide and separating the DNA-Protein complex from unbound DNA by electrophoresis through a non-denaturing, low ionic strength polyacrylamide gel. Under these conditions, covalent bonding between DNA and protein remains intact and the mobility of a DNA-Protein complex is significantly retarded, relative to unbound oligonucleotide, because of the size and charge of the protein. By addition of a large mass of alternating copolymer DNA (poly[dI.dC].poly[dI.dC]), specific binding of protein from a crude extract can be detected. The copolymer provides a large number of non-specific sites for protein to bind to, thus leaving labelled DNA sequences open to the specific binding of proteins of interest (Carthew et al, 1985; Singh et al, 1986).

In the previous chapter t(2;13) chromosome translocation breakpoints were cloned and sequenced. The sequences immediately adjacent to these breakpoints appear to be

significantly homologous to the consensus binding sequence for Translin protein (Aoki et al, 1995). To test whether these breakpoint flanking sites will indeed bind Translin protein the EMSA method was employed.

The ability for Translin protein to bind to DNA oligonucleotides by EMSA has been studied previously by Aoki et al (1995). They were able to demonstrate a number of features of Translin protein binding;

1. Translin will bind to both double-stranded and single-stranded DNA oligonucleotides, but with higher affinity to single-stranded oligonucleotides.
2. Translin will bind specifically to oligonucleotides based on sequences found adjacent to chromosome breakpoints in a range of lymphoid neoplasms.
3. A consensus translin binding sequence can be deduced from these sequences. Translin will have little or no binding activity to oligonucleotides with sequences unrelated to this consensus.
4. An antibody (raised by Aoki, et al) against Translin protein will interfere with oligonucleotide-protein complex formation.
5. Translin protein is found in the cytoplasm and nucleus of lymphoid cell lines such as RPMI8226, but only the cytoplasm of non-lymphoid cell lines such as HeLa, A431 and K562.

Translin protein has been purified to homogeneity using the gel mobility shift assay as a basis for fractionation and the gene coding for Translin was cloned and sequenced. It codes for a 228 amino acid protein which does not share homology to any other known protein. A 'leucine zipper' motif, five potential protein kinase C phosphorylation sites,

three tyrosine kinase phosphorylation sites and one N-glycosylation site are present in the Translin amino acid sequence (Aoki et al, 1995).

These data were used to design experiments to determine if Translin protein was present in RMS cell lines and, if so, whether it would bind to the sequences found flanking chromosome translocation breakpoints in the cell lines RH28, RH5 and TTC487.

6.1 Binding of protein from Rhabdomyosarcoma cell lines to oligonucleotides

Whole cell protein extracts were prepared from RD, an embryonal rhabdomyosarcoma cell line, Rh30, a t(2;13) bearing alveolar rhabdomyosarcoma cell line, and several other cell lines, HeLa, A431, K562, which previously have been shown to contain Translin protein (Aoki et al, 1995). EMSA assays are generally carried out with extracts of nuclear proteins as most transcription factors are located in the nucleus. To ensure that any Translin protein in these cell lines was present in the extract, regardless of its initial intracellular location (see point 5 above), whole cell extracts rather than nuclear extracts were prepared.

Extracts from cell lines RD, RH30 and HeLa were used in EMSA to look for binding of protein to four different oligonucleotides. Single-stranded oligonucleotides were used for this and all the following experiments. The oligonucleotides were;

1. 'consensus' 5'-ATTTCATGCAGGCCCTCCTAGTCA-3'. The middle 13 bases of this oligonucleotide constitute the consensus sequence for Translin protein DNA-binding.

2. '5'RH28' 5'-ATTTCATGAAGGACAATAAGTCA-3'. The middle 13 bases of this oligonucleotide constitutes a putative Translin binding site which is immediately 5' of the derivative chromosome 13 breakpoint in the t(2;13) chromosome translocation in the cell line Rh28.

3. '3'RH28' 5'-ATTTCTTTAAGACCTGCTAGTCA-3'. The middle 13 bases of this oligonucleotide constitutes a putative Translin binding site which is immediately 3' of the derivative chromosome 13 breakpoint in the t(2;13) chromosome translocation in the cell line Rh28.

4. 'nonspecific' 5'-ATTTCCACTATCGACTACAGTCA-3'. The middle 13 bases of this oligonucleotide are unrelated to any Translin binding site.

In EMSA, free labelled oligonucleotide probe will migrate rapidly to the bottom of the polyacrylamide gel. Any labelled oligonucleotide which has protein bound to it will migrate more slowly through the gel. In the initial experiment two slowly migrating bands are present (figure 6.1, bands A&B). The faster migrating band B is present in incubations with all four oligonucleotides and with all three extracts tested. The intensity of band B in the HeLa cell line extract is much less than that seen with the RD and RH30 cell line extracts. For all three cell lines tested, the slower migrating band A is present in the lanes containing the 'consensus' oligonucleotide and the lanes containing the '5'RH28' and '3'RH28' oligonucleotides. However this band is not present in the lane containing the 'nonspecific' oligonucleotide which is unrelated to a translin binding site.

Figure 6.1 Binding of protein in the electrophoretic mobility shift assay (EMSA). Extracts from cell lines RD, RH30 and HeLa were used in EMSA to look for binding of protein to four different oligonucleotides; one based on the consensus translin recognition sequence, two based on putative translin recognition sequences found flanking t(2;13) breakpoints, and one unrelated sequence. Unbound, ³²P-labelled oligonucleotide probe will migrate rapidly to the bottom of the polyacrylamide gel (band C). Any labelled oligonucleotide which has protein bound to it will migrate more slowly through the gel. In this experiment two slowly migrating bands are present (figure 4.1, bands A and B). The faster migrating band B is present in incubations with all four oligonucleotides and with all three extracts tested. The intensity of band B in the HeLa cell line extract is much less than that seen with the RD and RH30 cell line extracts. For all three cell lines tested, the slower migrating band A is present in the lanes containing the 'consensus' oligonucleotide and the lanes containing the '5'RH28' and '3'RH28' oligonucleotides. However this band is not present in the lane containing the 'nonspecific' oligonucleotide which is unrelated to a translin binding site.

RD

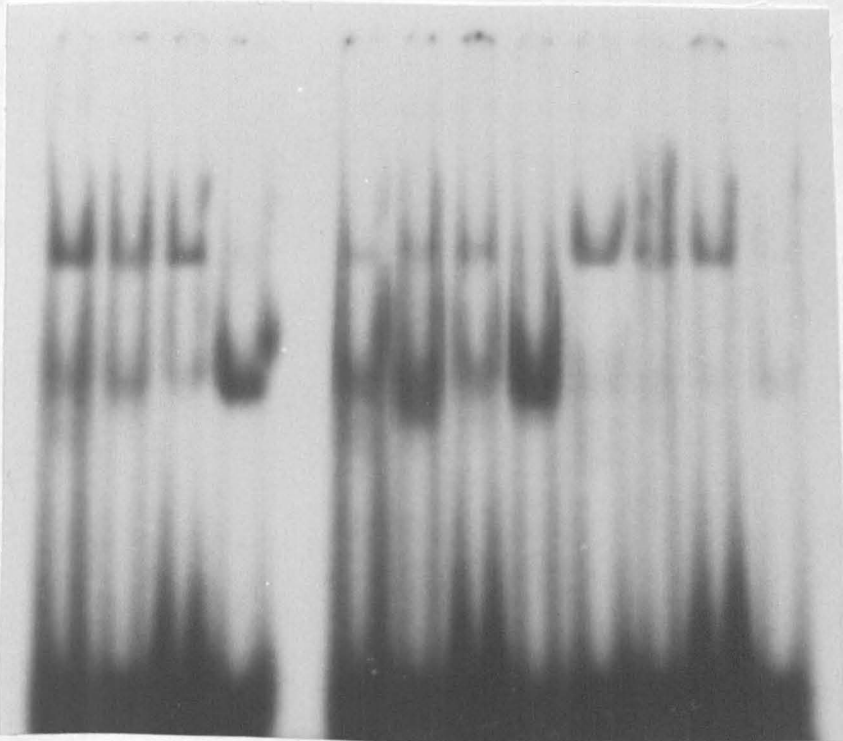
Rh30

Hela

A

B

C



HeLa cells are known to contain Translin protein which will bind to specific sequences in gel mobility shift studies (Aoki et al, 1995). Protein derived from rms cell lines RD and Rh30 will bind to the same sequence resulting in the appearance of bands which migrate at the same rate. Bands which migrate at the same rate also appear when all three extracts are exposed to oligonucleotides based on sequences found adjacent to the derivative chromosome 13 breakpoint in the cell line RH28.

6.2 Specificity of oligonucleotide-protein binding

Is the binding of protein to oligonucleotide in the previous experiment dependent on the sequence of bases? To answer this question competition experiments were carried out. EMSA reactions were set up as described above, but included in the incubations were increasing quantities of unlabelled oligonucleotide which will compete with the labelled oligonucleotide for protein. If the competing oligonucleotide has the same sequence as the labelled oligonucleotide the intensity of the gel shift band should be reduced. When the competing oligonucleotide has an unrelated sequence the intensity of the gel shift band may be reduced, indicating binding is not sequence specific. Alternatively, the intensity of the gel shift band will not be reduced indicating binding is sequence specific.

Two experiments were performed with either the 'consensus' oligonucleotide or the '5'RH28' oligonucleotide. The RD, RH30 and HeLa extracts plus the two other Translin bearing extracts A431 and K562 were tested. In each experiment, for each extract, a set of four EMSA reactions were performed. The first in each set of four

contained no competing unlabelled oligonucleotide. In the remaining three tubes of each set of four, a 40, 80 and 160 fold molar excess of competing unlabelled oligonucleotide was included. Figure 6.2 shows an experiment in which labelled 'consensus' oligonucleotide is competing with increasing quantities of unlabelled 'consensus' oligonucleotide. The band intensity is gradually reduced with increasing unlabelled oligonucleotide for all of the extracts tested. When the labelled 'consensus' oligonucleotide is competing with equivalent increasing quantities of unlabelled 'nonspecific' oligonucleotide band intensity is not reduced (see figure 6.3).

The competition experiments described here indicate that the binding of protein to oligonucleotide requires sequence specificity

6.3 Sizing of protein that binds to oligonucleotides

The above EMSA experiments were performed on non-denaturing polyacrylamide gels, which are not suitable for determining the molecular mass of the protein which is binding to oligonucleotide. The molecular mass of the particular protein in a whole cell extract that is binding to labelled oligonucleotide can be measured by irreversibly crosslinking the protein to the oligonucleotide using UV-light, and separating the complex on a denaturing polyacrylamide gel together with a set of molecular weight markers.

A series of incubations with RD protein extract and 'consensus' labelled oligonucleotide were exposed to UV light for varying lengths of time. The samples were separated by

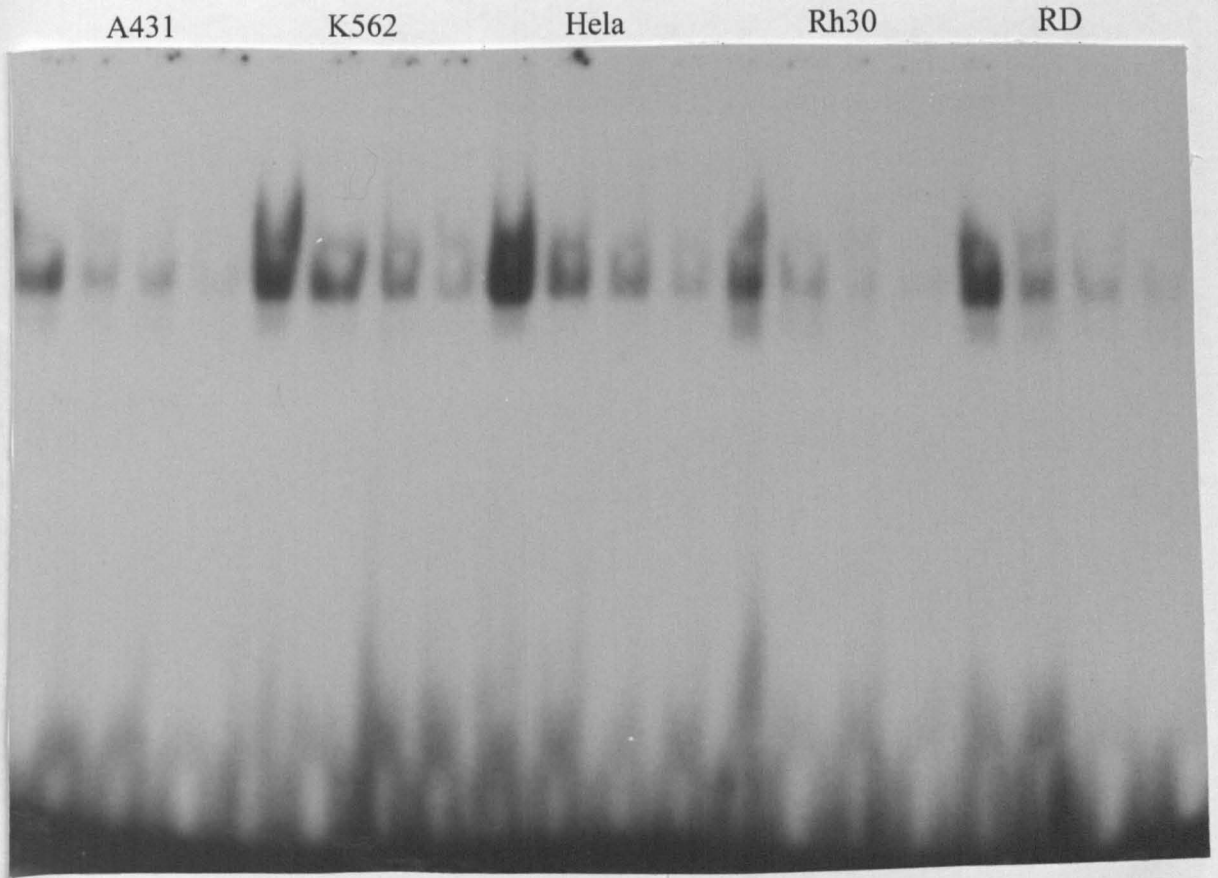


Figure 6.2 Specificity of protein/oligonucleotide interaction in EMSA - specific competitor In the first of each set of four lanes labelled oligonucleotide has been incubated with $7\mu\text{g}$ of whole cell extract from the cell lines A431, K562, HeLa, Rh30 and RD respectively. Lanes 2, 3 and 4 of each set include a 40, 80 and 160 molar excess of the same unlabelled oligonucleotide showing competition for protein. The final lane contains free labelled oligonucleotide only.

electrophoresis
acetic acid
single base
autoradiography

A431

K562

Hela

Rh30

RD

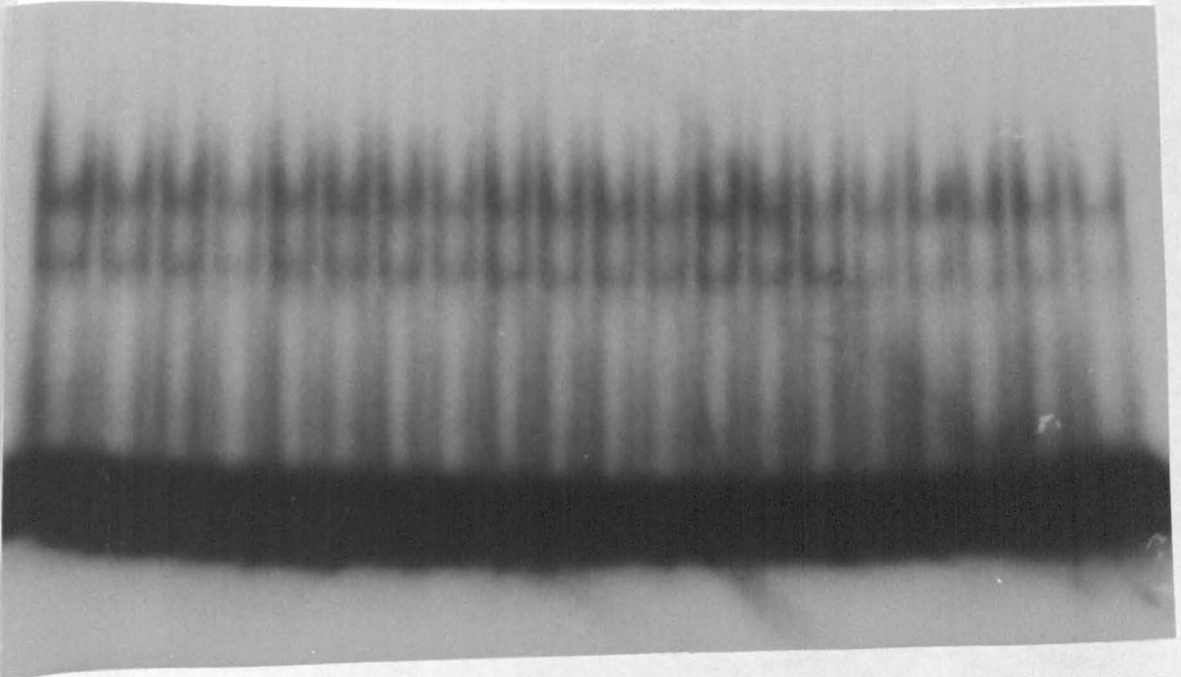


Figure 6.3 Specificity of protein/oligonucleotide interaction in EMSA - non-specific competitor. Addition of non-specific unlabelled oligonucleotide in the same molar excess as in figure 4.2 does not result in competition for protein.

electrophoresis through a denaturing polyacrylamide gel. The gel was fixed with 10% acetic acid/ 10% methanol, dried and autoradiographed. Rather than the expected single band, a range of bands of differing molecular mass appeared on the autoradiograph. Four repetitions of the same experiment resulted in the appearance of a number of bands of varying intensity.

A possible explanation for these observations is that because a high concentration of crude protein extract is present in close proximity to the labelled oligonucleotide, the UV light has caused crosslinking between nonspecifically binding proteins and labelled oligonucleotide. The molecular mass of the protein which binds to the test oligonucleotides specifically could be estimated if it was purified from cell extract. However, although a purification protocol had been developed for Translin protein this is a lengthy and difficult procedure. Therefore, other methods besides molecular mass estimation were used to try to identify the DNA-binding protein

6.4 Effect of oligonucleotide mutation on oligonucleotide-protein complex formation

In a further experiment to determine whether the sequences adjacent to the t(2;13) chromosome translocation breakpoint are Translin binding sites, a series of oligonucleotides were designed, based on the '5'RH28' oligonucleotide, which have an increasing number of base substitutions (underlined) so that they have a decreasing similarity to the Translin consensus binding site sequence;

1. 5'RH28 5'-ATTTCATGAAGGACAATAAGTCA-3'

2. 5'-ATTTCTTGAAGGACAATAAGTCA-3'
3. 5'-ATTTCTAGAAGGACAATAAGTCA-3'
4. 5'-ATTTCTACAAGGACAATAAGTCA-3'
5. 5'-ATTTCTACACGGACAATAAGTCA-3'
6. 5'-ATTTCTACACAGACAATAAGTCA-3'

The oligonucleotide series (starting with the '5'RH28' sequence) has homology to the consensus sequence ranging from 62% to below 25%.

Figure 6.4 below shows a gel mobility shift autoradiogram involving incubations of the above six oligonucleotides with whole cell extracts from the RD cell line. The first two lanes include the '5'RH28' oligonucleotide with and without cell extract. A strong protein-oligonucleotide complex band is present as in the above experiments. The remaining pairs of lanes involving the other oligonucleotides in the series show a significant reduction in protein-oligonucleotide complex formation.

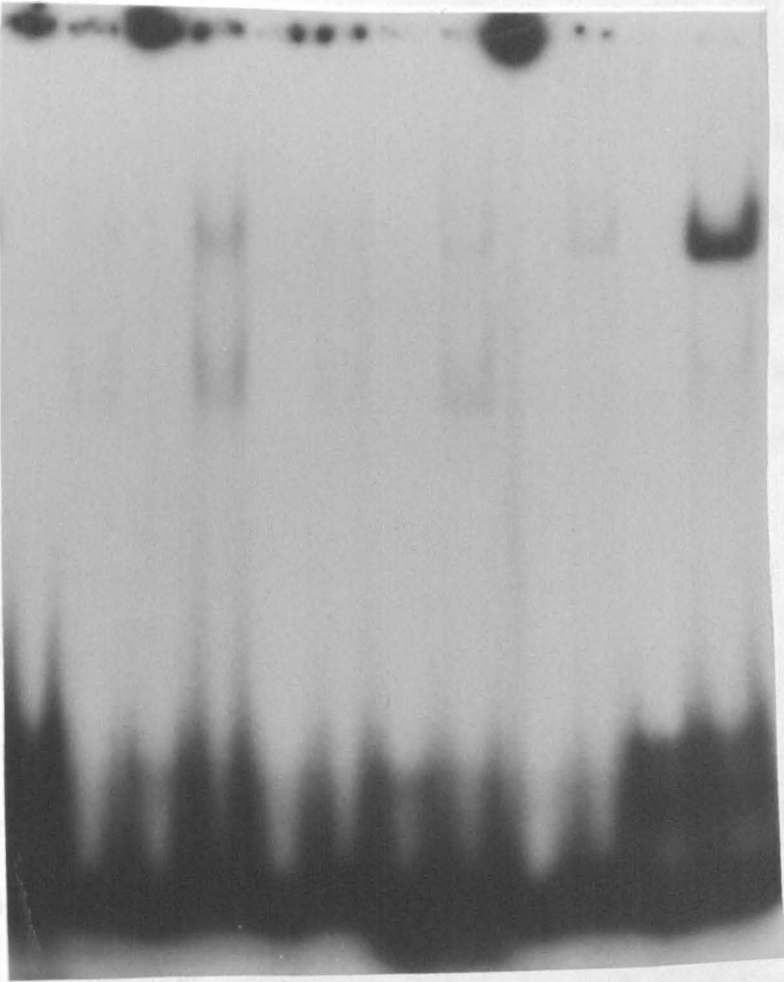
6.5 Effect of anti-translin antibody on oligonucleotide-protein complex formation

The above experiments suggest that a protein or protein complex is present in the rms cell lines tested which can bind to oligonucleotides based on the sequences found next to t(2;13) chromosome translocation breakpoints. To obtain direct evidence that this protein is Translin, an antibody raised against recombinant Translin protein was obtained (a gift from Dr M Kasai). Aoki et al had previously demonstrated that the epitope of the anti-translin antibody is close to the DNA-binding element of translin,

and therefore will block the binding of transcription factors.

in gel mobility shift assay: 5bc 4bc 3bc 2bc 1bc 5'Rh28

- + - + - + - + - + - +



The anti-Tp
RH30 and S
(Figure 6.5)
present. TP
control, a
was used to
presence of

In contrast
binding pro
oligonucleo
chromatogr

Oligonucleotide microarray and anti-TP
in Translin

Figure 6.4 Effect of oligonucleotide base changes on protein/oligonucleotide interaction in EMSA. Binding of protein from RD cell line extract to six different labelled oligonucleotides. Pairs of lanes either do contain extract (+) or do not contain extract (-). The oligonucleotide series, based on the '5'RH28' oligonucleotide, have an increasing number of base substitutions (1bc,2bc,3bc...) which reduce their similarities to the translin consensus recognition sequence.

and therefore will block the binding of recombinant Translin protein to oligonucleotides in gel mobility shift assays.

The anti-Translin antibody was included in experiments involving incubation of the RH30 and K562 protein extracts with '5'RH28' and '3'RH28' labelled oligonucleotides (Figure 6.5). A reduction in band intensity was apparent when the antibody was present. The reduction was greater when the amount of antibody was increased. As a control, a non-specific antibody fraction from the same animal as the translin antibody was used to show that the reduction in band intensity was specifically due to the presence of the translin antibody.

In conclusion, the experiments described in this chapter indicate that there is DNA-binding protein in rms cell lines which can bind specifically to oligonucleotides. The oligonucleotide sequences are based on those found adjacent to several t(2;13) chromosome translocation breakpoints found in alveolar rhabdomyosarcoma. Oligonucleotide mutagenesis and antibody interference studies suggest that this protein is Translin.

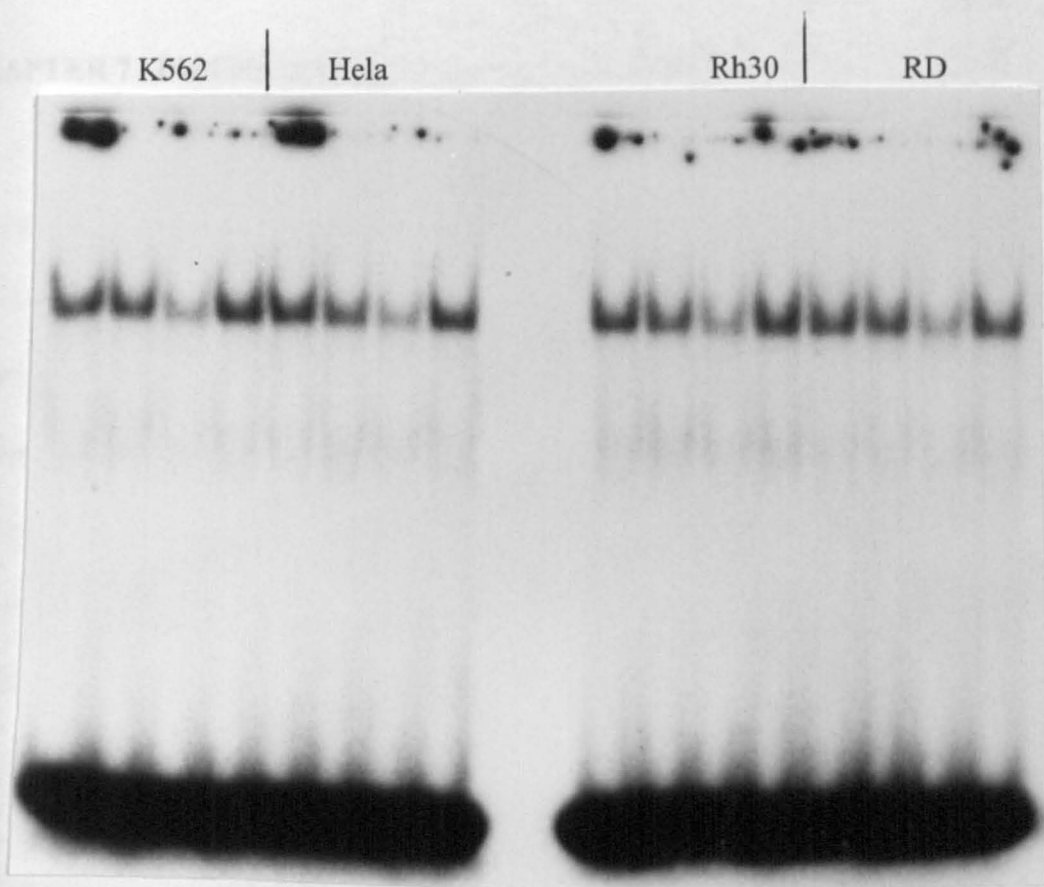


Figure 6.5 Effect of anti-translin antibody on protein/oligonucleotide interaction in EMSA.

In the first of each set of four lanes labelled oligonucleotide has been incubated with 7μg of whole cell extract from the cell lines K562, HeLa, Rh30 and RD respectively. Lanes 2 and 3 of each set include 0.1 and 0.5μg of anti-translin antibody respectively (rabbit IgG, gift from M.Kasai). The epitope of the antibody overlaps the DNA-binding element of translin and thus prevents the interaction between translin and labelled oligonucleotide in the gel shift assay. Reductions in band intensity are apparent in incubations containing translin antibody for all whole cell extracts tested. The fourth lane of each set includes 0.5μg of non-specific rabbit IgG antibody which has no effect on binding.

CHAPTER 7 DISCUSSION

7.1 AIMS OF THIS THESIS

1. Cytogenetic analysis of rhabdomyosarcoma tumours has shown that a translocation t(2;13)(q14;q35) is observed in tumours with alveolar histology. The locus involved on chromosome 2 is the transcription factor PAX3. The translocation t(2;13) is not seen in tumours with embryonal histology, but trisomy of chromosome 2 is a frequent finding. These data suggest that mutations of the PAX3 gene could also have a role in the genesis of the embryonal variant.

The first aim of this thesis was to test this hypothesis by screening tumour DNA from patients with embryonal rhabdomyosarcoma for intragenic mutations within the PAX3 gene. No disease associated mutations could be found in exons 2-8. The findings are discussed in section 7.2.

2. Despite the progress made in understanding the mechanisms which cause chromosome translocations involving juxtaposition of proto-oncogenes next to genes of the immune system, little is known about whether similar mechanisms cause other chromosome translocations involving gene fusion such as that seen in alveolar rhabdomyosarcoma. Thirty out of fifty-six different translocations or inversions described in a review (Rabbitts, 1994) do not involve immunoglobulin or T-cell receptor gene rearrangement.

The second aim of this project was to clone chromosome translocation breakpoints from alveolar rhabdomyosarcoma cell lines which bear the t(2;13) translocation in an attempt to elucidate the mechanism by which they occur.

Chapter 4 describes unsuccessful experiments designed to isolate breakpoint spanning restriction fragments from a cell line by cosmid-cosmid hybridisation. Chapter 5 describes vectorette PCR based experiments which led to the cloning of derivative chromosome 13 breakpoints in three cell lines, and the complete molecular description of a t(2;13) translocation. These findings are discussed in section 7.3 below.

7.2 MUTATION ANALYSIS IN THE PAX3 GENE IN EMBRYONAL AND ALVEOLAR RHABDOMYOSARCOMA

Chapter 3 includes a description of experiments in which the exons of the PAX3 gene were screened for mutations such as point mutations or small deletions in embryonal and alveolar rhabdomyosarcoma tumour samples. Mutation studies of the receptor tyrosine kinase gene RET in patients with medullary thyroid carcinoma (MTC) and multiple endocrine neoplasia type 2A (MEN 2A) have demonstrated point mutations within this gene as a possible cause of transformation (Hofstra et al, 1994;). The RET gene, like the PAX3 gene, is associated with tumour-specific chromosomal rearrangements in papillary thyroid carcinoma and a developmental anomaly, Hirschsprung's disease (Romeo et al, 1994; Ederly et al, 1994; Donis-Keller et al, 1993).

The SSCP method was used because it allows the quick and efficient screening of a large number of samples, and because the same method had been used previously to successfully screen the PAX3 gene for mutations in patients with Waardenburg syndrome. Possible mutations were found in exons 2 and 8, but the same alterations

were seen in a range of control samples, indicating that these sequence variations are polymorphisms rather than disease associated mutations.

The negative findings in chapter 3 do not exclude a possible role for the PAX3 gene in the aetiology of embryonal rhabdomyosarcoma. Position effect mutations or promoter element mutations may be responsible for the activation of the gene but are more difficult to analyse.

7.3 CLONING OF CHROMOSOME TRANSLOCATION BREAKPOINTS

The molecular cloning of the t(2;13) translocation breakpoints that lie in the PAX3 and FKHR genes, occurring on chromosomes 2 and 13 respectively, in three cell lines Rh28, Rh5 and TTC487, has shown that the breakpoints occur within the 16kb intron seven of the PAX3 gene and within the 130kb intron one of the FKHR gene. Although both introns represent large targets for chromosome translocation, the uniform structure of expressed chimeric PAX3/FKHR gene products suggests that there is an absolute requirement for translocations to occur between these two introns (Galili et al, 1993). The breakpoint regions within the PAX3 intron in several other translocation bearing cell lines have been mapped, and there is no indication of clustering of breakpoints within the intron (Barr et al, 1993; personal communication, FG Barr). Following the sequencing of the complete intron, no association between these breakpoint regions and the two ALU repeats and three microsatellite sequences, located within the intron, could be found (Macina et al, 1995).

The vectorette PCR technique was used to clone the translocation breakpoints as this technique permits the cloning of unknown sequences which are adjacent to known sequences. Within the t(2;13) translocation, unknown FKHR intron sequence is adjacent to known PAX3 intron sequence. In the technique, adaptors are ligated to the ends of restriction enzyme digested DNA. The sequence of the cloned region is then determined by PCR using an adaptor primer in combination with primer from the known sequence. The specificity of the PCR is, theoretically, dependent on the known sequence primer only.

There are some disadvantages to using the vectorette PCR method when cloning translocation breakpoints:

1. It is not possible to know if the required PCR product has been amplified until it has been cloned, sequenced and identified. Because the length of the expected PCR product is unknown, amplification products which are to be cloned and sequenced cannot be preselected by size on an agarose gel. Cloning and sequencing can be a lengthy procedure, particularly if the PCR reaction results in a number of amplification products.
2. If the normal chromosome 2 is present as well as the derivative chromosome 13, PCR primer sequences complementary to PAX3 intron sequence will hybridise both to the normal and derivative chromosomes. Amplification products from normal chromosome 2 were cloned and sequenced in several cell lines in this study.

3. During the process of ligating adaptors to the digested genomic DNA, it is possible to generate chimeric clones by ligation of multiple DNA fragments. Therefore, when analysing the sequence of cloned amplification products, it is important to ensure that no restriction sites at which genomic DNA had been originally digested are present. For example, if a vectorette library is constructed with *Sau3A* restriction enzyme and subsequently clones are isolated and sequenced from the library which contain the *Sau3A* restriction site, they are likely to be chimeric.

By using a vectorette PCR approach, the derivative chromosome 13 translocation breakpoints from three cell lines were cloned. The derivative chromosome 2 breakpoint and breakpoint spanning sequences on normal chromosome 13 were also cloned in one of the three cell lines. The complete description of the effects of the chromosome translocation in the cell line indicates that it is a complex process at the molecular level. A duplication of 101bp of sequence has occurred, since it is present in both derivative chromosomes. An insertion of material of undetermined size and unknown origin has occurred in derivative chromosome 2.

The derivative chromosome 13 breakpoint sequences have also been described in two other cell lines. In all three cell lines, binding sites for the protein translin were found flanking the breakpoints. Translin binding sites have also been found flanking translocation breakpoints in a range of lymphomas (Aoki et al, 1995). Chapter 6 describes further studies which suggest that RMS cell lines contain translin protein, and that translin will bind to the newly described breakpoint flanking sequences. The protein binding studies are discussed in section 7.4 below.

7.4 INTERACTION OF PROTEINS WITH TRANSLOCATION BREAKPOINT SEQUENCES

Translin protein was purified to homogeneity using the gel mobility shift assay as a basis for fractionation and the gene coding for Translin was cloned and sequenced (Aoki et al, 1995). It codes for a 228 amino acid protein. A 'leucine zipper' motif, five potential protein kinase C phosphorylation sites, three tyrosine kinase phosphorylation sites and one N-glycosylation site are present in the Translin amino acid sequence (Aoki et al, 1995).

Using gel mobility shift assays they were also able to demonstrate that:

1. Translin will bind to both double-stranded and single-stranded DNA oligonucleotides, but with higher affinity to single-stranded oligonucleotides.
2. Translin will bind specifically to oligonucleotides based on sequences found adjacent to chromosome breakpoints in a range of lymphoid neoplasm's.
3. A consensus translin binding sequence can be deduced from these sequences. Translin will have little or no binding activity to oligonucleotides with sequences unrelated to this consensus.
4. An antibody (raised by Aoki, et al) against Translin protein will interfere with oligonucleotide-protein complex formation.
5. Translin protein is found in both the cytoplasm and nucleus of lymphoid cell lines such as RPMI8226, but only the cytoplasm of non-lymphoid cell lines such as HeLa, A431 and K562.

The experimental data described in chapter 6 demonstrates that translin is present in rhabdomyosarcoma cell lines, and is capable of binding to the sequences which have been found flanking the t(2;13) translocation breakpoints. The electrophoretic mobility shift assay (EMSA) method used as a basis for the experiments has proven simple and versatile, generating several lines of evidence suggesting that it is translin protein rather than another protein which is binding to the flanking sequences:

1. A protein, from cell lines known to contain translin, binds to t(2;13) breakpoint flanking sequences and the consensus translin binding site.
2. A protein from RMS cell lines binds to the same sequences and with the same mobility on non-denaturing polyacrylamide gels.
3. Mutation of conserved bases in the translin binding site abrogates protein binding
4. The crucial evidence involved including an antibody in the EMSA reactions that had been raised against translin protein. Protein/DNA binding was almost completely blocked in the presence of 0.5 μ g of antibody despite the presence of a large quantity of whole cell extract in the reactions (7 μ g) which could have interfered with protein/antibody binding.

Since translin is present in RMS cells, and DNA sequences which bind translin are adjacent to chromosome breakpoints, it is likely that translin has a central role in the generation of the chimeric chromosomes. This suggestion raises several questions about translin:

1. What phenomenon initiates the activation of translin?
2. What is the normal function of translin?
3. How is the normal function of translin controlled?

Aoki et al, have recently isolated and cloned a further protein, TRAX, which interacts with and has extensive homology to, translin. TRAX has a potential leucine zipper DNA-binding motif like translin, but also contains nuclear targeting sequences which suggests a possible role in the active transport of translin into the nucleus.

It is unlikely that evolution has provided cells with a mechanism for producing mutations that ultimately will prove disadvantageous. It is much more likely that translin has a normal role in repair of DNA breakage, hence the chromosome translocations seen in RMS cells most probably represent inappropriate repair of DNA breaks.

Little is yet known about how translin mediates its normal function or about how it binds to its target sequences. One possibility would be that a DNA break is recognised and causes a signal to be passed to the cytoplasm which allows the transport of translin into the nucleus and to the DNA break site. A single molecule binds to the recognition sequence, and a translin multimer then develops with further binding along the DNA sequence with lower sequence specificity. Translin might function as 'scaffolding' around broken DNA ends, so bringing the ends into synapsis and facilitating the activity of other DNA repair enzymes. How might translin be integrated into the overall scheme of DNA break repair?

7.5 DOUBLE STRAND BREAK (DSB) REPAIR AND A MODEL FOR THE GENERATION OF CHROMOSOME TRANSLOCATIONS

There are three categories of DNA damage repair:

1. Excision repair of damaged bases and nucleotides which uses the information on the undamaged strand to repair the damaged one.

2. Mismatch repair which occurs after DNA replication in S phase, which relies on the parental strand to correct a misincorporation in the newly synthesised strand.
3. Double-strand break (dsb) repair, which occurs in two ways. Firstly, homologous recombination, in which the double-strand break (dsb) in one chromosome is repaired using the sequence information on the sister chromosome, and secondly, DNA end joining. In the yeast *Saccharomyces cerevisiae* homologous recombination is the prevalent mode of dsb repair whereas DNA end joining is prevalent in multicellular eukaryotes (Friedberg et al, 1995). A specialised form of DNA rejoining occurs in B- and T-lymphocytes called V(D)J recombination which is a process that normally brings together the 'variable', 'diversity' and 'joining' regions of immunoglobulin genes. The relationship between V(D)J recombination and the generation of chromosome translocations is well established (see chapter 1.9).

The importance of DNA end joining in the rearrangement of immunoglobulin genes is illustrated by mice that are spontaneous severe combined immunodeficiency (*scid*) mutants. These mice lack the ability to rejoin DNA in V(D)J recombination, but also are sensitive to ionizing radiation, indicating a general failure of DNA end joining (Fulop et al, 1990; Hendrikson et al, 1991).

7.5.1 *Proteins involved in dsb repair*

The first protein thought to be involved in dsb repair is termed Ku and was initially identified as an autoantigen from the sera of patients with scleroderma-polymyositis overlap syndrome (Mimori et al, 1981). Further studies have shown that Ku interacts with free ends of double-stranded DNA without any sequence specificity, and serves as

the DNA binding component of a protein named DNA-dependant protein kinase (DNA-PK) which phosphorylates a number of DNA binding proteins in the nucleus (Dvir et al, 1992; Gottlieb and Jackson, 1993). Further details of mechanisms for the repair of dsbs have been provided by the isolation of dsb repair defective mutants from bacteria, yeast and mammalian cells. Dsb repair mutants were selected on the basis of their extreme sensitivity to ionising radiation, such as X-rays, which are known to cause DNA dsbs (Jeggo, 1990). Eleven complementation groups of ionizing-radiation-sensitive mammalian cell mutants have been described, exhibiting a range of sensitivities to X-rays and other DNA-damaging agents (Collins, 1993). Members of four of these complementation groups (IR4-7) are specifically and exquisitely sensitive to dsbs suggesting that their mutant gene products interact in a single pathway of dsb repair. Mapping of the gene for DNA-PK has shown that it corresponds to the position for the murine *scid* defect and for the human gene that complements it (Siple et al, 1995; Miller et al, 1995; Blunt et al, 1995). Both DNA-PK and Ku have now been implicated in V(D)J recombination and dsb repair, DNA-PK and the two subunits of Ku, Ku80 and Ku70, also correspond to three of the ionizing-radiation-sensitive mammalian cell mutants indicated above (Lieber et al, 1997). Little is known about what brings the DNA ends into synapsis in preparation for rejoining following the ionizing radiation induced cleavage of DNA to create a double-strand gap.

An interesting parallel could be drawn from the similarities between V(D)J recombination and dsb repair, and the involvement of DNA binding sites for translin protein in two types of chromosome translocation (Those which appear to involve the V(D)J recombination mechanism and those which do not). Assuming that chromosome

translocations and inversions are most likely to occur because of incorrect repair of double-strand breaks, translin may play a role in the mechanisms that involve DNA-PK and Ku. Because it binds to DNA with sequence specificity and is able to multimerise, translin may perform a structural stabilization role near DNA ends around which the repair process can take place. By including an element of sequence specificity into the repair process it would be necessary to have a very large number of translin binding sites spread throughout the genome to prevent significant loss of genetic material. The translin binding sites described in this thesis found flanking the t(2;13) translocation and those flanking other translocations are only 40-70% homologous to the consensus sequence, suggesting that binding sites could be common. The hypothesis that translin plays an important role in dsb repair and the genesis of chromosome translocations could be tested by inactivating the translin gene in a transgenic mouse model using embryonic stem cell technology.

7.5.2 A model for the formation of chromosome translocations

It has been estimated that 15% of DNA double strand breaks are converted into chromosome breaks (Cornforth and Bedford, 1983). It is likely that a small subset of these chromosome breaks occur in the same cell, at the same time, and in close proximity to each other. A model for the generation of chromosome translocations or inversions maybe as follows:

Two double-strand breaks occur on different chromosomes at the same time and in close proximity to each other. The breaks may be complex including, for example, stretches of single-stranded DNA and base damage which cause the breaks to persist for longer

than usual. The four ends may diffuse apart and when they have been prepared for rejoining by the repair mechanisms outlined above the wrong ends are brought together. Consequently, chromosome translocations or inversions occur which may or may not affect gene loci on the misrepaired chromosomes. There is unlikely to be an adverse effect as a result of such an event in the majority of cases. However, if a chromosome translocation affects an important gene locus in a particular cell type, it may result in the development of neoplasia.

7.6 THE t(2;13) CHROMOSOME TRANSLOCATION AND POSITION EFFECT MUTATIONS

7.6.1 The role of the PAX3/FKHR gene product

The creation of the chimeric PAX3/FKHR gene product is important in the aetiology of rhabdomyosarcoma. Experiments demonstrate that the fusion protein is capable of binding to and transcriptionally activating model PAX3 binding sites and transient cotransfection assays using e5-CAT reporter plasmids indicate that the PAX3-FKHR fusion protein is a more potent transcription activator than PAX3 (see chapter 1.8; Fredericks et al, 1995; Sublett et al, 1995). The principal difference between PAX3 and PAX3-FKHR is the substitution of the carboxyl terminal domain of PAX3 for the putative transcriptional activation domain of FKHR. Experiments have been performed to demonstrate that it is this substitution that alters the transcriptional activation potential of the fusion protein (Bennicelli et al, 1995; Sublett et al, 1995). In similar experiments, evidence suggests that there are elements in the N-terminal region of

PAX3 which can inhibit transactivation of PAX3 but not PAX3-FKHR (Bennicelli et al,1996).

7.6.2 Expression of the PAX3/FKHR gene

Another possible consequence of the t(2;13) chromosome translocation in alveolar rhabdomyosarcoma is the deregulation of expression from the PAX3 promoter because of position effect mutations. These mutations alter gene expression through long-range effects on chromatin structure. Position effect mutations were first described in *Drosophila* as a consequence of chromosome rearrangements in which a gene, normally found in euchromatin, is placed in close proximity to heterochromatin. The expression of genes up to as much as a megabase from the heterochromatin can be affected by such a rearrangement (Henikoff, 1990). A number of germline and somatic position effect mutations have since been described in mouse and man (Bedell et al, 1996). For example, two cases of aniridia, which is caused by mutation of the PAX6 gene, have been described in which the gene has been silenced by a position effect. In one case, a chromosome inversion breakpoint is located 85 to 95kb 3' of the PAX6 gene, and in the other case a chromosome translocation is located upto 185kb 3' of the gene (Fantes et al, 1995). Other examples involve somatic position effect mutations which result in the activation or deregulation of cellular proto-oncogenes and are associated with tumour induction. A t(2;8) variant chromosome translocation in Burkitt's lymphoma results in the juxtaposition of immunoglobulin enhancer elements 140 to 300kb from the myc1 gene causing deregulated expression (Graham and Adams, 1986). Structural alterations of chromosome 3 up to 300kb 3' or 5' of the EVI1 proto-oncogene cause its aberrant

activation in 2% of patients with acute myelogenous leukaemia (Suzukawa et al, 1994; Levy et al, 1994).

Although the PAX3 and FKHR genes are structurally affected by the t(2;13) chromosome translocation in alveolar rhabdomyosarcoma, position effects maybe responsible for the activation of expression of the chimeric gene product. The translocation breakpoints are 80 to 90kb from the 5' end of the PAX3 gene. The FKHR gene on chromosome 13 is normally expressed in adult tissues whereas the PAX3 gene on chromosome 2 is normally only expressed during development. Chromosome 13 structural elements which allow expression of FKHR may be overriding the silencing of expression of PAX3 in adult tissue through position effects.

7.7 FURTHER STUDIES

To substantiate the suggestion that translin sites are found flanking translocation breakpoints it would be interesting to analyse further examples of alveolar RMS and other non-lymphoid tumour specific translocation breakpoints to see if the junction fragments in these tumours also display translin sequence homology. Further examples would indicate whether translin is involved in the generation of translocations in general.

Since translin protein is located in the cytoplasm of non-lymphoid cell types, it may be that its transport into the nucleus is caused by some physiological process. Because translin appears to bind to sequences adjacent to chromosome translocation breakpoints it may also be reasonable to assume that transport of translin into the nucleus is induced

when nuclear DNA has been damaged, and that translin is involved in the repair of such damage. It may be possible to exploit the cytoplasmic location of translin in experiments in which cells are exposed to DNA damaging agents. Measurement of the levels of translin protein in the cytoplasm and the nucleus could demonstrate transport of translin into the nucleus following such exposure . A DNA double strand break is the most likely cause of chromosome translocation so that studies using DNA damaging agents such as X-rays which cause this kind of damage should be informative.

REFERENCES

- Adams B, Dorfler P, Aguzzi A, et al: Pax-5 encodes the transcription factor BSAP and is expressed in B lymphocytes, the developing CNS, and adult testis. *Genes Dev* 1992;6:1589-1607.
- Alt FW, Yancopoulos GD, Blackwell TK, et al: Ordered rearrangement of immunoglobulin heavy-chain variable region segments. *Embo j* 1984;3:1209-1219.
- Amati B, Brooks MW, Levy N, Littlewood TD, Evan GI, Land H: Oncogenic activity of the c-myc protein requires dimerisation with max. *Cell* 1993;72:233-245.
- Ang SL, Wierda A, Wong D, et al: The formation and maintenance of the definitive endoderm lineage in the mouse: involvement of HNF3/forkhead proteins. *Development* 1993;119:1301-1315.
- Aoki K, Suzuki K, Sugano T, et al: A novel gene, Translin, Encodes a recombination hotspot binding protein associated with chromosomal translocations. *Nature Genetics* 1995;10:167-174.
- Aoki K, Ishida R, Kasai M: Isolation and characterisation of a cDNA encoding a translin-like protein, TRAX. *FEBS letters* 1997;401:109-112.
- Asano M, Gruss P: Pax-5 is expressed at the midbrain-hindbrain boundary during mouse development. *Mech Dev* 1992;39:29-39.
- Baer R, Davis M, Forster A, Rabbitts PH, Malcolm S, Rabbitts TH: c-myc gene activation and chromosomal translocation. *J Cell Science* 1984;S1:95-101.
- Baer R, Forster A, Lavenir I, Rabbitts TH: Immunoglobulin V_H genes are transcribed by T-cells in association with a new 5'exon. *J Exp Med* 1988;167:2011-2016.
- Baer R, Heppell A, Taylor AMR, Rabbitts PH, Boullier B, Rabbitts TH: The breakpoint of an inversion of chromosome 14 in a T-cell leukemia - sequences downstream of the immunoglobulin heavy chain locus are implicated in tumorigenesis. *PNAS* 1988;84:9069-9073.
- Bailly RA, Bosselut R, Zucman J, et al: DNA-binding and transcriptional activation properties of the EWS-Fli1 fusion protein resulting from the t(11;22). *Mol Cell Biol* 1994;14:3230-3241.
- Baldwin CT, Hoth CF, Amos JA, da Silva EO, Milunsky A: An exonic mutation in the HuP2 paired domain gene causes Waardenburg's syndrome [see comments]. *Nature* 1992;355:637-638.

Barr FG, Biegel JA, Sellinger B, Womer RB, Emanuel BS: Molecular and cytogenetic analysis of chromosomal arms 2q and 13q in alveolar rhabdomyosarcoma. *Genes Chromosom Cancer* 1991;3:153-161.

Barr FG, Holick J, Nycum L, Biegel JA, Emanuel BS: Localization of the t(2;13) breakpoint of alveolar rhabdomyosarcoma on a physical map of chromosome 2. *Genomics* 1992;13:1150-1156.

Barr FG, Nauta LE, Davis RJ, Schafer BW, Nycum LM, Beigel JA: In vivo amplification of the PAX3-FKHR and PAX7-FKHR fusion genes in alveolar rhabdomyosarcoma. *Hum Mol Genet* 1996;5:15-21.

Barr FG, Sellinger B, Emanuel BS: Localization of the rhabdomyosarcoma t(2;13) breakpoint on a physical map of chromosome 13. *Genomics* 1991;11:941-947.

Barr FG, X, Holick J, Biegel JA, Rovera G, Emanuel BS: Rearrangement of the PAX3 paired box gene in the paediatric solid tumour alveolar rhabdomyosarcoma. *Nat Genet* 1993;3:113-117.

Bedell MA, Jenkins NA, Copeland NG: Good genes in bad neighbourhoods. *Nature genet* 1996;12:229-232.

Bennicelli JL, Edwards RH, Barr FG: Mechanism for transcriptional gain of function resulting from chromosomal translocation in alveolar rhabdomyosarcoma. *Proc Natl Acad Sci USA* 1996;93:5455-5459.

Bennicelli JL, Fredericks WJ, Wilson RB, Rauscher FJ, Barr FG: Wild type PAX3 protein and the PAX3-FKHR fusion protein of alveolar rhabdomyosarcoma contain potent, structurally distinct transcriptional activation domains. *Oncogene* 1995;11:119-130.

Biegel JA, Meek RS, Parmiter AH, Conard K, Emanuel BS: Chromosomal translocation t(1;13)(p36;q14) in a case of rhabdomyosarcoma. *Genes Chromosom Cancer* 1991;3:483-484.

Biegel JA, Nycum LM, Valentine V, Barr FG, Shapiro DN: Detection of the t(2;13)(q35;q14) and PAX3-FKHR fusion in alveolar rhabdomyosarcoma by fluorescence in situ hybridization. *Genes Chromosom Cancer* 1995;12:186-192.

Biegel JA, Nycum LM, Valentine V, Barr FG, Shapiro DN: Detection of the t(2;13)(q35;q14) and PAX3-FKHR fusion in alveolar rhabdomyosarcoma by fluorescence in situ hybridisation. *Genes, Chromosom Cancer* 1995;12:186-192.

Bishop JM: Enemies within: the genesis of retroviral oncogenes. *Cell* 1981;23:5-6.

- Bishop JM: The molecular genetics of cancer. *Science* 1987;235:305-311.
- Blackwood EM, Eisenman RN: Max-A helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with myc. *Science* 1991;251:1211-1217.
- Blunt T, Finnie NJ, Taccioli GE, et al: Defective DNA-dependent protein kinase activity is linked to V(D)J recombination and DNA repair defects associated with the murine scid mutation. *Cell* 1995;80:813-823.
- Bodmer W, Bishop T, Karran P: Genetic steps in colorectal cancer. *Nature genet* 1994;6:217-219.
- Boehm T, Mengle-gaw L, Kees UR, et al: Alternating purine-pyrimidine tracts may promote chromosomal translocations seen in a variety of human lymphoid tumours. *EMBO J* 1989;8:2621-2631.
- Bopp D, Burri M, Baumgartner S, Frigerio G, Noll M: Conservation of a large protein domain in the segment ation gene *paired* and in functionally related genes off *Drosophila*. *Cell* 1986;47:1033-1040.
- Burri M, Tromvoukis Y, Bopp D, Frigerio G, Noll M: Conservation of the paired domain in metazoans and its structure in three isolated human genes. *EMBO J* 1989;8:1183-1190.
- Busslinger M, Klix N, Pfeffer P, Graninger PG, Kozmik Z: Deregulation of PAX5 by translocation of the Eu enhancer of the IgH locus adjacent to two alternative PAX5 promoters in a diffuse large-cell lymphoma. *Proc Natl Acad Sci USA* 1996;93:6129-6134.
- Butt J, Greenberg J, Winship I, Sellars S, Beighton P, Ramesar R: A splice junction mutation in PAX3 causes Waardenburg syndrome in a South African family. *Hum Mol Genet* 1994;3:197-198.
- Carthew RW, Chodosh LA, Sharp PA: An RNA polymerase II transcription factor binds to an upstream element in the adenovirus major late promotor. *Cell* 1985;43:439-448.
- Chalepakis G, Fritsch R, Fickenscher H, Deutsch U, Goulding M, Gruss P: The molecular basis of the undulated/Pax-1 mutation. *Cell* 1991;66:873-884.
- Chalepakis G, Goulding M, Read A, Strachan T, Gruss P: Molecular basis of splotch and Waardenburg Pax-3 mutations. *Proc Natl Acad Sci U S A* 1994;91:3685-3689.
- Chalepakis G, Jones FS, Edelman GM, Gruss P: Pax-3 contains domains for transcription activation and transcription inhibition. *Proc Natl Acad Sci U S A* 1994;91:12745-12749.

- Chalepakis G, Wijnholds J, Gruss P: Pax-3-DNA interaction: flexibility in the DNA binding and induction of DNA conformational changes by paired domains. *Nucleic Acids Res* 1994;22:3131-3137.
- Clark KL, Halay ED, Lai E, Burley SK: Co-crystal structure of the HNF-3/forkhead DNA-recognition motif resembles histone H5. *Nature* 1993;364:412-420.
- Cleary ML, Sklar J: Nucleotide sequence of a t(14;18) chromosomal breakpoint in follicular lymphoma and demonstration of a breakpoint-cluster region near a transcriptionally active locus on chromosome 18. *Proc Natl Acad Sci USA* 1985;82:7439-7443.
- Clevidence DE, Overdier DG, Tao W, et al: Identification of nine tissue-specific transcription factors of the hepatocyte nuclear factor 3/forkhead DNA-binding-domain family. *Proc Natl Acad Sci U S A* 1993;90:3948-3952.
- Collins AR: Mutant rodent cell lines sensitive to ultraviolet light, ionizing radiation and cross-linking agents: a comprehensive survey of genetic and biochemical characteristics. *Mutation Res* 1993;293:99-118.
- Cornforth MN, Bedford JS: X-ray induced breakage and rejoining of human interphase chromosomes. *Science* 1983;222:1141-1143.
- Costa RH, Grayson DR: Site-directed mutagenesis of hepatocyte nuclear factor (HNF) binding sites in the mouse transthyretin (TTR) promoter reveal synergistic interactions with its enhancer region. *Nucleic Acids Res* 1991;19:4139-4145.
- Cotton RGH, Rodrigues NR, Campbell RD. Reactivity of cytosine and thymine in single-base-mismatches with hydroxyamine and osmium tetroxide and its applications to the study of mutations. *Proc.Natl.Acad.Sci* 1988;85:4397-4401.
- Cowell JK, Mitchell CD: A somatic cell hybrid mapping panel for regional assignment of human chromosome 13 DNA sequences. *Cytogenet Cell Genet* 1989;52:1-6.
- Crist WM, Garnsey L, Beltangady MS, et al: Prognosis in children with rhabdomyosarcoma:a report of the intergroup Rhabdomyosarcoma studies I and II. *J Clin Oncol* 1990;8:443-452.
- Czerny T, Schaffner G, Busslinger M: DNA sequence recognition by Pax proteins: bipartite structure of the paired domain and its binding site. *Genes Dev* 1993;7:2048-2061.
- Davis RJ, Bennicelli JL, Macina RA, Nycum LM, Biegel JA, Barr FG: Structural characterisation of the FKHR gene and its rearrangement in alveolar rhabdomyosarcoma. *human molecular genetics* 1995;4:2355-2362.

Davis RJ, D'Cruz CM, Lovell MA, Biegel JA, Barr FG: Fusion of PAX7 to FKHR by the variant t(1;13)(p36;q14) translocation in alveolar rhabdomyosarcoma. *Cancer Res* 1994;54:2869-2872.

Davis RL, Weintraub H, Lassar AB: Expression of a single transfectant cDNA converts fibroblasts to myoblasts. *Cell* 1987;51:987-1000.

Delattre O, Zucman J, Plougastel B, et al: Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours. *Nature* 1992;359:162-165.

Denny CT, Yoshikai Y, Mak TW, Smith SD, Hollis GF, Kirsch IR: A chromosome 14 inversion in a T-cell lymphoma is caused by site specific recombination between immunoglobulin and T-cell receptor loci. *Nature* 1986;320:549-551.

Desiderio SV, Yancopoulos GD, Paskind M, et al: Insertion of N regions into heavy-chain genes is correlated with expression of terminal deoxytransferase in B cells. *Nature* 1984;311:752-755.

Dias P, Kumar P, Marsden HB, Gattamaneni HR, Heighway J, Kumar S: N-myc gene is amplified in alveolar rhabdomyosarcomas (RMS) but not in embryonal RMS. *Int J Cancer* 1990;45:593-596.

Donis-Keller H, Dou S, Chi D, Carlson KM, Toshima K, Lairmore TC, Wells SA. Mutations in the RET protooncogene are associated with MEN 2A and FMTC. *Hum. Mol. Genet.* 1993;2:851-856.

Douglass EC, Rowe ST, Valentine M, et al: Variant translocations of chromosome 13 in alveolar rhabdomyosarcoma. *Genes Chromosom Cancer* 1991;3:480-482.

Douglass EC, Valentine M, Etcubanas E, et al: A specific chromosomal abnormality in rhabdomyosarcoma [published erratum appears in *Cytogenet Cell Genet* 1988;47(4): following 232]. *Cytogenet Cell Genet* 1987;45:148-155.

Dressler GR, Wilkinson JE, Rothenpieler UW, Patterson LT, Williams Simons L, Westphal H: Deregulation of Pax-2 expression in transgenic mice generates severe kidney abnormalities. *Nature* 1993;362:65-67.

Dressler GR, Wilkinson JE, Rothenpieler UW, Patterson LT, Williams L, Westphal H: Deregulation of Pax2 expression in transgenic mice generates severe kidney abnormalities. *Nature* 1993;362:65-67.

Dvir A, Peterson SR, Knuth MW, Lu H, Dynan WS: Ku autoantigen is the regulatory component of a template-associated protein kinase that phosphorylates RNA polymerase II. *Proc Natl Acad Sci* 1992;89:11920-11924.

Edery P, Lyonnet S, Mulligan LS, Ponder BAJ, Munnich A. Mutations of the RET proto-oncogene in Hirschsprung's disease. *Nature* 1994;367:378-380.

Epstein DJ, Vekemans M, Gros P: Splotch (Sp2H), a mutation affecting development of the mouse neural tube, shows a deletion within the paired homeodomain of Pax-3. *Cell* 1991;67:767-774.

Epstein DJ, Vogan KJ, Trasler DG, Gros P: A mutation within intron 3 of the Pax-3 gene produces aberrantly spliced mRNA transcripts in the splotch (Sp) mouse mutant. *Proc Natl Acad Sci USA* 1993;90:532-536.

Epstein JA, Lam P, Jepeal L, Maas RL, Shapiro DN: Pax3 inhibits myogenic differentiation of cultured myoblast cells. *J Biol Chem* 1995;270:11719-11722.

Fantes J, Redeker B, Breen M, et al: Aniridia-associated cytogenetic rearrangements suggest that a position effect may cause the mutant phenotype. *Hum mol genet* 1995;4:415-422.

Fields S, Song OK: A novel genetic system to detect protein protein interactions. *Nature* 1989;340:245-246.

Fredericks WJ, Galili N, Mukhopadhyay S, et al: The PAX3-FKHR fusion protein created by the t(2;13) translocation in alveolar rhabdomyosarcomas is a more potent transcriptional activator than PAX3. *Mol Cell Biol* 1995;15:1522-1535.

Friedberg EC, Walker GC, Siede W: *DNA repair and mutagenesis*, Washington DC, ASM press; 1995:

Friend SH, Bernards R, Rogelj S, et al: A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature* 1986;323:643-646.

Fulop GM, Phillips RA: the scid mutation in mice causes a general defect in DNA repair. *Nature* 1990;347:479-482.

Galas D, Schmitz A: Dnase footprinting: a simple method for the detection of protein-DNA binding specificity. *Nucleic Acids res* 1978;5:3157-3170.

Galili N, Davis RJ, Fredericks WJ, et al: Fusion of a fork head domain gene to PAX3 in the solid tumour alveolar rhabdomyosarcoma [published erratum appears in *Nat Genet* 1994 Feb;6(2):214]. *Nature Genetics* 1993;5:230-235.

- Glaser T, Walton DS, Maas RL: Genomic structure, evolutionary conservation and aniridia mutations in the human PAX6 gene. *Nat Genet* 1992;2:232-239.
- Glavac D, Dean M. Optimisation of the SSCP technique for detection of point mutations *Hum. Mutation* 1993;2:404-414.
- Glaser T, Walton DS, Mass RL: Genomic structure, evolutionary conservation and aniridia mutations in the human PAX6 gene. *Nature genet* 1992;2:232-239.
- Gottlieb TM, Jackson SP: The DNA-dependent protein kinase: requirement of DNA ends and association with Ku antigen. *Cell* 1993;72:131-142.
- Goulding M, Sterrer S, Fleming J, et al: Analysis of the Pax-3 gene in the mouse mutant splotch. *Genomics* 1993;17:355-363.
- Goulding MD, Chalepakis G, Deutsch U, Erselius JR, Gruss P: Pax-3, a novel murine DNA binding protein expressed during early neurogenesis. *EMBO J* 1991;10:1135-1147.
- Goulding MD, Lumsden A, Gruss P: Signals from the notochord and floor plate regulate the region-specific expression of two pax genes in the developing spinal cord. *Development* 1993;117:1001-1016.
- Graham M, Adams JM: Chromosome 8 breakpoint far 3' of the c-myc oncogene in a Burkitts lymphoma 2;8 variant translocation is equivalent to a murine pvt-1 locus. *EMBO J* 1986;5:2845-2849.
- Grieco M, Santoro M, Fusco A, Vecchio G. PTC is a novel rearranged form of the ret proto-oncogene and is frequently detected in vivo in human thyroid papillary carcinomas. *Cell* 1990;60:557-563.
- Gruss P, Walther C: Pax in development. *Cell* 1992;69:719-722.
- Han K, Manley JL: Transcriptional repression by the Drosophila even-skipped protein: definition of a minimal repression domain. *Genes dev* 1993;7:491-503.
- Hanson IM: *Hum Mol Genet* 1993;2:915-917.
- Hanson IM, Fletcher JM, Jordan T, et al: Mutations at the PAX6 locus are found in heterogeneous anterior segment malformations including Peters' anomaly. *Nat Genet* 1994;6:168-173.
- Hendrikson EA, Qin S, Bump EA, Schatz DG, Oettinger M, Weaver DT: A link between double-strand break repair and V(D)J recombination: the scid mutation. *Proc Natl Acad Sci* 1991;88:4061-4065.
- Henikoff s: Position-effect variegation after 60 years. *tigs* 1990;6:422-426.

Henry I, Bonaiti-Pellie C, Chehensse V, et al: Uniparental parental isodisomy in a genetic cancer predisposing syndrome. *Nature* 1991;351:665-667.

Henry I, van Heyningen V, Puech A, et al: Reassessment of breakpoints in chromosome 11p15. *Cytogenet Cell Genet* 1993;62:52-53.

Hodgson SV, Saunders KE: A probable case of the homozygous condition of the aniridia gene. *J clin inv* 1980;47:478-481.

Hofstra RMW, Landsvater RM, Ceccherini I, Stulp RP, Lips CJM, Buys CHCM. A mutation in the RET protooncogene associated with multiple endocrine neoplasia type 2B and sporadic medullary thyroid carcinoma. *Nature* 1994;367:375-377.

Horn RC, Enterline HT: Rhabdomyosarcoma: a clinicopathological study; identification and classification. *cancer* 1958;11:181-199.

Hoth CF, Milunsky A, Lipsky N, Sheffer R, Clarren SK, Baldwin CT: Mutations in the paired domain of the human PAX3 gene cause Klein-Waardenburg syndrome (WS-III) as well as Waardenburg syndrome type I (WS-I). *Am J Hum Genet* 1993;52:455-462.

Hromas R, Costa R: The hepatocyte nuclear factor-3/forkhead transcription regulatory family in development, inflammation and neoplasia. *Crit Rev Oncol hematol* 1995;20:129-140.

Jeggo PA, Taccioli GE, Jackson SP: Menage a trois: double strand break repair, V(D)J recombination and DNA-PK. *Bioessays* 1995;17:949-957.

Jordan T, Hanson I, Zaletayev D, et al: The human PAX6 gene is mutated in two patients with aniridia. *Nat Genet* 1992;1:328-332.

Kessel K, Gruss P: Murine developmental control genes. *Science* 1990;249:374-379.

Kleinfield R, Hardy RR, Tarlinton D, Dangl J, Herzenberg LA, Weigert M: Recombination between an expressed immunoglobulin heavy-chain gene and a germline variable gene segment in a Ly1+ B-cell lymphoma. *Nature* 1986;322:843-846.

Knudsen AG: Mutation and human cancer. *Adv cancer res* 1973;17:317-352.

Koi M, Johnson LA, Kalikin LM, Little PF, Nakamura Y, Feinberg AP: Tumor cell growth arrest caused by subchromosomal transferable DNA fragments from chromosome 11. *Science* 1993;260:361-364.

- Koufos A, Hansen MF, Copeland NG, Jenkins NA, Lampkin BC, Cavenee WK: Loss of heterozygosity in three embryonal tumours suggests a common pathogenetic mechanism. *Nature* 1985;316:330-334.
- Kozmik Z, Sure U, Ruedi D, Busslinger M, Aguzzi A: Deregulated expression of PAX5 in medulloblastoma. *Proc Natl Acad Sci U S A* 1995;92:5709-5713.
- Lai E, Prezioso VR, Smith E, Litvin O, Costa RH, Darnell JE: HNF-3A, a hepatocyte-enriched transcription factor of novel structure is regulated transcriptionally. *Genes dev* 1990;4:1427-1436.
- Lai E, Prezioso VR, Tao W, Chen WS, Darnell JE: Hepatocyte nuclear factor 3a belongs to a gene family in mammals that is homologous to the drosophila homeotic gene *forkhead*. *Genes dev* 1991;5:416-427.
- Lessnick SL, Braun BS, Denny CT, May WA: Multiple domains mediate transformation by the Ewing's sarcoma EWS/FLI-1 fusion gene. *Oncogene* 1995;10:423-431.
- Levy ER, Parganas E, Morishita K, et al: DNA rearrangements proximal to the EVI1 locus associated with the 3q21q26 syndrome. *Blood* 1994;83:1348-1354.
- Li L, Olson EN: Regulation of muscle cell growth and differentiation by the MyoD family of helix-loop-helix proteins. *Adv cancer res* 1992;58:95-119.
- Lieber MR, Grawunder U, Wu X, Yaneva M: Tying loose ends: roles of Ku and DNA-dependent protein kinase in the repair of double-strand breaks. *Current opinions in genetics and development* 1997;7:99-104.
- Loh WE, Jr., Scoble HJ, Livanos E, et al: Human chromosome 11 contains two different growth suppressor genes for embryonal rhabdomyosarcoma. *Proc Natl Acad Sci U S A* 1992;89:1755-1759.
- Lu M, Zhang N, Raimondi S, Ho AD: S1 nuclease hypersensitive sites in an oligopurine/oligopyrimidine DNA from the t(10;14) breakpoint cluster region. *Nucleic acids res* 1991;20:263-266
- Macina RA, Barr FG, X, Riethman HC: Genomic organization of the human PAX3 gene: DNA sequence analysis of the region disrupted in alveolar rhabdomyosarcoma. *Genomics* 1995;26:1-8.
- Mao X, Miesfeldt S, Yang H, Leiden JM, Thompson CB: The FLI-1 and chimeric EWS-FLI-1 oncoproteins display similar DNA binding specificities. *J Biol Chem* 1994;269:18216-18222.
- Maulbecker CC, Gruss P: The oncogenic potential of Pax genes. *EMBO J* 1993;12:2361-2367.

May WA, Gishizky ML, Lessnick SL, et al: Ewing sarcoma 11;22 translocation produces a chimeric transcription factor that requires the DNA-binding domain encoded by FLI1 for transformation. *Proc Natl Acad Sci U S A* 1993;90:5752-5756.

May WA, Lessnick SL, Braun BS, et al: The Ewing's sarcoma EWS/FLI-1 fusion gene encodes a more potent transcriptional activator and is a more powerful transforming gene than FLI-1. *Mol Cell Biol* 1993;13:7393-7398.

Miller RD, Hogg J, Gell J, Jackson SP, Riblet R: Gene for the catalytic subunit of mouse DNA-PK maps to the scid locus. *Proc Natl Acad Sci* 1995;92:10792-10795.

Mimori T, Akizuki M, Yamagata H, Inada S, Yoshida S, Homma M: Characterisation of a high molecular weight acidic nuclear protein recognised by autoantibodies in sera from patients with polymyositis-scleroderma overlap syndrome. *J Clin Invest* 1981;68:611-620.

Mitchell CD, Ventris JA, Warr TJ, Cowell JK: Molecular definition in a somatic cell hybrid of a specific 2:13 translocation breakpoint in childhood rhabdomyosarcoma. *Oncogene* 1991;6:89-92.

Mitchell PJ, Tjian R: Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* 1989;245:371-378.

Monaghan AP, Kaestner KH, Grau E, Schutz G: Postimplantation expression patterns indicate a role for the mouse forkhead/HNF-3 alpha, beta and gamma genes in determination of the definitive endoderm, chordamesoderm and neuroectoderm. *Development* 1993;119:567-578.

Morell R, Friedman TB, Moeljopawiro S, Hartono, Soewito, Asher JH, Jr. A frameshift mutation in the HuP2 paired domain of the probable human homolog of murine Pax-3 is responsible for Waardenburg syndrome type 1 in an Indonesian family. *Hum Mol Genet* 1992;1:243-247.

Mulligan LM, Kwok JBJ, Healey CS, Ponder BAJ. Germline mutations of the RET proto-oncogene in multiple endocrine neoplasia type 2A. *Nature* 1993;363:458-460.

Newsham I, Claussen U, Ludecke HJ, et al: Microdissection of chromosome band 11p15.5: characterisation of probes mapping distal to the HBBC locus. *genes chrom cancer* 1991;3:108-116.

Niggli FK, Powell JE, Parkes SE, et al: DNA ploidy and proliferative activity(S-phase) in childhood soft tissue sarcomas;their value as prognostic indicators. *Br J Cancer* 1994;69:1106-1110.

Noll M: Evolution and role of Pax genes. *Curr Opin Genet Dev* 1993;3:595-605.

Nowell PC, Hungerford DA. A minute chromosome in human chronic granulocytic leukaemia. *Science* 1960;132:1497.

Ohno T, Ouchida M, Lee L, Gatalica Z, Rao VN, Reddy ES: The EWS gene, involved in Ewing family of tumors, malignant melanoma of soft parts and desmoplastic small round cell tumors, codes for an RNA binding protein with novel regulatory domains. *Oncogene* 1994;9:3087-3097.

Ohno T, Rao VN, Reddy ES: EWS/Fli-1 chimeric protein is a transcriptional activator. *Cancer Res* 1993;53:5859-5863.

Olegard C, Mandahl N, Heim S, Willen H, Leifsson B, Mitelman F: Embryonal rhabdomyosarcoma with 100 chromosomes but no structural aberrations. *Cancer Genet Cytogenet* 1992;60:198-201.

Olson EN: MyoD family: a paradigm for development? *Genes devel* 1990;4:1454-1461.

Olson EN, Klien WH: bHLH factors in muscle development: deadlines and commitments, what to leave in and what to leave out. *Genes devel* 1994;8:1-8.

Orita M, Suzuki Y, Sekiya K, Hayashi K: Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics* 1989;5:874-879.

Padgett RA, Grabowski P, Konarska M, Sieler S, Sharp P: Splicing of messenger RNA precursors. *Ann Rev Biochem* 1986;55:1119-1150.

Pandya A, Xia XJ, Lnada BL, et al: Phenotypic variation in Waardenburg syndrome-mutational heterogeneity, modifier genes or polygenic background. *Hum Mol Genet* 1996;5:497-502.

Pani L, Overdier DG, Porcella A, Qian X, Lai E, Costa RH: Hepatocyte nuclear factor 3 beta contains two transcriptional activation domains, one of which is novel and conserved with the Drosophila fork head protein. *Mol Cell Biol* 1992;12:3723-3732.

Pappo AS, Crist WM, Kuttesch J, et al: Tumor-cell DNA content predicts outcome in children and adolescents with clinical group III embryonal rhabdomyosarcoma. The Intergroup Rhabdomyosarcoma Study Committee of the Children's Cancer Group and the Pediatric Oncology Group [published erratum appears in *J Clin Oncol* 1994 Feb;12(2):440]. *J Clin Oncol* 1993;11:1901-1905.

- Parry P, Wei Y, Evans G: Cloning and characterization of the t(X;11) breakpoint from a leukemic cell line identify a new member of the forkhead gene family. *Genes Chromosom Cancer* 1994;11:79-84.
- Pasteris NG, Trask BJ, Sheldon S, Gorski JL: Discordant phenotype of two overlapping deletions involving the PAX3 gene in chromosome 2q35. *Hum Mol Genet* 1993;2:953-959.
- Potluri VR, Gilbert F: A cytogenetic study of embryonal rhabdomyosarcoma. *Cancer genet cytogenet* 1985;14:169-173.
- Prendergast GC, Lawe D, Ziff E: *Cell* 1991;65:395-407.
- Prendergast GC, Lawe D, Ziff EB: Association of Myn, the murine homolog of max, with c-myc stimulates methylation sensitive DNA binding and ras cotransformation. *Cell* 1991;65:395-407.
- Rabbitts TH: Chromosomal translocations in human cancer. *Nature* 1994;372:143-149.
- Rabbitts TH, Boehm T, Menglegaw L: Chromosomal abnormalities in lymphoid tumours - mechanism and role in tumour pathogenesis. *TIGS* 1988;11:300-304.
- Rainier S, Johnson LA, Dobry CJ, Ping AJ, Grundy PE, Feinberg AP: Relaxation of imprinted genes in human cancer. *Nature* 1993;362:747-749.
- Raney BR, Hays DM, Tefft M, Triche TJ: Rhabdomyosarcoma and the undifferentiated sarcomas, in Pizzo PA, Poplack DG (eds): *Principles and practice of paediatric oncology*. Philadelphia, J.B.Lippincott; 1989:635-658.
- Ren R, Mayer BJ, Cicchetti P, Baltimore D: Identification of a ten-amino acid proline-rich SH3 binding site. *Science* 1993;259:1157-1161.
- Riley J, Butler R, Ogilvie D, et al: A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. *Nucleic acids research* 1990;18:2887-2890.
- Romeo G, Ronchetto P, Luo Y, Barone V, Martucciello G. Point mutations affecting the tyrosine kinase domain of the RET proto-oncogene in Hirschsprung's disease. *Nature* 1994;367:377-378.
- Rowley JD. A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and giemsa staining. *Nature* 1973;243:290-293.

- Sakai T, Toguchida J, Ohtani N, Yandell DW, Rapaport JM, Dryja TP: Allele-specific hypermethylation of the retinoblastoma tumor-suppressor gene. *Am J Hum Genet* 1991;48:880-888.
- Sanyanusin P, Schimmenti LA, McNoe LA, et al: Mutation of the PAX2 gene in a family with optic nerve colobomas, renal anomalies and vesicoureteral reflux. *Nat genet* 1995;9:358-363.
- Sasaki H, Hogan BLM: Differential expression of multiple forkhead related genes during gastrulation and axial pattern formation in the mouse embryo. *Development* 1993;118:47-59.
- Sawyers CL, Denny CT, Witte ON: Leukemia and the disruption of normal hematopoiesis. *Cell* 1991;64:337-350.
- Schedl A, Ross A, Lee M, et al: Influence of PAX6 gene dosage on development: overexpression causes severe eye abnormalities. *Cell* 1996;86:71-82.
- Scheidler S, Fredericks WJ, Rauscher FJ, Barr FG, Vogt PK: The hybrid PAX3-FKHR fusion protein of alveolar rhabdomyosarcoma transforms fibroblasts in culture. *Proc Natl Acad Sci* 1996;93:9805-9809.
- Schurr E: Mapping of Col3a1 and Col6a3 to proximal murine chromosome 1 identifies conserved image of structural protein genes between murine chromosome 1 and human chromosome 2q. *Genomics* 1990;8:477-486.
- Scrabble H, Cavenee W, Ghavimi F, Lovell M, Morgan K, Sapienza C: A model for embryonal rhabdomyosarcoma tumorigenesis that involves genome imprinting. *Proc Natl Acad Sci USA* 1989;86:7480-7484.
- Scrabble H, Witte D, Shimada H, et al: Molecular differential pathology of rhabdomyosarcoma. *Genes Chromosom Cancer* 1989;1:23-35.
- Scrabble HJ, Johnson DK, Rinchik EM, Cavenee WK: Rhabdomyosarcoma associated locus and MyoD1 are syntenic but separate loci on the short arm of human chromosome 11. *Proc Natl Acad Sci USA* 1990;87:2182-2186.
- Scrabble HJ, Witte D, Lampkin BC, Cavenee WK: Chromosomal localization of the human rhabdomyosarcoma locus by mitotic recombination mapping. *Nature* 1987;329:645-647.
- Seidal T, Mark J, Hagmar B, Angervall L: Alveolar rhabdomyosarcoma: a cytogenetic and correlated cytological and histological study. *Acta Pathol Microbiol Immunol Scand* 1982;90:345-354.

Shapiro DN, Parham DM, Douglass MD, et al: Relationship of tumor cell ploidy to histological subtype and treatment outcome in children and adolescents with unresectable rhabdomyosarcoma. *J Clin Invest* 1991;9:159-166.

Shapiro DN, Sublett JE, Li B, Downing JR, Naeve CW: Fusion of PAX3 to a member of the forkhead family of transcription factors in human alveolar rhabdomyosarcoma. *Cancer Res* 1993;53:5108-5112.

Shapiro DN, Valentine MB, Sublett JE: Chromosomal sublocalisation of the 2;13 translocation breakpoint in alveolar rhabdomyosarcoma. *Genes chrom cancer* 1992;4:241-249.

Sheffield VC, Cox DR, Lerman LS, Myers RM. Attachment of a 40 base pair G+C rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes. *Proc.Natl.Acad.Sci* 1989;86:232-236.

Singh H, Sen R, Baltimore D, Sharp PA: A nuclear factor that binds to a conserved sequence motif in transcriptional control elements of immunoglobulin genes. *Nature* 1986;319:154-158.

Siple JC, Menninger JC, Hartley KO, Ward DC, Jackson SP, Anderson CW: Gene for the catalytic subunit of the human DNA-activated protein kinase maps to the site of the XRCC7 gene on chromosome 8. *Proc Natl Acad Sci* 1995;92:7515-7519.

Sorensen P, H., Shimada H, Liu XF, Lim JF, Thomas G, Triche TJ: Biphenotypic sarcomas with myogenic and neural differentiation express the Ewing's sarcoma EWS/FLI1 fusion gene. *cancer research* 1995;55:1385-1392.

Stapleton P, Weith A, Urbanek P, Kozmik Z, Busslinger M: Chromosomal localization of seven PAX genes and cloning of a novel family member, PAX-9. *Nat Genet* 1993;3:292-298.

Steingrimsson E, Moore KJ, Lamoreux ML, Burley SK, Copeland NG, Jenkins NA: Molecular basis of mouse *microphthalmia (mi)* mutations helps explain their developmental and phenotypic consequences. *Nat genet* 1994;8:256-263.

Stuart ET, Haffner R, Oren M, Gruss P: Loss of p53 function through PAX-mediated transcription repression. *EMBO J* 1995;14:5638-5645.

Sublett JE, Jeon I, Shapiro DN: The alveolar rhabdomyosarcoma PAX3/FKHR fusion protein is a transcription activator. *Oncogene* 1995;11:545-552.

Suzukawa K, Parganas E, Gajjar A, Abe T, Takahashi S, Tani K, Asano S, Asou H, Kamada N, Yokota J, Morishita K, Ihle JN. Identification of a breakpoint cluster region

3' of the ribophorin I gene at 3q21 associated with the transcriptional activation of the EVI1 gene in acute myelogenous leukaemias with inv(3)(q21q26). *Blood* 1994;84:2681-2688.

Tapscott SJ, Thayer MJ, Weintraub H: Deficiency in rhabdomyosarcoma of a factor required for MyoD activity and myogenesis. *Science* 1993;259:1450-1453.

Tapscott SJ, Weintraub H: MyoD and the regulation of myogenesis by helix-loop-helix proteins. *J Clin Invest* 1991;87:1133-1138.

Tassabehji M, Newton VE, Leverton K, et al: PAX3 gene structure and mutations: close analogies between Waardenburg syndrome and the Splotch mouse. *Hum Mol Genet* 1994;3:1069-1074.

Tassabehji M, Read AP, Newton VE, et al: Waardenburg's syndrome patients have mutations in the human homologue of the *pax-3* paired box gene. *Nature* 1992;355:635-636.

Tonegawa S: Somatic generation of antibody diversity. *Nature* 1983;302:575-581.

Tsujimoto Y, Croce CM: Molecular cloning of a human immunoglobulin gamma-chain variable sequence. *Nuc Acids res* 1984;12:8407-8414.

Tsujimoto Y, Jaffe E, Cossman J, Gorham J, Nowell PC, Croce CM: Clustering of breakpoints on chromosome 11 in human B-cell neoplasms with the t(11;14) chromosome translocation. *Nature* 1985;315:340-343.

Tsujimoto Y, Yunis J, Onoratoshowe L, Erikson J, Nowell PC, Croce CM: Molecular cloning of the chromosomal breakpoint of B-cell lymphomas and leukemias with the t(11;14) chromosome translocation. *Science* 1984;224:1403-1406.

Tsukamoto K, Tohma T, Ohta T, Yamakawa K, Fukushima Y, Niiikawa N: Cloning and characterisation of the inversion breakpoint at chromosome 2q35 in a patient with Waardenburg syndrome type I. *Hum Mol Genet* 1992;1:315-317.

Tycko B: Genomic imprinting: mechanisms and role in human pathology. *Am J Pathol* 1994;144:431-443.

Urbanek P, Wang ZQ, Fetka I, Wagner EF, Busslinger M: Complete block of early B-cell differentiation and altered patterning of the posterior midbrain in mice lacking PAX5/BSAP. *Cell* 1994;79:901-912.

Valentine M, Douglass EC, Look AT: Closely linked loci on the long arm of chromosome 13 flank a specific 2;13 translocation breakpoint in childhood rhabdomyosarcoma. *Cytogenet Cell Genet* 1989;52:128-132.

- Vogan KJ, Epstein DJ, Trasler DG, Gros P: The splotch-delayed (Spd) mouse mutant carries a point mutation within the paired box of the Pax-3 gene. *Genomics* 1993;17:364-369.
- Voullaire LE, Petrovic V, Sheffield LJ, Campbell P: Two forms of ring 13 in a child with rhabdomyosarcoma. *Am J Med Genet* 1991;39:285-287.
- Wallin J, Mizutani Y, Imai K, et al: A new Pax gene, Pax-9, maps to mouse chromosome 12. *Mamm Genome* 1993;4:354-358.
- Walther C, Guenet JL, Simon D, et al: Pax: a murine multigene family of paired box-containing genes. *Genomics* 1991;11:424-434.
- Waters SH, Saikh KU, Stavnezer J: A B-cell specific nuclear protein that binds to DNA sites 5' to immunoglobulin S-alpha tandem repeats is regulated during differentiation. *Mol Cell Biol* 1989;9:5594-5601.
- Weigel D, Jurgens G, Kuttner F, Seifert E, Jackle H: The homeotic gene fork head encodes a nuclear protein and is expressed in the terminal regions of the Drosophila embryo. *Cell* 1989;57:645-658.
- Weintraub H, Davis R, Tapscott S, et al: The myoD gene family:nodal point during specificatin of the muscle lineage. *Science* 1991;251:761-766.
- Weissman BE, Saxon PJ, Pasquale SR, Jones GR, Geiser AG, Stanbridge EJ: Introduction of a normal chromosome 11 into a Wilm's tumor cell line controls its tumorigenic expression. *Science* 1987;236:175-180.
- Whang Peng J, Knutsen T, Theil K, Horowitz ME, Triche T: Cytogenetic studies in subgroups of rhabdomyosarcoma. *Genes Chromosom Cancer* 1992;5:299-310.
- White MB, Carvalho M, Derse D, O'Brien SJ, Dean M. Detecting single base substitutions as heteroduplex polymorphisms. *Genomics* 1992;12:301-306.
- Wijnaendts LC, van der Linden JC, van Diest P, et al: Prognostic importance of DNA flow cytometric variables in rhabdomyosarcomas. *J Clin Pathol* 1993;46:948-952.
- Xu W, Rould MA, Jun S, Desplan C, Pabo CO: Crystal structure of a paired domain-DNA complex at 2.5 A resolution reveals structural basis for Pax developmental mutations. *Cell* 1995;80:639-650.
- Yancopoulos GD, Alt FW: Developmentally controlled and tissue specific expression of unrearranged Vh gene segments. *Cell* 1985;40:271-281.

Youil R, Kemper BW, Cotton RGH. Screening for mutations by enzyme cleavage of mismatch using T4 endonuclease VII. *Am.J.Hum.Genet.* 1993;53:s1257.

Zhong S, Wolf C, Spurr N: Chromosomal assignment and linkage analysis of the human glutathione S-transferase gene (GSTM1) using intron specific polymerase chain reaction. *Human Genetics* 1992;90:435-439.

Zlotogora J, Lerer I, Bar David S, Ergaz Z, Abeliovich D: Homozygosity for Waardenburg syndrome. *Am J Hum Genet* 1995;56:1173-1178.

Translin recognition site sequences flank chromosome translocation breakpoints in alveolar rhabdomyosarcoma cell lines

Jeremy G Chalk¹, Frederic G Barr² and Christopher D Mitchell¹

¹University of Oxford, Department of Paediatrics, John Radcliffe Hospital, Oxford OX3 9DU, UK; ²Divisions of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA 19104-6082, USA

Alveolar rhabdomyosarcoma is characterized by a t(2;13)(q35;q14) chromosome translocation, which leads to the fusion of the PAX3 and the FKHR genes. The resulting fusion gene encodes a chimeric protein which has aberrant transcriptional activity. We report the molecular definition of the genomic breakpoints on both derivative chromosomes in one case and the derivative chromosome 13 breakpoints in two other cases. The DNA sequences adjacent to the breakpoints on the derivative chromosome 13 are remarkable for their resemblance to recognition sequences for the protein translin. Gel shift analyses confirm that these sequences bind translin. These findings suggest that translin may not only be important in the genesis of chromosomal translocations in lymphoid malignancy, but also in translocations found in solid tumours.

Keywords:

Introduction

Numerous cytogenetic studies have demonstrated that most human neoplasms carry chromosomal aberrations. Leukemias, lymphomas and solid tumours have been shown to carry specific chromosomal abnormalities. A variety of studies have shown that these chromosome translocations result in the altered expression and/or structure of cellular gene products, leading to functional activation which contributes to the initiation or progression of the neoplastic state (Rabbitts, 1994).

Rhabdomyosarcoma, a malignant tumour arising from skeletal muscle precursors, is the commonest form of soft-tissue sarcoma in childhood. The majority of alveolar rhabdomyosarcomas are characterized by a specific chromosome translocation – t(2;13)(q35;q14) (Douglass *et al.*, 1987). This translocation has been shown to cause the fusion of 5' DNA-binding domains of the PAX3 gene to 3' regions of the FKHR gene, resulting in the expression of a chimeric protein product (Barr *et al.*, 1993; Galili *et al.*, 1993). The genomic breakpoints appear to occur consistently in the 17.5 kb intron 7 of the PAX3 gene on chromosome 2 and a large 130 kb intron 1 of the FKHR gene on chromosome 13 (Davis *et al.*, 1995).

Two consequences of the t(2;13) translocation have been described. Firstly, although the DNA binding of

PAX3/FKHR protein is impaired relative to that of PAX3 protein, the fusion protein is a much more potent transcriptional activator and will transform under conditions in which wild-type PAX3 is not (Fredericks *et al.*, 1995; Scheidler *et al.*, 1996). Secondly, both PAX3 and the PAX3/FKHR fusion protein inhibit the myogenic differentiation of myoblasts, with the latter demonstrating enhanced inhibitory ability. This ability is dependent on both the paired box and the homeodomain being intact (Epstein *et al.*, 1995).

Chromosome translocations in malignant tumours may arise by a variety of mechanisms. The translocations seen in some lymphomas and leukemias involve the immunoglobulin or T cell receptor genes (Rabbitts, 1994). The presence of the Ig/TCR recombinase signal sequence on both affected chromosomes indicates that errors in normal gene rearrangement are responsible for translocations in these cases. The breakpoint junctions of chromosome translocations in some other lymphoid malignancies are characterized by another consensus sequence. A protein, purified from the Lymphoid cell line DND-41, binds to oligonucleotides based on sequences from a number of different breakpoint junctions of this kind (Aoki *et al.*, 1995). The gene, named Translin, which codes for this protein has been identified and cloned, and is characterized by a leucine zipper DNA-binding motif, eight potential phosphorylation sites and one N-glycosylation site. Translin protein is located in the nucleus and cytoplasm in lymphoid cell lines, but in the cytoplasm only of non-lymphoid and non-haematopoietic cell lines (Aoki *et al.*, 1995).

Similar motifs have not yet been reported at translocation breakpoints in solid tumours, and the mechanisms by which translocations in these tumours occur remain obscure. We wished to know if there were sequence motifs flanking the breakpoints in the t(2;13) translocation which might predispose to rearrangements occurring in these particular regions.

Results

Cloning of the derivative chromosome 13 translocation breakpoint in Rh28

Vectorette PCR (Riley *et al.*, 1990) was used to clone novel unknown chromosome 13 sequences that have been juxtaposed to known, chromosome 2, PAX3 intronic sequences as a result of the t(2;13) chromosome translocation in rhabdomyosarcoma cell lines Rh28, Rh5 and TTC487. Genomic DNA was digested with an appropriate restriction enzyme, and the

resulting fragments then ligated to compatible vector-ette 'adapters' to produce a vectorette 'library'. DNA fragments spanning the translocation breakpoint were then identified by using an appropriately sited PAX3 intron primer together with a vectorette adapter primer in PCR. To confirm that any novel sequence cloned in this way is derived from chromosome 13, primers based on this sequence could be used to amplify a PCR product from a somatic cell hybrid which contains human chromosome 13 as its only human component (PGME1 or 289) (Cowell and Mitchell, 1989; Zhong et al., 1992).

The genomic translocation breakpoint of the Rh28 cell line has been mapped previously to a 0.6 kb region between a SacI and a ApaI site in the 17.5 kb intron 7 of the PAX3 gene (see Figure 1b). The presence of these two restriction sites was confirmed in the sequence of the intron in the Genbank database (accession no. U12259) (Macina et al., 1993). To verify that intron sequences 5' of the SacI site were still present in the derivative 13 chromosome, genomic DNA from Rh28 and the associated derivative chromosome 13 bearing RHF14 somatic cell hybrid (see Materials and methods) were used as template to amplify a 484 base-pair fragment from positions 12794 to 13278 in the intron. This product includes a PstI site at position 12845 (see Figure 1c). The expected products were amplified and subsequent digestion of these products with PstI restriction enzyme resulted in two fragments of the predicted size, 51 and 433 base-pairs.

Four vectorette libraries were constructed in total from genomic DNA from Rh28 or RHF14 cell lines, one pair using HaeIII restriction enzyme and another pair using Sau3A restriction enzyme. Nested primers were designed based on sequences immediately 5' of the SacI site and were used in conjunction with vectorette adapter primers. Primers 12800F and UVP were used

in a first round of amplification, and nested primer 13300F and USP were used in a second round amplification. A 200 bp product was amplified from the Rh28/HaeIII and the RHF14/HaeIII vectorette libraries. A 400 bp product was amplified from the Rh28/Sau3A and the RHF14/Sau3A vectorette libraries (see Figure 1c). All four products were cloned and sequenced. All four sequencing experiments indicate that the genomic breakpoint occurred at position 13402/3 of the PAX3 intron. Novel sequence presumed to represent FKHR intron on chromosome 13, was present 3' of the breakpoint in both Sau3A and HaeIII library clones. Two hundred and sixty nine base-pairs of novel sequence were identified in the Sau3A clones, of which the first 66 base-pairs exactly matched those present in the HaeIII clones (see Figure 2a).

Cloning of chromosome 13 sequences which span the translocation breakpoint in Rh28

The novel sequences found 3' of the derivative translocation breakpoint were used to design a second pair of nested primers. Primers 50R and UVP were used in a first round of amplification using the same vectorette libraries as above. Nested primers 30R and USP were used in second round amplification. A 150 bp product was amplified from the Rh28/HaeIII vectorette library and a similar 800 bp product was amplified from the RHF14/HaeIII vectorette library (see Figure 1d). The 800 bp products were the size predicted if amplification was from the derivative 13 chromosome template. The 150 bp product, present in Rh28 only, ought to represent normal chromosome 13 sequence. No product could be identified from the Rh28/Sau3A or the RHF14/Sau3A vectorette libraries. The 150 bp amplification product was then cloned and sequenced. The sequence immediately adjacent to the 30R primer matches that of the clones from the derivative chromosome 13 at the breakpoint. The clone also contained 111 bp of unique sequence 5' of the breakpoint. Together with the sequences 3' of the breakpoint on the derivative chromosome 13, a total of 380 bp of novel sequence have been cloned. To confirm that this novel sequence represents a contiguous region of chromosome 13, a primer pair spanning the breakpoint in this unique sequence were designed. The expected 313 bp product could be amplified from Rh28, Placenta, PGME1 or 289 genomic DNA. Sequence of the product confirmed the fidelity of the amplification reaction. Figure 2b includes the entire sequence obtained from chromosome 13.

Cloning of the derivative chromosome 2 translocation breakpoint in Rh28

To determine whether there have been any insertions and/or deletions associated with the chromosome translocation we used a similar approach as above to clone the reciprocal translocation breakpoint on the derivative chromosome 2. Nested primers, 12800F and 13500R, were designed, based on sequences in the PAX3 intron immediately 3' of position 13402 now established as the translocation breakpoint on the derivative chromosome 13.

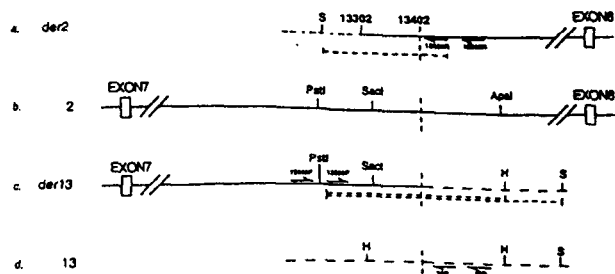


Figure 1 Schematic representation of the two normal and two derivative chromosome breakpoint regions in the Rh28 cell line. (a) Derivative chromosome 2. The vectorette PCR product extends from primer 13500R to a Sau3A restriction site (S). The derivative 2 breakpoint is at position 13302, 101 bp from the derivative 13 breakpoint at position 13402. (b) Normal chromosome 2. The breakpoint region in Rh28 has been mapped to a 0.6 kb region between a SacI and a ApaI site within the 17.5 kb PAX3 intron 7. (c) Derivative chromosome 13. PCR was used to verify the presence of sequences immediately 5' of the SacI site, including a PstI site at position 12845, in the derivative chromosome 13. Vectorette PCR products span the breakpoint from primer 13300F to either the HaeIII (H) or the Sau3A (S) site. The breakpoint is at position 13402 in the PAX3 intron. (d) Normal chromosome 13. The vectorette PCR product from primer 30R to a HaeIII site (H) spans the breakpoint region on normal chromosome 13

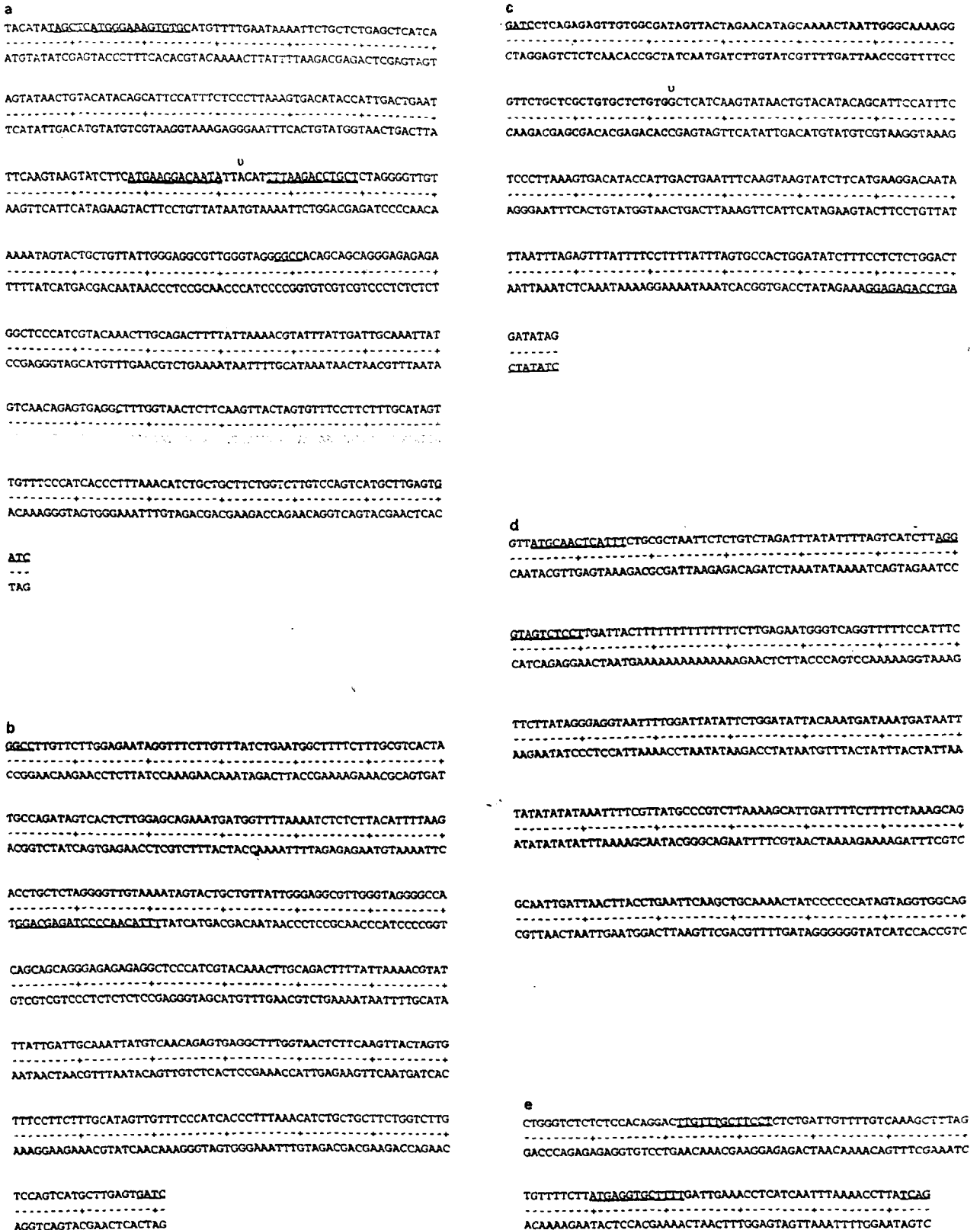


Figure 2 Sequences of vectorette PCR clones which span the (t;13) chromosome translocation breakpoints in the cell lines Rh28, Rh5 and TTC487. (a) Rh28 derivative chromosome 13 breakpoint sequence from PAX3 intron primer 13300F to Sau3A restriction site (both underlined). The breakpoint occurs at position 13402 of the PAX3 intron (u). 269 bp of novel chromosome 13 sequence follows the breakpoint. Putative Translin recognition sequences occur immediately 5' and 3' of the breakpoint (double underlined). (b) Normal chromosome 13 sequence which spans the Rh28 derivative chromosome 13 breakpoint. Chromosome 13 based primer 30R and HaeIII restriction site are underlined. 111 bp of novel sequence 5' of the derivative 13 breakpoint have been cloned. This together with the novel sequence cloned above represents a contiguous 380 bp of chromosome 13. (c) Rh28 derivative chromosome 2 breakpoint sequence from PAX3 intron primer 13500R to Sau3A restriction site (both underlined). This breakpoint occurs at position 13301 of the PAX3 intron (u) indicating that 101 bp of PAX3 intron sequence has been duplicated. Somatic cell hybrid PCR suggests the remaining 82 bp of sequence (bold) is not derived from either chromosome 2 or chromosome 13. (d) Rh5 derivative chromosome 13 sequence. The breakpoint occurs at position 11099 of the PAX3 intron (u). Putative Translin recognition sequences are double underlined. (e) TTC487 derivative chromosome 13 sequence. The breakpoint occurs at position 4557 of the PAX3 intron (u). Putative Translin recognition sequences are double underlined

possible that these sequences are deleted from derivative chromosome 2 but there is no straightforward way of testing this hypothesis as a somatic cell hybrid cell line containing derivative chromosome 2 material only is not available.

Amplification of the Rh28/HaeIII and Rh28/Sau3A vectorette libraries was carried out with primers 13600R and UVP, followed by nested 13500R and USP primers. A 300 bp product was amplified from the Rh28/Sau3A vectorette library and a 900 bp product amplified from the Rh28/HaeIII vectorette library (see Figure 1a). The 900 bp product was of the size predicted for normal chromosome 2. The 300 bp product, presumed to be derived from the derivative chromosome 2, was cloned and sequenced (see Figure 2c). The first 164 bp of sequence extending from the 13500R primer matched the original chromosome 2 sequence indicating that the translocation breakpoint in the derivative 2 chromosome is at position 13501/2 of the PAX3 intron. The breakpoint in the derivative 13 chromosome is at position 13402/3. As 101 bp of chromosome 2 sequence is present in both derivative chromosomes, there appears to have been a duplication of this region. The final 82 bp sequence of this clone is again novel and does not match the newly-derived chromosome 13 sequence described above. The 82 bp sequence could be amplified from Rh28 and placental genomic DNA, but not from somatic cell hybrids containing chromosome 2 or 13 as their only human components (GM10826B, 289 and PGME1). Since the 82 bp sequence is not derived from either chromosome 2 or chromosome 13 an insertion of material from some other chromosome may have occurred.

Cloning the derivative 13 breakpoints in two other t(2;13) bearing cell lines Rh5 and TTC487

The breakpoints in the PAX3 gene for the cell lines Rh5 and TTC487 have been mapped and lie in different regions of intron 7 to that of Rh28. The Rh5 breakpoint region lies in a 869 bp region between a XbaI site at position 11078 and a EcoRI site at position 11947, approximately 2 kb 5' of the Rh28 breakpoint within intron 7. The TTC487 breakpoint lies between a BglI site at position 4221 and a AvaII site at position 4741. This 520 bp region is 7 kb 5' of the Rh5 breakpoint region and 9 kb 5' of the Rh28

breakpoint. Vectorette libraries were constructed from Rh5 genomic DNA using Sau3A restriction enzyme, and from TTC487 genomic DNA using TaqI restriction enzyme. Nested primers were designed from sequences immediately 5' of each breakpoint region and used in conjunction with vectorette primers in a nested PCR protocol as above. A 300 bp product from the Rh5 reaction and a 400 bp product from the TTC487 reaction were cloned and sequenced.

The sequence of the 300 bp clone from the Rh5 library matches the published PAX3 intron 7 sequence up to position 11099 followed by 316 bp of novel sequence. PCR with primers based on this sequence on somatic cell hybrids, PGME and 289, confirmed that the 316 bp of novel sequence derives from chromosome 13. The sequence of the 400 bp clone from the TTC487 library matches the published PAX3 intron 7 sequence up to position 4557 followed by 51 bp of novel sequence. The 51 bp of novel sequence in this clone is not large enough to analyse by PCR to confirm whether it is derived from chromosome 13 or not.

Analysis of sequences around the translocation breakpoints

Figure 2a includes the entire 423 base pairs sequence from the derivative chromosome 13 junction fragment in Rh28 and Figure 2c the derivative chromosome 2 junction fragment in Rh28. Figures 2d and 2e include the derivative 13 sequences from cell lines Rh5 and TTC487 respectively. A search of the sequences around the translocation breakpoints did not reveal the presence of dyad symmetries, short direct repeats interspersed repetitive sequences, oligopurine/oligopyrimidine sites or recombinase signal sequences. We noted possible homology with the recently described consensus recognition sequence for Translin protein adjacent to the derivative chromosome 13 junction (Table 1). For example, the sequence in the PAX3 intron from position 13387 to 13399 is ATGAAGGA CAATA which has a 62% homology to the translin consensus sequence. This sequence is just 3 nucleotides from the derivative 13 breakpoint in Rh28. Three nucleotides on the 3' side of the Rh28 derivative 13 breakpoint is the sequence TTTAAGACCTGCT which again displays 62% homology to the translin consensus sequence. This finding implies that Translin could have a role in the genesis of the translocation

Table 1 Homology between the translin consensus sequence binding site and sites found flanking chromosome translocation breakpoints. The consensus sequence was derived by Aoki *et al* from the chromosome breakpoint junctions in 91 cases of human lymphoid neoplasm including the three sequences indicated below. These three sequences were also shown to bind translin protein in gel shift assays (Aoki *et al*, 1995)

| | Breakpoint position | Translin site | % homology | Reference |
|-----------|---------------------|---------------------|------------|----------------------------|
| Consensus | - | ATGCAGGCC (A/T) CCT | - | Aoki <i>et al.</i> , 1995 |
| bcr/abl | - | TTGCAGTGA G CCG | 62 | Chen <i>et al.</i> , 1989 |
| TCR/TTG2 | - | GTGAAGGCT C CGC | 62 | Boehm <i>et al.</i> , 1988 |
| IgH/bcl6 | - | CTGCACGTG C GAT | 43 | Baron <i>et al.</i> , 1993 |
| Rh28 | 13402 | ATGAAGGAC A ATA | 62 | |
| | | TTTAAGACC T GCT | 62 | |
| Rh5 | 11099 | ATGCAACTC A TTT | 62 | |
| | | AGGGTAGTC T CCT | 62 | |
| TTC487 | 4557 | TTGTTTGCT T CCT | 62 | |
| | | ATGCATATA T GTA | 43 | |

t(2;13)(q35;q14). Thus, we sought to determine if these DNA sequences could bind Translin protein.

Protein binding studies

Whole cell protein extracts were prepared from RD, an embryonal rhabdomyosarcoma cell line, Rh30, a t(2;13) bearing alveolar rhabdomyosarcoma cell line, and several other cell lines which previously have been shown to contain Translin protein (Aoki *et al.*, 1995). Gel mobility shift assays have shown that translin protein will bind to oligonucleotides which are based on sequences immediately adjacent to chromosome translocation breakpoints in a range of lymphoid tumours (Aoki *et al.*, 1995). An oligonucleotide, based on the sequence found 5' of the (2;13) translocation breakpoint in Rh28, was used in gel mobility shift assays. Protein from all the cell lines tested, including the rhabdomyosarcoma lines, bound to oligonucleotide. When unlabelled oligonucleotide was included as specific competitor at 40–160-fold molar excess over labelled probe, it was evident that formation of stable labelled complex with protein was considerably reduced (Figure 3a). When a non-specific unlabelled oligonucleotide was included as a competitor, no reduction in complex formation was observed. Specificity of complex formation was further characterized by preincubation with either preimmune IgM or anti-translin antibody (a gift of Dr M Kasai). The epitope of the translin antibody is at or near the DNA-binding element of translin protein and thus prevents its binding to labelled oligonucleotide in this assay. A small reduction is apparent when 0.1 μg of translin antibody is included in the reaction and a large reduction is apparent when 0.5 μg of antibody is included. There is no reduction in binding when the non-specific antibody is included. Binding of protein to oligonucleotide was blocked with the anti-translin antibody but was unaffected by preimmune antibody, further supporting the notion that translin binds sequences found next to the t(2;13) breakpoint in Rh28 (Figure 3b).

Discussion

We report the molecular cloning of the t(2;13) translocation breakpoints that lie in the PAX3 and FKHR genes, occurring on chromosomes 2 and 13 respectively, in the cell lines Rh28, Rh5 and TTC487. Breakpoints within both genes are intronic. At the molecular level, the rearrangement in Rh28 is complex, and involves both a duplication of 101 bp of PAX3 intron sequence adjacent to the breakpoint and a possible insertion of undetermined size. The functional consequences of this translocation mandate that the breakpoint in PAX3 lies in intron 7 and in FKHR in intron 1, as the resulting chimeric gene on the derivative chromosome 13 requires PAX3 exons 1–7 and FKHR exons 2 and 3 for activity. Is there any evidence that DNA sequence flanking the breakpoints predisposes to translocations involving these two genes? The presence of putative binding sites for the protein translin was noted. Translin has been previously identified by its ability to bind to conserved sequences adjacent to chromosome translocation

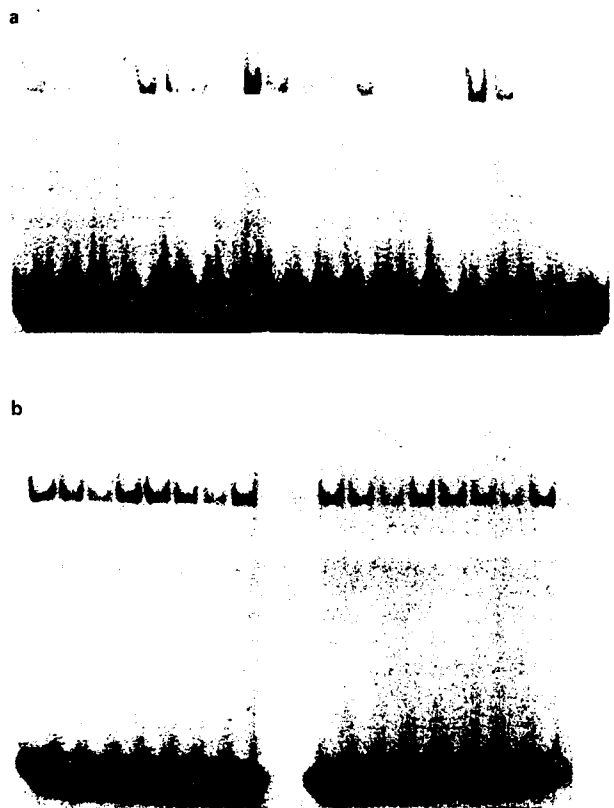


Figure 3 Binding of protein in a gel mobility shift assay to a single stranded, ^{32}P -labelled oligonucleotide which is based on the sequence 5' of the Rh28 derivative 13 chromosome. (a) In the first of each set of four lanes labelled oligonucleotide has been incubated with 7 μg of whole cell extract from the cell lines A431, K562, Hela, Rh30 and RD respectively. Lanes 2, 3 and 4 of each set include a 40, 80 and 160 molar excess of the same unlabelled oligonucleotide showing competition for protein. The final lane contains free labelled oligonucleotide only. Addition of non-specific unlabelled oligonucleotide does not result in competition for protein (not shown). (b) Specific inhibition of DNA binding activity. In the first of each set of four lanes labelled oligonucleotide has been incubated with 7 μg of whole cell extract from the cell lines K562, Hela, Rh30 and RD respectively. Lanes 2 and 3 of each set include 0.1 and 0.5 μg of anti-translin antibody respectively (rabbit IgG, gift from M Kasai). The epitope of the antibody overlaps the DNA-binding element of translin and thus prevents the interaction between translin and labelled oligonucleotide in the gel shift assay. Reductions in band intensity are apparent in incubations containing translin antibody for all whole cell extracts tested. The fourth lane of each set includes 0.5 μg of non-specific rabbit IgG antibody which has no effect on binding

breakpoints in a variety of lymphoid malignancies. Homology is seen in both components of the derivative chromosome 13, unlike the situation in lymphoid tumours where the recognition sequence is only normally present on the 5' side of the translocation. Typically, the DNA sequences seen in lymphoid tumours have a 50–70% homology to the consensus recognition sequence. We have shown that Translin protein is present in rhabdiod cell lines and that it will bind to oligonucleotides that are based on the putative translin binding sites found flanking chromosome translocation breakpoints.

These findings raise the possibility that the chromosomal rearrangement seen in alveolar rhabdomyosarcoma arises in part as a result of an interaction of damaged DNA with activity of Translin or a related

protein. Translin protein is located specifically in the cytoplasm of non-lymphoid cell lines (Aoki *et al.*, 1995). The data presented here raise the intriguing possibility that active transport of Translin protein is associated with a nuclear mechanism which may result in the formation of a chromosome translocation.

It is the first time such an observation has been made in a solid tumour. Clearly, it is now essential to analyse further examples of alveolar RMS and other non-lymphoid tumour specific translocation breakpoints to see if the junction fragments in these tumours also display Translin sequence homology.

Materials and methods

Rh28, Rh5 and TTC487 are alveolar rhabdomyosarcoma cell lines which contain the translocation t(2;13)(q35;q14) (Douglas *et al.*, 1987; Galili *et al.*, 1993; Sorensen *et al.*, 1995). RHF14 is a somatic cell hybrid which contains the derivative 13 chromosome from Rh28 but not the derivative chromosome 2 or normal chromosome 2 (Mitchell *et al.*, 1991). PGME1 and 289 are somatic cell hybrids with human chromosome 13 as their only human component (Cowell and Mitchell, 1989; Zhong *et al.*, 1992). GM10826B is a somatic cell hybrid with human chromosome 2 as its only human component. 289 and GM10826B genomic DNA were supplied by the UK HGMP Resource Centre. Rh28, RHF14 and PGME1 genomic DNA was extracted and purified using a Nucleon genomic DNA extraction kit (Scotlab).

Construction of vectorette libraries

Five μ g of each DNA was digested in buffer containing 5 units of restriction enzyme (Boehringer Mannheim) in a final volume of 50 μ l. Digests were then heat denatured at 70°C for 10 min. One μ l of each digest was ligated together with 3 pmol of appropriate vectorette adaptor (Genosys) in the presence of 1 mM ATP and 10 units of T4 DNA ligase (Amersham). After 4 h at room temperature, 60 μ l of water was added. Vectorette libraries were stored in aliquots at -20°C (Riley *et al.*, 1990).

Primer design

Primers that are specific to chromosome 2 have been based on the intron 7 sequence of the PAX3 gene deposited in Genbank (accession number U12259) (Macina *et al.*, 1995). The universal vectorette primer (UVP) and the universal sequencing primer (USP) were both supplied by Genosys. Other primers were synthesized by phosphoramidite chemistry on an Applied Biosystems 380B instrument: 12800F 5'-GCTGTGCTGAGTGCTGTGG-3'; 13300F 5'-TAGCTCATGGGAAAGTGTGC-3'; 50R 5'-GGTTGC-GGAGGGTTATTGTGC-3'; 30R 5'-AAATGTTGGGG-ATCTCGTC-3'; 13600R 5'-CATGAATAAGTTGTG-TAAGC-3'; 13500R 5'-CTATATCAGTCCAGAGAGG-3'; Rh5F 5'-GTAAATTAGAAGAACATCTGCC-3'; Rh5FN 5'-GTTATGCAACTCATTCTGC-3'; TTC F 5'-GCTG-

TTCAAACTCTCTGAAG-3'; TTC FN 5'-GCAATTATA-GATTGCAACC-3'

Amplification of vectorette libraries

A 5 μ l aliquot of each vectorette library was amplified in a total volume of 50 μ l containing 10 mM Tris-CL pH 8.3, 50 mM KCL, 1.5 mM MgCl₂, 0.1% gelatin, 100 μ M dNTP and 1 pmole of each primer. Reaction mixtures were denatured at 94°C for 5 min before adding 1 unit of Taq DNA polymerase (Perkin-Elmer Cetus) and 1 unit of Perfectmatch enhancer (Stratagene). Subsequent cycles consisted of 1 min at 94°C, 1 min at 58°C and 2 min at 72°C for a total of 40 cycles. One μ l of a 1:100 dilution of each reaction mixture was used in a further round of amplification under the same reaction conditions (but excluding Perfectmatch enhancer) using nested primers. Aliquots of the reactions were analysed on 1.5% agarose gels.

Amplification products were cloned directly into pCR11 plasmid vector (Invitrogen). Plasmid DNA from clones was isolated using Wizard Miniprep (Promega), and then alkali denatured prior to sequencing. For each sequencing reaction 2 μ g of double stranded plasmid DNA and 6 ng of primer were used in the Sequenase system ver 2.0 (USB/Amersham). In all sequencing experiments four independent clones were sequenced with vector primers from both strands.

Gel mobility shift assays

Whole cell extracts were prepared from between 5×10^6 and 10^7 cells which were washed in ice cold PBS and either scrapped or pelleted into eppendorf tubes and resuspended in 400 μ l of ice cold hypotonic buffer (10 mM HEPES-KOH pH 7.9, 1.5 mM MgCl₂, 10 mM KCl, 0.5 mM DTT, 0.2 mM PMSF) and incubated on ice for 10 min. Cells were then resuspended in 100 μ l ice cold hypertonic buffer (20 mM HEPES-KOH pH 7.9, 25% glycerol, 420 mM NaCl, 1.5 mM MgCl₂, 0.2 mM EDTA, 0.5 mM DTT, 0.2 mM PMSF) and incubated for a further 20 min. Cellular debris was removed by centrifugation for 2 min at 4°C, the supernatant was aliquoted and stored at -70°C. Protein concentration was determined by the Lowry method (Sigma).

Oligonucleotides were end-labelled with T4-polynucleotide kinase and [γ -³²P]ATP (both Amersham). Seven μ g of whole cell extracts were incubated at room temperature for 20 min with 10 000 c.p.m. of ³²P-labelled oligonucleotides (10 pg), in 20 μ l of binding buffer containing 10 mM Tris-HCl (pH 7.5), 50 mM NaCl, 1 mM EDTA, 1 mM DTT, 5% glycerol and 4 μ g of poly(dI-dC)(dI-dC). DNA-protein complexes were separated on 5% non-denaturing polyacrylamide gels. Gels were fixed, dried and autoradiographed.

Acknowledgements

This research and JC are funded by the Cancer Research Campaign. FGB study support provided by NIH grant CA61935 and ACS grant JFRA-492. Technical assistance provided by Lauren Nauta.

We thank the families and friends of Jason Campbell and Graham Thomas for their support. Dr Ann Harris, Dr Pete Little and Dr Richard Womer have provided help and advice

References

- Aoki K, Suzuki K, Sugano T, Tasaka T, Nakahara K, Kuge O, Omori A and Kasai M. (1995). *Nature Genetics.*, 10, 167-174.
- Baron B, Nucifora G, McCabe N, Espinosa R and LeBeau M. (1993). *Proc. Natl. Acad. Sci.*, 90, 5262-5266.
- Barr FG, Galili N, Holick J, Biegel JA, Rovera G and Emanuel BS. (1993). *Nat. Genet.*, 3, 113-117.
- Boehm T, Bulewela L, Williams D, White L and Rabbitts T. (1988). *EMBO J.*, 7, 2011-2017.

- Chen S. (1989). *Oncogene*, **4**, 195-202.
- Cowell JK and Mitchell CD. (1989). *Cytogenet. Cell. Genet.*, **52**, 1-6.
- Davis RJ, Bennicelli JL, Macina RA, Nycum LM, Biegel JA and Barr FG. (1995). *Human Molecular Genetics.*, **4**, 2355-2362.
- Douglass EC, Valentine M, Etcubanas E, Parham D, Webber BL, Houghton PJ and Green AA. (1987). *Cytogenet. Cell. Genet.*, **45**, 148-155.
- Epstein JA, Lam P, Jepeal L, Maas RL and Shapiro DN. (1995). *J. Biol. Chem.*, **270**, 11719-11722.
- Fredericks WJ, Galili N, Mukhopadhyay S, Rovera G, Bennicelli J, Barr FG and Rauscher FJ. (1995). *Mol. Cell. Biol.*, **15**, 1522-1535.
- Galili N, Davis RJ, Fredericks WJ, Mukhopadhyay S, Rauscher FJ, Emanuel BS, Rovera G and Barr FG. (1993). *Nat. Genet.*, **5**, 230-235.
- Macina RA, Barr FG, Galili N and Riethman HC. (1995). *Genomics*, **26**, 1-8.
- Mitchell CD, Ventris JA, Warr TJ and Cowell JK. (1991). *Oncogene*, **6**, 89-92.
- Rabbitts TH. (1994). *Nature*, **372**, 143-149.
- Riley J, Butler R, Ogilvie D, Finniear R, Jenner D, Powell S, Anand R, Smith J and Markham A. (1990). *Nucleic Acids Research*, **18**, 2887-2890.
- Scheidler S, Fredericks WJ, Rauscher FJ, Barr FG and Vogt PK. (1996). *Proc. Natl. Acad. Sci. USA*, **93**, 9805-9809.
- Sorensen PH, Shimada H, Liu XF, Lim JF, Thomas G and Triche TJ. (1995). *Cancer Research*, **55**, 1385-1392.
- Zhong S, Wolf C and Spurr N. (1992). *Human Genetics*, **90**, 435-439 (Abstract).