



**AN INVESTIGATION OF THE (4;11)(q21;p15)  
TRANSLOCATION IN ACUTE LYMPHOCYTIC  
LEUKAEMIA**

by

Damian J. Hussey, B.Sc. (Hons)

Thesis submitted for the degree of  
Doctor of Philosophy  
to  
The University of Adelaide  
Department of Medicine  
(The Queen Elizabeth Hospital)

June 2000

This copy is printed on archival paper



**Dual nuclear and cytoplasmic localisation of nrg**



**Lorna Catherine Hussey**  
5/5/1931 - 8/7/1996



**Vincent James Hussey**  
20/7/1925 - 23/5/1996

**This thesis is dedicated to my parents**

# Table of Contents

|   |          |
|---|----------|
| Table of Contents .....   | iv       |
| List of Tables .....  | ix       |
| List of Figures .....   | ix       |
| Abstract .....  | x        |
| Declaration .....   | xii      |
| Acknowledgments .....   | xliii    |
| Publications .....  | xv       |
| Selected conference presentations arising from this thesis .....                    | xvi      |
| Abbreviations .....   | xviii    |
| <br>  |          |
| <b>Chapter 1</b>  |          |
| <b>Introduction .....</b>   | <b>2</b> |
| 1.1 Haematopoiesis .....  | 2        |
| 1.1.1 Classification of the leukaemias .....  | 3        |
| 1.1.1.1 Morphological and cytochemical criteria for leukaemia classification .....  | 4        |
| 1.1.1.2 Immunophenotyping .....   | 5        |
| 1.1.1.3 Cytogenetics .....  | 6        |
| 1.1.1.4 Proposed revised classification of the haematological malignancies .....    | 6        |
| 1.1.2 Incidence of leukaemic subtypes .....   | 7        |
| 1.2 Chromosome translocations and leukaemia .....                                   | 9        |
| 1.2.1 Chromosome translocations alter normal patterns of gene expression .....      | 9        |
| 1.2.1.1 Deregulating the expression of a proto-oncogene .....                       | 10       |
| 1.2.1.2 Production of an in-frame fusion gene .....                                 | 11       |
| 1.2.2 Common targets of chromosome translocations .....                             | 13       |
| 1.2.2.1 Disruption of transcription factors .....                                   | 13       |
| 1.2.2.2 Disruption of protein kinases .....   | 14       |
| 1.2.3 Rearrangements associated with T-cell ALL .....                               | 15       |
| 1.3 Identifying genes at chromosome translocation breakpoints .....                 | 17       |
| 1.3.1 Positional cloning approach .....   | 17       |
| 1.3.2 Candidate gene approach .....   | 18       |
| 1.4 Aims .....  | 19       |
| 1.4.1 Identification of a novel T-cell ALL translocation, t(4;11)(q21;p14-15) ..... | 20       |
| 1.4.2 Preliminary molecular analysis of t(4;11)(q21;p14-15) .....                   | 22       |
| 1.4.3 Narrowing the chromosome 11 breakpoint region .....                           | 23       |

## Chapter 2

|  |           |
|--|-----------|
| <b>Materials and Methods</b> .....   | <b>25</b> |
| 2.1 Materials .....  | 25        |
| 2.1.1 Buffers and Solutions:.....  | 25        |
| 2.1.2 Media .....  | 27        |
| 2.2 Methods.....   | 27        |
| 2.2.1 Basic nucleic acid isolation and manipulation procedures.....                          | 27        |
| 2.2.1.1 Isolation of Lymphocytes/ mononuclear cells from peripheral blood.....               | 27        |
| 2.2.1.2 Thawing of frozen samples .....  | 28        |
| 2.2.1.3 DNA isolation from cell suspensions.....   | 28        |
| 2.2.1.4 RNA isolation procedure.....   | 29        |
| 2.2.1.5 Plasmid DNA isolation .....  | 29        |
| 2.2.1.6 DNA ligation.....  | 30        |
| 2.2.1.7 Preparation of competent <i>E. coli</i> and transformation with plasmid DNA.....     | 30        |
| 2.2.1.8 Sequencing.....  | 31        |
| 2.2.2 Cell culture.....  | 31        |
| 2.2.2.1 Thawing of cells for cell culture .....  | 31        |
| 2.2.2.2 Subculturing of NIH-3T3 mouse fibroblast cells .....                                 | 32        |
| 2.2.2.3 Transfection of NIH3T3 cells with plasmid DNA .....                                  | 32        |
| 2.2.3 Screening the $\lambda$ gt11 bacteriophage library.....                                | 33        |
| 2.2.4 Southern Analysis .....  | 34        |
| 2.2.5 Northern Analysis .....  | 36        |
| 2.2.6 Polymerase chain reaction (PCR) protocols.....   | 36        |
| 2.2.6.1 Standard PCR conditions .....  | 36        |
| 2.2.6.2 PCR conditions using Expand™ Long template and Expand™ Hi-Fi PCR System.....         | 37        |
| 2.2.6.3 Sequence of oligonucleotides for hybridisation, PCR and sequencing applications..... | 37        |
| 2.2.6.4 Restriction endonuclease digestion of PCR products.....                              | 40        |
| 2.2.6.5 PCR product purification .....   | 40        |
| 2.2.6.6 Reverse transcription (RT)- PCR conditions using MMLV.....                           | 41        |
| 2.2.6.7 Reverse transcription (RT)- PCR conditions using Superscript II .....                | 41        |
| 2.2.6.8 DNase I treatment of RNA.....  | 42        |
| 2.2.7 Procedure for 3' RACE.....   | 42        |

## Chapter 3

|  |           |
|--|-----------|
| <b>Investigation of the candidate gene, <i>ZNF195</i></b> .....  | <b>45</b> |
| 3.1 Confirmation of the 11p15.5 breakpoint region.....   | 47        |
| 3.2 Subcloning of zinc finger containing sequences from cosmid Z104.....                                     | 48        |
| 3.3 Are the zinc finger coding sequences present within Z104 transcribed? .....                              | 51        |
| 3.3.1 Northern analysis of adult and foetal tissues using the 336 bp <i>Eco</i> RI fragment as a probe ..... | 53        |
| 3.3.2 Assessing the specificity of the 336 bp <i>Eco</i> RI fragment for Z104.....                           | 53        |
| 3.3.3 Elevated expression of a 4.3 kb transcript in the t(4;11) patient.....                                 | 57        |
| 3.4 Isolation of lambda phage clones from a HUT-78 cDNA library.....   | 57        |
| 3.5 Assembly of restriction fragment and phage sequence into a contiguous sequence .....                     | 58        |
| 3.6 Description of the <i>ZNF195</i> gene present within Z104 .....  | 59        |
| 3.6.1 Determination of the intron/exon boundaries of <i>ZNF195</i> .....                                     | 61        |
| 3.6.2 Positioning of <i>ZNF195</i> within the Z104 cosmid.....   | 63        |
| 3.7 Alternative splicing of exons 4a and 4b.....   | 64        |
| 3.8 Revised Northern analysis with a probe from the 3'UTR of <i>ZNF195</i> .....                             | 66        |
| 3.9 Discovery of a deletion/insertion polymorphism in the <i>ZNF195</i> 3'UTR.....                           | 69        |
| 3.10 Assessing the imprinting status of <i>ZNF195</i> .....  | 71        |
| 3.11 Discussion.....   | 75        |
| 3.11.1 Northern analysis of <i>ZNF195</i> expression.....  | 75        |
| 3.11.2 Alternative splicing of <i>ZNF195</i> .....   | 76        |
| 3.11.3 What are the functions of <i>ZNF195</i> ?.....  | 76        |
| 3.11.4 Genomic position of <i>ZNF195</i> .....   | 77        |
| 3.11.5 Is <i>ZNF195</i> the 11p15.5 breakpoint gene in t(4;11)?.....   | 78        |

## Chapter 4

|   |           |
|---|-----------|
| <b>Identification of the t(4;11)(q21;p15) breakpoint genes</b> .....  | <b>81</b> |
| 4.1 Investigating <i>NUP98</i> as a candidate breakpoint gene in t(4;11)(q21;p15).....                                  | 82        |
| 4.1.1 Northern analysis .....   | 82        |
| 4.1.2 Southern analysis .....   | 84        |
| 4.1.3 Testing der(4) and der(11) for retention of <i>NUP98</i> sequences flanking the t(7;11)(p15;p15) breakpoint. .... | 85        |
| 4.1.3.1 Primer pairs 5' of the published <i>NUP98</i> breakpoint .....  | 86        |
| 4.1.3.2 Primer pairs 3' of the published <i>NUP98</i> breakpoint .....  | 87        |
| 4.2 Identification of the chromosome 4 breakpoint gene.....   | 89        |
| 4.2.1 3' RACE to determine the sequence fused to <i>NUP98</i> .....   | 91        |
| 4.2.1.1 Characterisation of smaller RACE products .....   | 96        |

|  |            |
|--|------------|
| 4.3 RT-PCR.....  | 98         |
| 4.4 Northern analysis .....  | 100        |
| 4.5 Discussion .....   | 103        |
| 4.5.1 Common features of t(4;11) ALL patients.....   | 105        |
| 4.5.2 t(4;11)(q21;p15) fuses the <i>NUP98</i> and <i>RAP1GDS1</i> genes .....                                    | 105        |
| 4.5.3 <i>NUP98</i> breakpoints.....  | 106        |
| 4.5.4 Fusion partners of <i>NUP98</i> .....  | 107        |
| 4.5.5 Retention of the FG repeat region in fusions involving nucleoporins.....                                   | 109        |
| 4.5.6 The role of smgGDS .....   | 109        |
| <br>   |            |
| <b>Chapter 5</b>   |            |
| <b>Cellular localisation of nrg and its components .....</b>   | <b>114</b> |
| 5.1 Creating constructs for studying the cellular location of nrg .....  | 115        |
| 5.1.1 Discovery of a novel exon in the predominant isoform of <i>RAP1GDS1</i> .....                              | 117        |
| in PBMNC .....   | 117        |
| 5.1.1.1 Other sequence variations in <i>RAP1GDS1</i> .....   | 121        |
| 5.1.2 Three fragment ligation for the creation of the <i>gfp-nrg</i> expression construct.....                   | 122        |
| 5.1.3 <i>NUP98</i> cloning.....  | 124        |
| 5.2 Cellular localisation of <i>gfp</i> tagged <i>nup98</i> , <i>nup98t</i> , <i>smgGDS</i> and <i>nrg</i> ..... | 130        |
| 5.3 Discussion.....  | 133        |
| 5.3.1 The predominant isoform of <i>smgGDS</i> in PBMNC contains 12 armadillo repeats .....                      | 133        |
| 5.3.2 Localisation of <i>gfp-nup98</i> .....   | 134        |
| 5.3.3 Localisation of <i>gfp-nup98t</i> .....  | 136        |
| 5.3.3.1 Possible disruption of <i>crml</i> mediated nucleocytoplasmic transport .....                            | 137        |
| 5.3.3.2 Possible disruption of <i>rae1</i> mediated nucleocytoplasmic transport.....                             | 139        |
| 5.3.4 Localisation of <i>gfp-smgGDS</i> .....  | 140        |
| 5.3.5 Localisation of <i>gfp-nrg</i> .....   | 141        |
| 5.3.5.1 <i>Nrg</i> in the nucleus .....  | 142        |
| 5.3.5.2 The role of SMAP .....   | 146        |
| 5.3.5.3 <i>Nrg</i> in the cytoplasm .....  | 147        |
| 5.4 Conclusion .....   | 148        |

## Chapter 6

|  |            |
|--|------------|
| <b>Conclusions and future studies</b> .....  | <b>150</b> |
| 6.1 A common theme for nup98 fusions .....   | 151        |
| 6.1.1 Cbp and p300 are key regulators of haematopoietic differentiation .....        | 152        |
| 6.1.1.1 Cbp/p300 interacts with aml1 .....   | 152        |
| 6.1.1.2 Haploinsufficiency of cbp causes haematopoietic defects .....                | 153        |
| 6.2 What is the role of the fusion partners of nup98? .....                          | 153        |
| 6.2.1 Why is nrg specific to T-cell ALL? .....                                       | 155        |
| 6.3 Future studies .....   | 156        |
| 6.3.1 Identifying proteins that interact with nrg and pathways affected by nrg ..... | 156        |
| 6.3.1.1 A common pathway for T-cell ALL.....   | 157        |
| 6.3.2 Does nrg effectively cause increased expression/activity of smgGDS?.....       | 158        |
| 6.3.3 Further studies on nrg localisation .....                                      | 159        |
| 6.3.4 Assessing the transforming properties of nrg.....                              | 159        |
| 6.3.4.1 <i>in vitro</i> transformation studies.....                                  | 160        |
| 6.3.4.2 <i>in vivo</i> transformation studies .....                                  | 161        |
| <b>References</b> .....  | <b>164</b> |
| <b>Appendix Journal Papers</b> .....   | <b>191</b> |



## List of Tables

|  |     |
|--|-----|
| Table 1.1 South Australian leukaemia cases over the 22 year period Jan 1977-Dec 1998 .....   | 8   |
| Table 1.2 Chromosomal rearrangements specifically associated with T-cell ALL.....  | 16  |
| Table 2.1 Oligonucleotides used in this study.....   | 38  |
| Table 4.1: Sequence characterisation of 3' RACE products .....   | 97  |
| Table 4.2: Clinical features, cytogenetics, immunophenotype and gene rearrangements of patients with a<br>t(4;11)(q21p14-15) ..... | 99  |
| Table 4.3 Translocations involving the <i>NUP98</i> gene.....  | 108 |

## List of Figures

|   |     |
|---|-----|
| Figure 1.1 t(4;11)(q21;p14-15) in a 21 year old adult male with T-cell ALL.....   | 21  |
| Figure 3.1 PCR analysis of 11p15.5 markers on the der(4) and der(11) containing somatic cell hybrid DNA ....                              | 49  |
| Figure 3.2 Selection of Z104 fragments for cloning and sequencing.....  | 50  |
| Figure 3.3 Northern analysis of T-cell lines and t(4;11) patient samples using the Z104 336 bp <i>Eco</i> RI fragment<br>as a probe ..... | 52  |
| Figure 3.4 Expression of Z104 related and <i>ZNF195</i> transcripts in adult and foetal tissues .....                                     | 54  |
| Figure 3.5 Southern analysis using the Z104 336 bp <i>Eco</i> RI fragment as a probe.....   | 56  |
| Figure 3.6a Method used PCR amplify insert DNA from g11 HUT-78 cDNA clones.....   | 60  |
| Figure 3.6b <i>ZNF195</i> Sequencing strategy.....  | 60  |
| Figure 3.7 The mRNA and putative amino acid sequence of <i>ZNF195</i> (Genbank Accession AF003540).....                                   | 62  |
| Figure 3.8 Alternative splicing of exons 4a and 4b in T-cell lines and normal donors.....   | 65  |
| Figure 3.9 <i>ZNF195</i> imprinting analysis .....  | 70  |
| Figure 4.1 Testing <i>NUP98</i> as a candidate breakpoint gene in t(4;11).....  | 83  |
| Figure 4.2 PCR analysis of the der(4) and der(11) containing somatic cell hybrids.....  | 88  |
| Figure 4.3 3' RACE to identify the <i>NUP98</i> fusion partner.....   | 93  |
| Figure 4.4 RT-PCR analysis and sequence of <i>NRG</i> .....   | 95  |
| Figure 4.5 Northern analysis of <i>NRG</i> expression.....  | 102 |
| Figure 4.6 Multiple tissue Northern analysis of <i>RAP1GDS1</i> .....   | 104 |
| Figure 4.7 Schematic representation of the nup98, smgGDS, nrg and nrg2 proteins .....   | 110 |
| Figure 5.1 Subcloning strategy for creating gfp tagged proteins .....   | 116 |
| Figure 5.2 Agarose gel electrophoresis of <i>RAP1GDS1</i> and <i>NRG</i> RT-PCR products .....  | 118 |
| Figure 5.3 The predominant isoform of <i>RAP1GDS1</i> in PBMNCs contains a novel 147 bp exon .....  | 120 |
| Figure 5.4 Three fragment cloning strategy for correcting <i>NRG</i> .....  | 125 |
| Figure 5.5 Subcellular localisation of gfp tagged proteins in NIH3T3 cells.....   | 127 |
| Figure 5.6 The additional 147 bp exon in <i>RAP1GDS1</i> encodes an armadillo repeat.....   | 135 |
| Figure 5.7 A model for the transforming mechanism of nrg in the nucleus.....  | 143 |

## Abstract

This thesis describes the results of an investigation to determine the molecular basis of an uncharacterised (4;11)(q21;p15) translocation in a patient with T-cell acute lymphocytic leukaemia (ALL). The chromosome 11 breakpoint had been narrowed to the region between the markers *D11S860* and *D11S470*.

A new zinc finger gene near the breakpoint region was cloned and characterised. This gene, subsequently named *ZNF195*, encodes an N terminal KRAB domain and 14 tandemly repeated Krüppel type zinc finger motifs. *ZNF195* was subsequently localised distal to *D11S470* and therefore distal to the chromosome 11 breakpoint in the (4;11)(q21;p15) translocation. This excluded disruption of *ZNF195* by the translocation.

Subsequently the nucleoporin 98 gene (*NUP98*) was identified at an acute myeloid leukaemia translocation. *NUP98* maps distal to *D11S470* and therefore within the breakpoint region. Analysis of somatic cell hybrids segregating the t(4;11) translocation chromosomes showed that the chromosome 11 breakpoint occurred within *NUP98*. The fusion partner of *NUP98* was identified as the *RAP1GDS1* gene using 3' RACE. In the *NUP98-RAP1GDS1* fusion transcript (abbreviated *NRG*), the 5' end of the *NUP98* gene is joined in frame to the coding region of the *RAP1GDS1* gene. This joins the phenylalanine-glycine (FG) repeat rich region of *nup98* to smgGDS (the most common name for the protein encoded by *RAP1GDS1*) which largely consists of tandem armadillo repeats. *NRG* fusion transcripts were detected in the leukaemic cells of two other adult T-cell ALL patients with a t(4;11)(q21;p15) translocation. This is the first report of a *NUP98* translocation in lymphocytic leukaemia and the first time that *RAP1GDS1* has been implicated in any human malignancy.

The cellular localisation of nrg was determined and compared to that of its protein components nup98t (t for truncated) and smgGDS. cDNAs were cloned into the pEGFP-C2 mammalian expression vector to create green fluorescent protein (gfp) tagged proteins. The location of the gfp tagged proteins within transfected NIH-3T3 mouse fibroblast cells was visualised using confocal microscopy. Gfp-nup98t was located in a punctate pattern around the nuclear envelope. Gfp-smgGDS was found throughout the cytoplasm and was absent from the nucleus. The hybrid protein gfp-nrg was present throughout the cytoplasm but was also visible within specific subnuclear domains. Therefore the formation of nrg results in nuclear localisation of the normally cytoplasmic smgGDS protein. Nup98t has been shown by others to have strong transcriptional transactivation ability while smgGDS has been shown to be important in blocking apoptosis in thymocytes. It is therefore possible that nrg promotes leukemogenesis through two independent pathways. In the nucleus nrg may act to deregulate transcription pathways critical to normal T-cell development, while in the cytoplasm, nrg may act to increase T-cell survival by promoting an anti-apoptotic phenotype.

## Declaration

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

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

Signed..

Date.....8/6/00.....

## Acknowledgments

I am very grateful to my supervisor, Alexander Dobrovic, for providing the opportunity to study in his laboratory. I would like to thank Alex for his extremely helpful, friendly, enthusiastic and supportive approach and for allowing me the freedom to determine the direction of my research. Thanks also for proof reading thesis drafts rapidly and providing helpful comments.

Thanks to all the medical staff of the Department of Haematology/ Oncology at the Queen Elizabeth Hospital (TQEH). In particular, thanks to Professor Ed Sage for fostering scientific research in the department. Thanks to TQEH Research Foundation and TQEH Department of Haematology/Oncology for providing my PhD scholarship and maintenance allowance.

Thanks to Ravi Krishnan, my co-supervisor, for helpful discussion and for providing cell lines. Thanks to all members of the Surgery, Obstetrics and Gynaecology, Rheumatology, Medicine, Cardiology and Transplantation Immunology laboratories for sharing their resources.

So many people gave generously of their time to assist various aspects of my research. In particular I would like to thank Jennie Finch, Greg Peters, Jenny Hardingham, Jing Xian Mi, Terry Gooley, Nigel Parker, Elizabeth Algar, Sarah Moore, Mario Nicola, Nick Wickham, Ian Lewis, Belinda Farmer, Emanuel Raniolo, Glen Smythe, Graham Webb, and Andreas Evdokiou. Thanks also to Petranel Ferrao, Tony Cambareri, Leonie Ashman and Tom Gonda for helpful discussions. Thanks to Professor Alec Morley (Department of Haematology/Oncology, Flinders Medical Centre) and Professor Bik To (Department of Haematology/Oncology, Institute of Medical and Veterinary Science) for access to clinical specimens. Thanks to Nat Albanese for assistance with plasmid sequencing and maintenance of cell cultures used for the protein localisation experiments of this study.

Thanks to Tina Bianco, for graciously sharing an office with me and for listening to me talk (often to myself) about experimental results. Thanks also for our many lively office (and lab) antics. I am indebted to Tina for her all of the help she has given me, including generous assistance with preparation of figures for publications and conference presentations. Thanks to Sarah Swinburne for sharing her knowledge on mRNA splicing. Thanks also to Sarah for making work a fun place to be and for organising some great social events.

Thankyou to my parents, to whom this thesis is dedicated, for their love, support and guidance. Thanks to my sister, her family and my brother for your support and for being there, you know how much you mean to me. Thanks to all the Chamalaun's and our friends for providing lots of encouragement and for helping out on many occasions.

Finally, I would like to thank my wife, Nicole, and my daughter, Mikayla. Nicole, thanks for making home a nice place to visit while I was living in the lab and a nice place to live while I was writing this thesis. I am very grateful for your constant support, involvement and encouragement throughout my PhD. Thankyou for proof reading the early drafts of this thesis

and for playing an instrumental role in the formatting, printing and collation of the final version. Dearest Mikayla, you are nearly 2 now. You have made the last 2 years of my PhD brighter and more meaningful. You have perfected the phrase "daddy, working, 'puter", and I'm sorry that you've had to say it so often lately.

## Publications

### *Publications arising directly from this thesis*

- Hussey DJ**, Parker NJ, Hussey ND, Little PF, Dobrovic A. (1997). Characterization of a KRAB family zinc finger gene, ZNF195, mapping to chromosome band 11p15.5. *Genomics* **45**:451-5.
- Hussey DJ**, Nicola M, Moore S, Peters GB, Dobrovic A. (1999). A new recurrent translocation t(4;11)(q21;p15.5) in T-cell leukaemia involving the *NUP98* and *RAP1GDS1* genes. *Blood* **94**:2072-79.
- Hussey DJ**, Albanese NO, Dobrovic A. Fusion of *NUP98* and *RAP1GDS1* in T-cell ALL results in aberrant nuclear localisation of the smgGDS protein. Manuscript in preparation, for submission to *Blood*.

### *Other Publications*

- Bianco, T., **Hussey, D.** and Dobrovic, A. (1999). Methylation-sensitive, single-strand conformation analysis (MS-SSCA): A rapid method to screen for and analyze methylation. *Hum Mutat*, **14**, 289-93.
- Cooper, S.J., Murphy, R., Dolman, G., **Hussey, D.** and Hope, R.M. (1996). A molecular and evolutionary study of the beta-globin gene family of the Australian marsupial *Sminthopsis crassicaudata*. *Mol Biol Evol*, **13**, 1012-22.
- Hussey, N.D., Donggui, H., Froiland, D.A., **Hussey, D.J.**, Haan, E.A., Matthews, C.D. & Craig, J.E. (1999). Analysis of five Duchenne muscular dystrophy exons and gender determination using conventional duplex polymerase chain reaction on single cells. *Mol Hum Reprod*, **5**, 1089-94.

## Selected conference presentations arising from this thesis

American Association of Cancer Research Special Meeting, Molecular Genetics of Cancer, Oxford, England, September 1997. **Hussey DJ**, Parker NJ, Hussey ND, Little PF, Dobrovic A. Characterisation of a KRAB family zinc finger gene, *ZNF195*, mapping to the *WT2* region of chromosome band 11p15.5.

Annual Meeting of the American Society for Haematology, San Diego, December, 1997. Dobrovic A, **Hussey, DJ** and Sage RE. *NUP98* is the gene at the chromosome 11 breakpoint of the translocation, t(4;11)(q21;p15.5) in a patient with acute lymphocytic leukemia. Blood 90 (suppl 1) 318a (1997).

Annual Meeting of the Haematology Society of Australia Sydney, September 1998. **Hussey DJ**, Sage RE, Dobrovic A. *NUP98* is the gene at the chromosome 11 breakpoint of the translocation, t(4;11)(q21;p15.5), in a patient with acute lymphocytic leukaemia. (chosen for the Presidential Symposium).

Queen Elizabeth Hospital Research Day, Adelaide, October 1998. Winner of the best presentation prize in the category of Higher Degrees (Basic Science). **Hussey DJ**, Sage RE, Dobrovic A. Cloning of the breakpoint genes of a recurrent chromosome translocation which defines a new clinical entity in leukaemia.

Annual Meeting of the American Association of Cancer Research, Philadelphia, April 1999. **Hussey DJ**, Nicola M, Moore S, Dobrovic A. A new recurrent translocation t(4;11)(q21;p15) in T-cell acute lymphocytic leukemia involving the *NUP98* and *RAP1GDS1* genes. Proc Amer. Assoc. Cancer Research 40:38-39 (1999).

Queen Elizabeth Hospital Research Day, Adelaide, October 1999. Winner of the best presentation prize in the category of Higher Degrees (Basic Science). **Hussey DJ**, Albanese N, Dobrovic A. Fusion of the *NUP98* & *RAP1GDS1* genes is recurrent in T-cell acute lymphocytic leukaemia and results in an altered cellular localisation of the RAP1GDS1 protein.



Lorne Cancer Conference, February 2000. **Hussey DJ**, Albanese N, Dobrovic A. Fusion of the *NUP98* and *RAP1GDS1* genes is recurrent in T-cell acute lymphocytic leukaemia and results in a fusion protein localising to the nucleus.

6th Annual Australian Society of Cytogeneticists Meeting, Adelaide, March, 2000. Hussey D, Albanese N, Nicola M, Moore S, Peters G, Dobrovic A. The “Adelaide chromosome”- a new recurrent translocation t(4;11) (q21;p15) in T-cell ALL involves the *NUP98* and *RAP1GDS1* genes.

Australian Society for Medical Research, South Australian Division, Annual Scientific Meeting, Adelaide, June 2000. **Hussey D**, Albanese N, Dobrovic A. Fusion of the *NUP98* and *RAP1GDS1* genes results in a protein localising to the nucleus.

## Abbreviations

|                 |  |
|-----------------|--|
| ALL             | acute lymphocytic leukaemia  |
| AML             | acute myeloid leukaemia  |
| BLAST           | basic local alignment search tool  |
| BLAST N         | Compares a nucleotide query sequence against a nucleotide sequence database.                               |
| BLAST X         | Compares a nucleotide query sequence translated in all reading frames against a protein sequence database. |
| BLAST P         | Compares an amino acid query sequence against a protein sequence database.                                 |
| bp              | base pair(s)   |
| CDNA            | DNA complementary to mRNA  |
| CML             | chronic myeloid leukaemia  |
| dATP            | 2'-deoxyadenosine-5'-triphosphate  |
| EST             | expressed sequence tag(s)  |
| g               | gram force of gravity  |
| GEF             | guanine nucleotide exchange factor   |
| gfp             | green fluorescent protein  |
| mRNA            | messenger RNA  |
| NE              | nuclear envelope   |
| NPC             | nuclear pore complex   |
| nrg             | nup98-smgGDS fusion protein  |
| <i>NRG</i>      | <i>NUP98-RAP1GDS1</i> fusion gene or mRNA  |
| nt              | nucleotide(s)  |
| <i>NUP98</i>    | gene encoding a nucleoporin of 98 kDa  |
| nup98           | nucleoporin of 98 kDa  |
| ORF             | open reading frame   |
| PBMNC           | peripheral blood mononuclear cells   |
| PCR             | polymerase chain reaction  |
| polyA RNA       | polyadenylated RNA   |
| 3' RACE         | rapid amplification of cDNA ends   |
| <i>RAP1GDS1</i> | gene encoding the rap1/smg guanine nucleotide disassociation stimulator                                    |
| RT              | reverse transcription  |
| UTR             | untranslated region  |

SI (Système International) units and the international code for DNA bases are used throughout

In accordance with the guidelines established by the Human Genome Nomenclature Committee, all genes referred to in this thesis are named in uppercase italics and all proteins are named in lower case plain font.

**Chapter 1**  
**Introduction**



# Chapter 1

## Introduction

The aim of the study described in this thesis was to determine the genes at the breakpoints of a chromosome translocation identified in an adult patient with acute T-cell leukaemia. Chromosome translocations, which involve exchange of DNA between chromosomes, are a common event in leukaemia. A large number of genes, many with key roles in haematopoiesis, have been identified by studying the breakpoints of chromosome translocations.

As the knowledge of the distinct genetic aberrations associated with particular forms of leukaemia increases, we can expect the development of therapies which are specific for the patient's malignancy. The associations between chromosome translocations and particular subtypes of leukaemia, and the therapeutic implications of these associations, have been exhaustively reviewed (Campana and Pui, 1995; Faderl *et al.*, 1998; Gilliland, 1998; Look, 1997; Pui, 1998; Rabbitts, 1994; Rowley, 1999; Sawyers, 1997; Uckun *et al.*, 1998).

This chapter provides an introduction to chromosome translocations and leukaemia and highlights, through specific examples, the value of studying chromosome translocations in order to identify novel genes involved in haematopoiesis. This chapter is particularly focussed on acute lymphocytic leukaemia (ALL) because the translocation reported in this thesis is recurrent in T-cell ALL.

### 1.1 Haematopoiesis

Haematopoiesis is an intricate process that leads to the production of all of the normal cellular components of blood. For normal haematopoiesis to occur, a very complex and delicate

balance of cell renewal, differentiation and death needs to be maintained. This balance is dependent upon the capacity of pluripotential stem cells to continually replicate in a regulated manner. Haematopoiesis also requires the appropriate differentiation of stem cells in response to various growth factors from the stroma and the surrounding environment.

Peripheral blood is composed of cells of two main lineages, the myeloid and lymphoid lineages. In the early stages of haematopoiesis, a pluripotent stem cell commits to one of these two lineages. The pathway from a committed lineage precursor to a fully differentiated cell type is subdivided into a number of intermediate steps. The myeloid lineage includes erythrocytes, platelets, basophils, eosinophils, neutrophils and monocytes. The lymphoid lineage comprises T-lymphocytes B-lymphocytes and natural killer cells.

### **1.1.1 Classification of the leukaemias**

Leukaemia is a malignant disease of the haematopoietic system, which is characterised by uncontrolled proliferation or expansion of haematopoietic cells that do not retain the capacity to differentiate normally to mature blood cells (definition from Sawyers *et al.*, 1991). Leukaemia occurs when there is a block or disruption along the differentiation pathways involved in normal haematopoiesis. The type of any particular leukaemia is related to the point in the differentiation pathway at which the leukaemic cell arises. Thus, the leukaemias have been divided into two main categories according to the predominant cell type; lymphoid and myeloid. The terms chronic and acute refer to the severity of the disease, ie the acute leukaemias are typically aggressive at presentation whereas the chronic leukaemias are relatively benign at presentation.

There are haematological malignancies that do not strictly classify as leukaemias because they do not involve both a block in differentiation and an uncontrolled cellular proliferation. For

example, in the myeloproliferative disorders, there is a clonal expansion of cells that are still able to differentiate. Alternatively, in the myelodysplastic syndromes there is no uncontrolled proliferation but myeloid cells are blocked at some stage along the maturation pathway. Both the myeloproliferative disorders and the myelodysplastic syndromes can progress to acute leukaemia (reviewed in Sawyers *et al.*, 1991).

#### **1.1.1.1 Morphological and cytochemical criteria for leukaemia classification**

Leukaemia can be classified in part by the morphology and cytochemical staining pattern of the leukaemic blast cell. These features are used to classify leukaemias according to a system devised by the French-American-British (FAB) cooperative group of haematologists. Features such as the size of the cell, the shape of the nucleus and number of nucleoli and the presence and staining pattern of cytoplasmic granules, are all considered in this system (reviewed in Bennett *et al.*, 1985).

Morphology and cytochemical staining are usually, but not always, sufficient to initially classify leukaemia as lymphoid or myeloid. There are eight FAB subtypes for AML (acute myeloid leukaemia), ie AML M0-M7, and three subtypes for ALL, ie ALL L1-L3. Subtype L1 represents the most common ALL morphology and offers the best prognosis. It occurs in about 85% of children and 30% of adults with ALL. Subtype L2 is the most common type in adults with ALL; 64% of adults with ALL have this subtype compared with only 15% of children. L2 morphology conveys a poorer prognosis than L1, even in children at low or intermediate risk of complications. Subtype L3 is the least common and is found in 7% of adults and 2% of children with ALL. Until recently, patients with ALL L3 had a poorer prognosis than those with L1 and L2. However, cytogenetic studies have shown a close link between ALL L3 and Burkitt's lymphoma. As a result, ALL L3 patients are now treated with a unique lymphoma protocol that has improved long-term survival rates (Bain, 1999).

The FAB classification of the AMLs is imprecise because it attempts to assign a single cell lineage to each leukaemia. In the case of the ALLs, the FAB criteria do not discriminate between B-cell lineage and T-cell lineage disease. Other features such as the immunophenotype and cytogenetics of the leukaemic cells are essential for accurate classification.

#### **1.1.1.2 Immunophenotyping**

Immunophenotyping categorises leukaemias by the cell-surface and cytoplasmic antigens expressed on the blast cells. These antigens are indicative of the cell lineage and maturation stage of the blast cell (reviewed in Bene *et al.*, 1999). Immunophenotyping is an invaluable tool in cases where the blast cells are developmentally immature and the myeloid and lymphoid lineages cannot be differentiated by morphology alone. There are still rare cases of leukaemia that do not express lineage-associated markers, these are referred to as acute undifferentiated leukaemias.

In AML, immunophenotyping assists the classification of undifferentiated leukaemia (FAB M0), erythroleukaemia (M6), and megakaryoblastic leukaemia (M7). In ALL immunophenotyping is essential to distinguish the B-cell and T-cell subtypes, and has allowed subclassification of 2 categories of T-cell ALL and 4 categories of B-cell ALL, each with a characteristic prognosis. In pre T-cell ALL, leukaemic transformation of a progenitor cell occurs at an earlier stage of the T-cell maturation pathway than in T-cell ALL. Pre T-cell ALL can be distinguished from T-cell ALL on the basis of immunophenotyping because in pre T-cell ALL the blast cells are surface positive for CD7 only, whereas T-cell ALL cells are surface positive for both CD7 and CD2. Both types express CD3 cytoplasmically (reviewed in Russell, 1997).

### **1.1.1.3 Cytogenetics**

In addition to the impact of immunophenotyping, improvements in cytogenetic and molecular biology techniques have greatly contributed to more accurate classification and improved understanding of the various leukaemias. It has become clear that recurring and consistent changes in chromosome number and structural alterations such as deletions, inversions and especially translocations are frequently associated with leukaemia (discussed below and reviewed in Look, 1997; Rabbitts, 1994). There is an association between specific chromosomal translocations in acute leukaemia and the stage of development at which normal cellular differentiation is blocked (reviewed in Look, 1997). This association often means that distinct subtypes of leukaemia can be classified by the specific genetic alterations associated with them. For example, the (15;17) translocation is typically found in the leukaemic cells of acute promyelocytic leukaemia (FAB M3) and not other forms of leukaemia (reviewed in Grignani *et al.*, 1994). Similarly, the (1;19) translocation is found in the leukaemic cells of pre-B acute lymphoblastic leukaemia, and no other forms of leukaemia (reviewed in Hunger, 1996).

### **1.1.1.4 Proposed revised classification of the haematological malignancies**

The combined impact of advances in immunophenotyping, cytogenetics and molecular biology on classification and prognostic categorisation of the leukaemias have caused the FAB classifications to become somewhat outdated. Leukaemic subtypes that can be grouped according to their cytogenetics, immunophenotype and clinical features do not always correlate perfectly with the FAB categories. For this reason, the World Health Organisation (WHO) have proposed a new set of classifications for haematological malignancies (Harris *et al.*, 2000). The report on the proposed classifications has only just been published and the classifications have not been adopted by the scientific/medical community as yet.



According to the WHO classifications, the haematological malignancies are stratified primarily according to lineage into 4 categories: myeloid neoplasms, lymphoid neoplasms, mast cell disorders, and histiocytic neoplasms. The WHO used the Revised European-American Classification of Lymphoid Neoplasms (Lister *et al.*, 1995) as a model for the revised classifications. Accordingly, within each of the 4 categories, distinct diseases are defined according to a combination of morphology, immunophenotype, genetic features and clinical syndromes. The overall goal of the re-classification was to determine a set of uniform criteria that can be used by pathologists and will have clinical relevance.

Importantly, the WHO decided that the terms L1, L2 and L3 were no longer relevant for the acute lymphoid malignancies, since L1 and L2 morphology do not predict immunophenotype, genetic abnormalities or clinical behaviour. It was decided that the precursor B-cell lymphoblastic malignancies should be grouped into 4 cytogenetic groups: t(9;22)(q34;q11), 11q23 rearrangements, t(1;19)(q23;p13) and t(12;21)(p12;q22), each with unique prognostic factors. Although T-cell ALL is associated with chromosomal rearrangements (Section 1.2.3), these do not correlate well with specific subtypes of T-cell ALL. Accordingly, the WHO made no recommendations for the incorporation of cytogenetic data in the classification of T-cell ALL. ALL L3, characterised by t(8;14)(q24;q32) and variants or *c-myc* rearrangement, is now to be diagnosed as Burkitt lymphoma in leukaemic phase, ie Burkitt cell leukaemia.

### **1.1.2 Incidence of leukaemic subtypes**

ALL is divided into childhood and adult categories with an arbitrary cut off point of 15 years of age separating the groups. In the state of South Australia, ALL accounts for approximately 84% of childhood leukaemias, and approximately 28% of all childhood cancers (Table 1.1). In contrast, ALL only accounts for about 7% of adult leukaemias and 0.2% of all adult cancers in South Australia. T-cell ALL accounts for 25% of adult (reviewed in Copelan and McGuire, 1995) and 15% of childhood (reviewed in Pui *et al.*, 1993) ALL cases. Despite its

relatively low overall incidence in adults, ALL has a bimodal distribution, with a second peak around age 50 and a low but steady rise in incidence with increasing age (reviewed in Faderl *et al.*, 1998).

Current cure rates show that adults with ALL have a far worse prognosis than children with this disease (reviewed in Pui and Evans, 1998). Nearly 80% of all childhood ALL cases are curable compared to only 30-40% of adult cases. The poor prognosis of adult ALL is due, in part, to the higher frequency of specific chromosome translocations in the leukaemic cells of adults (reviewed in Faderl *et al.*, 1998).

**Table 1.1 South Australian leukaemia cases over the 22 year period Jan 1977-Dec 1998.**

The data was taken from “Epidemiology of Cancer in South Australia, South Australian Cancer Registry, July 1999” and summarised here. CML is chronic myeloid leukaemia. CLL is chronic lymphocytic leukaemia.

|                                   | Age 0-14 |                                      |                                 |                             | Age 15-85+ |                                      |                             |                             |
|-----------------------------------|----------|--------------------------------------|---------------------------------|-----------------------------|------------|--------------------------------------|-----------------------------|-----------------------------|
|                                   | No.      | Percent of childhood cancers (t=941) | Percent of childhood leukaemias | Percent of total leukaemias | No.        | Percent of adult cancers (t=116,984) | Percent of adult leukaemias | Percent of total leukaemias |
| <b>ALL</b>                        | 261      | 27.7                                 | 83.7                            | 6.6                         | 241        | 0.2                                  | 6.6                         | 6.6                         |
| <b>CLL</b>                        | 1        | 0.1                                  | 0.3                             | 0.03                        | 1521       | 1.3                                  | 41.5                        | 38.2                        |
| <b>AML</b>                        | 42       | 4.5                                  | 13.5                            | 1.1                         | 1163       | 1                                    | 31.7                        | 29.2                        |
| <b>CML</b>                        | 5        | 0.5                                  | 1.6                             | 0.1                         | 428        | 0.4                                  | 11.7                        | 10.8                        |
| <b>Other specified leukaemia</b>  | 1        | 0.1                                  | 0.3                             | 0.03                        | 268        | 0.2                                  | 7.3                         | 6.7                         |
| <b>Unspecified cell leukaemia</b> | 2        | 0.2                                  | 0.6                             | 0.05                        | 46         | 0.04                                 | 1.3                         | 1.2                         |

## **1.2 Chromosome translocations and leukaemia**

A variety of structural cytogenetic abnormalities are associated with leukaemia. These include deletions, inversions and translocations. In a deletion, part of a chromosome is lost and if the deletion is recurrent it suggests the presence of a tumour suppressor gene. Inversions occur when a section of a chromosome is turned through 180 degrees. Translocations are the most common genetic abnormality associated with leukaemia and they occur in up to 65% of the acute leukaemias (reviewed in Look, 1997; Rabbitts, 1994). A translocation occurs when part of one chromosome is joined to part of another chromosome. More complex translocations can involve 3 or more chromosomes. The most frequently occurring translocation in leukaemia is a reciprocal translocation, which involves a direct exchange of DNA between chromosomes. The resulting abnormal chromosomes produced by the translocation are called derivative (*der*) chromosomes. Derivative chromosomes are stably inherited during mitosis as their centromeres are intact. It is customary to name a derivative chromosome by the origin of the centromere that it retains, eg a *der(11)* chromosome is a translocation chromosome that retains the centromere of chromosome 11.

### **1.2.1 Chromosome translocations alter normal patterns of gene expression**

Translocations and inversions involve chromosome breakage and rejoining at two or more sites, thereby creating new combinations of genetic material. More rarely, deletions also act to create new gene combinations. Examples of new gene combinations resulting from chromosomal deletion include the AML associated *MLL-LARG* fusion, which arises from an interstitial deletion at 11q23 (Kourlas *et al.*, 2000), and the T-cell ALL associated *SIL-TAL* fusion, which arises from an interstitial deletion at 1p32 (Aplan *et al.*, 1990). Inversions, deletions and translocations are commonly grouped under the general heading of chromosome rearrangements. Chromosome rearrangements can cause abnormal cell behaviour and result in leukaemia by two general mechanisms; deregulation of gene expression or production of a

chimaeric fusion gene. The first two translocations to be extensively studied at the molecular level illustrate both of these mechanisms.

#### **1.2.1.1 Deregulating the expression of a proto-oncogene**

Deregulated expression of a gene involves removal of a gene from its normal regulatory elements and juxtaposition with regulatory elements from another gene. Usually this occurs in lymphocytic leukaemias when either a T-cell receptor gene or an immunoglobulin gene is translocated near a proto-oncogene, and leads to activation of a proto-oncogene. An example is juxtaposition of the *c-myc* proto-oncogene to an immunoglobulin heavy chain (IgH) enhancer element, in t(8;14)(q24;q32) associated Burkitt cell leukaemia (Taub *et al.*, 1982). This translocation and its variants, t(2;8)(p12q24) and t(8;22)(q24;q11), result in 2-5 fold overexpression of *c-myc* due to increased transcription mediated by the IgH enhancer (Maguire *et al.*, 1983).

C-myc dimerizes with a protein called max, which can bind to DNA and to proteins (reviewed in Rabbitts, 1994). Overexpression of *c-myc* disrupts the normal equilibrium of the *c-myc*-max transcription network by favouring the creation of *c-myc*-max dimers which activate transcription. It is believed that this shift in equilibrium results in the transcriptional activation of the oncogene targets of *c-myc*.

Overexpression of *c-myc* results in a transformed phenotype not only by increasing the expression of growth promoting proteins, but also by repressing the expression of growth suppressing proteins (reviewed in Amati *et al.*, 1998). To give an example of both of these interactions, *c-myc* activates transcription of the cell cycle gene, *CDC25A* (Galaktionov *et al.*, 1996), whereas it represses transcription of the tumour suppressor gene, *GAS1* (Lee *et al.*, 1997).

### 1.2.1.2 Production of an in-frame fusion gene

The other event which can arise due to a chromosome translocation is the in frame fusion of gene segments from two chromosomes to produce a novel hybrid gene. In the vast majority of translocations that give rise to fusion genes, the breakpoints occur within the introns of each gene. The classic example of gene fusion arising from chromosome translocation is  $t(9;22)(q34;q11)$ , which fuses the majority of the *ABL* gene from chromosome 9 to the 5' exons of the *BCR* gene on chromosome 22 (de Klein *et al.*, 1982; Groffen *et al.*, 1984).

The *bcr-abl* fusion protein is encoded on the *der(22)* chromosome, called the Philadelphia chromosome, and can be expressed as two distinct isoforms,  $p210^{bcr-abl}$  or  $p190^{bcr-abl}$ , depending on the position of the breakpoint in the *BCR* gene. Typically the  $p210^{bcr-abl}$  isoform is associated with CML, while the  $p190^{bcr-abl}$  isoform is associated with ALL (reviewed in Dobrovic *et al.*, 1991). The reciprocal hybrid protein from *der(9)*, *abl-bcr*, is not expressed in all patients with the translocation (Melo *et al.*, 1993b), indicating that it is not the principal leukemogenic protein arising from  $t(9;22)$ . In support of this, experiments involving the reconstitution of mouse bone marrow expressing  $p210^{bcr-abl}$  (Pear *et al.*, 1998), and the creation of transgenic mice expressing  $p210^{bcr-abl}$  (Honda *et al.*, 1998) or  $p190^{bcr-abl}$  (Heisterkamp *et al.*, 1990), have shown that both of these isoforms are sufficient to cause the similar leukaemias to those that they are associated with in humans.

$t(9;22)(q34;q11)$  occurs in up to 95% of patients with CML, in about 1 to 2% of AML patients, as well as in up to 5% of children and 15% to 30% of adults with ALL (reviewed in Faderl *et al.*, 1998). This translocation is associated with a very poor prognosis in both childhood and adult ALL, with low remission duration and survival time. It requires drastic treatment, ie transplantation of patient bone marrow with allogeneic stem cells, and the treatment has a high risk of failure (reviewed in Look, 1997). Recent evidence suggests that

p190<sup>bcr-abl</sup> targets a primitive haematopoietic progenitor cell (Cobaleda *et al.*, 2000), with a high capacity for self renewal, and this undoubtedly contributes to the poor prognosis in t(9;22) ALL.

Despite the fact that it is now close to 20 years since t(9;22) was first characterised at the molecular level, there is still much to be learnt about the biological role of bcr-abl. The normal physiological role of bcr is not well understood, but it has been shown that the first exon of *BCR* encodes a protein segment with serine/threonine kinase activity (Maru and Witte, 1991). Bcr also contains domains involved in the regulation of rho family GTPases (reviewed in Narumiya, 1996), these are small G proteins with diverse functions ranging from cytoskeletal maintenance to transcriptional regulation (reviewed in Kjoller and Hall, 1999). Abl is a nuclear/cytoplasmic shuttling protein that functions as a non-receptor tyrosine kinase in the cytoplasm and is involved in regulating gene expression in the nucleus (Van Etten *et al.*, 1989; Taagepera *et al.*, 1998). In addition to its tyrosine kinase domain, abl has protein-protein and protein-DNA interaction motifs which are involved in cell cycle regulation and can induce cell growth arrest and apoptosis (reviewed in Shaul, 2000).

In contrast to normal abl, p210<sup>bcr-abl</sup> is exclusively cytoplasmic (Amson *et al.*, 1989) and exhibits increased tyrosine kinase activity (Konopka *et al.*, 1984). Both p210<sup>bcr-abl</sup> and p190<sup>bcr-abl</sup> inhibit apoptosis through combined activation of the growth promoting c-myc protein and the anti-apoptotic bcl-2 protein (reviewed in Sanchez-Garcia, 1997). The ras signal transduction pathway is activated by bcr-abl through its ability to recruit adaptor proteins, eg crkl and grb2, that connect it to positive regulators of ras activity (reviewed in Sawyers, 1997). p210<sup>bcr-abl</sup> also promotes growth by inhibiting rasGAP, a negative regulator of ras activity, and causing increased signalling through the ras signalling pathway (Kashige *et al.*, 2000).

## 1.2.2 Common targets of chromosome translocations

Through the study of novel chromosome translocations, many novel genes of importance in cancer have been discovered, and most of these genes show a very high level of conservation throughout evolution (Look, 1997). Two distinct types of pathways are targeted by chromosome translocations; transcription and cellular signalling. Translocations exert their oncogenic effect by directly altering gene transcription to promote disordered or blocked cellular differentiation, and/or by promoting aberrant transduction of growth signals through constitutive kinase activation (reviewed in Rabbitts, 1994).

### 1.2.2.1 Disruption of transcription factors

The most common gene type to be disrupted by chromosome translocations are those that encode DNA binding transcription factors (reviewed in Look, 1997). The *AML1* and *MLL* genes, on chromosomes 21 and 11 respectively, are both key transcriptional regulators of haematopoiesis and were both initially identified because of their involvement in chromosome translocations (reviewed in Rowley, 1999).

*AML1* was first identified by cloning the translocation genes of the AML associated t(8;21)(q22;q22) translocation (Erickson *et al.*, 1992). *AML1* is related to the *Drosophila* gene, *runt*, which encodes a sequence specific DNA binding protein (Erickson *et al.*, 1992). *Aml1* is a key regulator of a number of genes involved in haematopoietic development, and knockout of *AML1* results in embryonic lethality associated with severe defects in fetal liver haematopoiesis (Okuda *et al.*, 1996). Haploinsufficiency of *AML1*, caused by inactivating point mutations, is associated with familial thrombocytopenia and a predisposition to develop AML (Song *et al.*, 1999). Four translocations involving *AML1* have been cloned and it is thought that they all share a common dominant negative mode of action which involves repressing the transcription of *aml1* target genes (reviewed in Rowley, 1999). It was recently

shown that the (8;21) translocation modifies the intranuclear targeting of *aml1*, and this is also thought to contribute to its oncogenic properties (McNeil *et al.*, 1999).

Another transcription factor gene, *MLL* (also called *HRX*, *Htrx* and *ALL-1*) was discovered by cloning the translocation genes of the AML and ALL associated translocation t(4;11)(q21;q23). *MLL* is related to the *Drosophila* gene, *trithorax*, which is involved in maintenance of transcriptional regulation (reviewed in Gebuhr *et al.*, 2000). Like *AML1*, *MLL* has been shown through knockout experiments to be essential for normal haematopoiesis. *MLL* is involved in nearly 40 distinct leukaemia associated translocations, and 19 of these have been cloned and found to cause in-frame gene fusions (reviewed in Rowley, 1999). The normal role of *MLL* and its target genes is poorly understood at present and this has hampered the understanding of the way in which the fusion proteins operate. Recently, the *mll-enl* fusion protein, created by the t(11;19) translocation associated with B-cell or B-myeloid leukaemia, was shown to act through a gain of function involving the DNA binding ability of *mll* and the transcriptional transactivation capacity of *enl* (Slany *et al.*, 1998).

#### **1.2.2.2 Disruption of protein kinases**

Genes encoding protein kinases are also targeted by chromosome translocations (reviewed in Rabbitts, 1994). The t(9;22) translocation involves disruption and constitutive activation of the *abl* tyrosine kinase, as discussed in Section 1.2.1.2. In T-cell ALL, inappropriate expression of the *lck* and *tan-1* tyrosine kinases in T-cells is caused by translocations which juxtapose these genes with T-cell receptor enhancer elements (see Table 1.2). *Lck* is a *src* family tyrosine kinase and its overexpression has been shown to cause thymomas and peripheral blood malignancies in mice (Abraham *et al.*, 1991; Wildin *et al.*, 1991).



Other examples of translocations involving protein kinases are the t(5;12)(q33;p13) which occurs in a subset of chronic myelomonocytic leukaemia patients (Golub *et al.*, 1994), and the t(9;12)(q34;p13) which occurs in some patients with acute undifferentiated leukaemia and AML (Papadopoulos *et al.*, 1995). Both of these translocations involve fusion of the TEL gene from chromosome 12 to tyrosine kinase proteins. The translocation (5;12)(q33;p13) fuses the amino-terminal half of tel, which contains a DNA binding domain, to the transmembrane and tyrosine kinase domains of the PDGFR $\beta$  protein (Golub *et al.*, 1994). Similarly, the (5;12) translocation fuses the amino-terminal half of tel to the abl tyrosine kinase (Papadopoulos *et al.*, 1995). These translocations all result in constitutive activation of tyrosine kinase activity, which is mediated through the fusion partner, and it is thought that this common feature reveals a common oncogenic mechanism, ie constitutive positive signalling through the ras pathway via interactions with adaptor proteins such as crkl and grb2 (reviewed in Sawyers, 1997).

### **1.2.3 Rearrangements associated with T-cell ALL**

No specific karyotypic abnormality can be associated with a distinct clinical subtype of T-cell ALL, but there are a number of distinct chromosomal translocations that occur in association with T-cell ALL (Table 1.2). One of the most common genetic defects in T-cell ALL is not actually a translocation but an interstitial deletion on chromosomal region 1p32. This leads to the production of a fusion transcript called *SIL-TAL1* (Aplan *et al.*, 1990), and this is estimated to occur with a frequency of 16 to 26% in T-cell ALL (reviewed in Uckun *et al.*, 1998).

Translocations that deregulate proto-oncogene expression are frequent in T-cell ALL. The deregulated expression is a molecular consequence of the juxtaposition of genes that are

often silent in T-cells, with the enhancer elements of T cell receptor genes (reviewed in Rabbitts, 1994).

**Table 1.2 Chromosomal rearrangements specifically associated with T-cell ALL**

Translocations that are associated with T-cell ALL but also occur in other forms of haematological malignancy have been omitted from the table.

| <b>Rearrangement</b>                  | <b>Genes involved</b>   | <b>Reference</b>   |
|---------------------------------------|---|--|
| t(8;14)(q24;q11)                      | <i>c-MYC</i> 8q24<br><i>TCR-<math>\alpha</math></i> 14q11                   | (Shima <i>et al.</i> , 1986)   |
| t(7;19)(q35;p13)                      | <i>LYL1</i> 19p13<br><i>TCR-<math>\beta</math></i> 7q35                     | (Mellentin <i>et al.</i> , 1989)   |
| t(1;14)(p32;q11)<br>t(1;7)(p32;q35)   | <i>TAL1/SCL</i> 1p32<br><i>TCR-<math>\alpha</math></i> 14q11                | (Begley <i>et al.</i> , 1989)  |
| t(7;9)(q35;q32)                       | <i>TAL2</i> 9q32<br><i>TCR-<math>\beta</math></i> 7q35                      | (Baer, 1993)<br>review   |
| t(11;14)(p15;q11)                     | <i>RBTN1/Ttg1/LMO1</i> 11p15<br><i>TCR-<math>\delta</math></i> 14q11        | (Boehm <i>et al.</i> , 1991)   |
| t(11;14)(p13;q11)<br>t(7;11)(q35;p13) | <i>RBTN2/Ttg2/LMO2</i> 11p13<br><i>TCR-<math>\delta/\alpha/\beta</math></i> | (Boehm <i>et al.</i> , 1991)   |
| t(10;14)(q24;q11)<br>t(7;10)(q35;q24) | <i>HOX11</i> (10q24)<br><i>TCR-<math>\alpha/\beta</math></i>                | (Dube <i>et al.</i> , 1991;<br>Hatano <i>et al.</i> , 1991;<br>Kennedy <i>et al.</i> , 1991;<br>Lu <i>et al.</i> , 1991) |
| t(7;9)(q34;q34)                       | <i>TANI</i> (9q34)<br><i>TCR-<math>\beta</math></i> (7q34)                  | (Ellisen <i>et al.</i> , 1991)   |
| t(1;7)(p34;q34)                       | <i>LCK</i> (1p34)<br><i>TCR-<math>\beta</math></i> (7q34)                   | (Burnett <i>et al.</i> , 1991)<br>(Tycko <i>et al.</i> , 1991)   |
| t(14;21)(q11;q22)                     | <i>BHLHB1</i> (21q22)<br><i>TCR<math>\alpha</math></i> (14q11)              | (Wang <i>et al.</i> , 2000)  |
| del(1p32)                             | <i>SIL</i> (1p32)<br><i>TAL1</i> (1p32)                                     | (Aplan <i>et al.</i> , 1990)   |

## **1.3 Identifying genes at chromosome translocation breakpoints**

If one of the genes involved in a translocation is known then the identification of the partner gene is relatively straightforward. However, if both genes are unknown, there are two methods used for identifying them; the candidate gene approach and the positional cloning approach. Usually a combination of both methods is employed.

### **1.3.1 Positional cloning approach**

The first step in positional cloning of a translocation breakpoint gene is the identification of known genetic markers that flank the breakpoint region on one of the chromosomes involved in the translocation. Prior to the use of fluorescence in situ hybridisation (FISH), the simplest way to do this was to segregate the derivative chromosomes of the translocation into somatic cell hybrids. The relative position of the breakpoint with respect to the markers could then be determined by mapping the markers to one or the other derivative chromosome. With the invention of FISH it then became possible to map markers directly onto the derivative chromosomes of patient metaphase spreads.

Once the closest markers flanking the breakpoint region have been identified, they can be used as probes in Southern analysis of normal DNA and patient DNA, to test whether they detect novel restriction fragments in the patient DNA. The presence of novel restriction fragments indicates a chromosome rearrangement in or near the probe. This method was used in preliminary characterisation of the chromosome 21 breakpoint in the (8;21)(q22;q22) translocation (Gao *et al.*, 1991), which is associated with 15% of all AML cases (reviewed in Rowley, 1999).

The flanking markers can be used as probes to identify large genomic clones, eg yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs) and P1 artificial

chromosomes (PACs). If a genomic clone contains both flanking markers then it should span the breakpoint site. Once a genomic clone containing the breakpoint region has been identified it is subcloned into smaller, more manageable fragments, using cosmid or  $\lambda$  vectors. The subclones are then rescreened to identify the clone spanning the breakpoint. When the clone of interest has been identified there are a number of methods that can be used to identify genes present within it, including cDNA selection and exon trapping (reviewed in Monaco, 1994). Two common leukaemia associated translocations,  $t(4;11)(q21;q23)$  and  $t(8;21)(q22;q22)$ , were characterised using YACs containing the breakpoint regions on chromosomes 4 and 21 respectively (Gao *et al.*, 1991; Rowley *et al.*, 1990).

The final step in the positional cloning approach is testing the genes, identified from the genomic clones, for their disruption in the leukaemic cells. This step is essentially the candidate gene approach, which is described below.

### **1.3.2 Candidate gene approach**

The candidate gene approach tests all genes assigned to the breakpoint region, especially those genes which are already known or implicated for their involvement in cancer. A classic example of this approach is the identification of *ABL* as the chromosome 9 breakpoint gene in the  $(9;22)(q34;q11)$  translocation (reviewed in Dobrovic *et al.*, 1991). This translocation was first mistakenly identified as a recurring deletion of chromosome 22 in CML (Nowell and Hungerford, 1960). Over 10 years later, improvements in cytogenetic banding techniques revealed that the shorter chromosome 22 arose from a translocation, not a deletion, between chromosomal regions 9q34 and 22q11 (Rowley, 1973). Approximately ten years later, the *ABL* oncogene was mapped to 9q34, making it an obvious candidate breakpoint gene in  $t(9;22)$  (de Klein *et al.*, 1982). Soon after the mapping of *ABL*, Southern analysis was used to demonstrate that *ABL* was disrupted by  $t(9;22)$  (Groffen *et al.*, 1984).

Since this approach requires pinpointing of a candidate region, it usually requires some degree of positional cloning before candidate breakpoint genes can be assigned. However, now that a large number of translocation breakpoint genes have been cloned and are known to be involved in multiple translocations, the positional information obtained from karyotyping is now often sufficient to identify one of the breakpoint genes in novel translocations (reviewed in Rowley, 1999).

Recently, a powerful method involving mouse gene disruption through proviral integration, has been used to identify genes involved in leukemogenesis (Li *et al.*, 1999). This method identified more than 90 candidate leukaemia genes, some of which were already known to be involved in leukaemia, and some that were novel. Many of the genes identified through proviral disruption of the mouse genome have human orthologues. The chromosomal location of many of the human orthologues has been established and this will facilitate the candidate gene approach to identifying novel translocation breakpoint genes. Furthermore, the Human Genome Project has led to the availability of huge amounts of sequence and mapping data, which will also facilitate the candidate gene approach.

## **1.4 Aims**

The study described in this thesis involved an investigation of a novel T-cell ALL translocation, t(4;11)(q21;p15). At the onset of this project the translocation was thought to be comparatively rare. The aim of the study was to identify the genes disrupted by the translocation and potentially reveal novel genes with importance in cancer, or already characterised genes whose involvement in cancer had not been previously observed. The following discussion begins with the initial identification and molecular study of the translocation and finishes with the state of knowledge at the beginning of this study.

### 1.4.1 Identification of a novel T-cell ALL translocation, t(4;11)(q21;p14-15)

The translocation, t(4;11)(q21;p14-15), was observed in a 21 year old male with a large mediastinal mass and a white cell count of  $420 \times 10^9$ / ml. This translocation was the primary cytogenetic alteration at presentation apart from two marker chromosomes (Fig. 1.1a). Cell surface marker and DNA analysis was consistent with a diagnosis of T cell ALL (Hardingham *et al.*, 1991). Most of the peripheral blood blasts were positive for the T-cell markers CD7, CD2, and CD5 but did not express the mature T-cell markers CD4 and CD8. Most blasts were also positive for CD10, which is a B-cell lineage marker occasionally expressed in T-cell leukaemia, but not for CD19 which is a marker of more mature B-cells. Most of the blasts also expressed the stem cell marker CD34. A significant subpopulation of peripheral blood blasts expressed the T-cell marker CD3, the myeloid marker CD33 and the myelomonocytic marker CD11b. The mixed lineage phenotype was not considered to be the result of a biclonal expansion of leukaemic cells because most of the leukaemic blasts at presentation were positive for all of the following markers-CD2, CD7, CD10, CD34 and HLA-DR. Furthermore, all metaphases examined at presentation had the same karyotype.

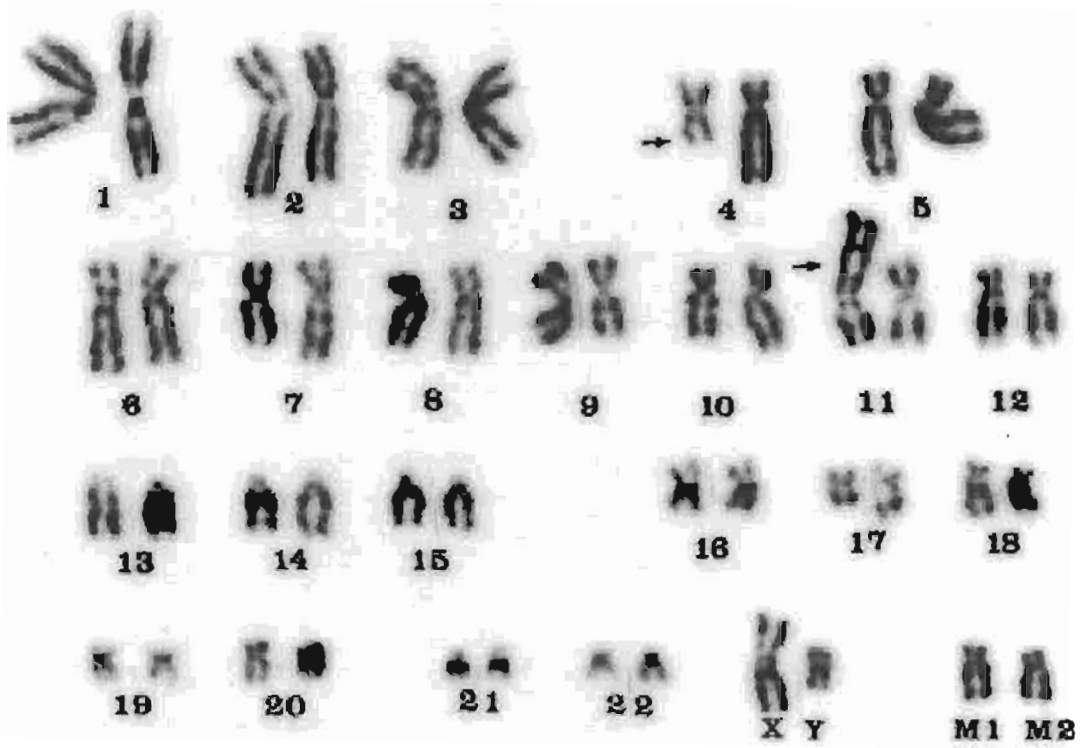
At the time it could not be excluded that the t(4;11)(q21;p14-15) was a cryptic variant of either the t(4;11)(q21;q23) which is principally associated with B-cell leukaemia (Childs *et al.*, 1988), or the T-cell leukaemia associated t(11;14)(p15;q11) (Le Beau *et al.*, 1986). Such variants have been reported in CML where the *BCR-ABL* fusion has occurred despite a lack of cytogenetic evidence for t(9;22) (reviewed in Dobrovic *et al.*, 1991). The clinical picture of the t(4;11)(q21;p14-15) patient does not well match that described by Pui *et al.* (1991) for the t(4;11)(q21;q23), suggesting that the t(4;11)(q21;p14-15) translocation is not merely a cryptic variant, but a distinct translocation. Furthermore, after the initial report on t(4;11)(q21;p14-15), the molecular basis of the t(4;11)(q21;q23) was described. The t(4;11)(q21;q23) fuses the *MLL* gene from 11q23 to the *AF-4* gene from 4q21 (Domer *et al.*, 1993; Gu *et al.*, 1992). This

**Figure 1.1 t(4;11)(q21;p14-15) in a 21 year old adult male with T-cell ALL**

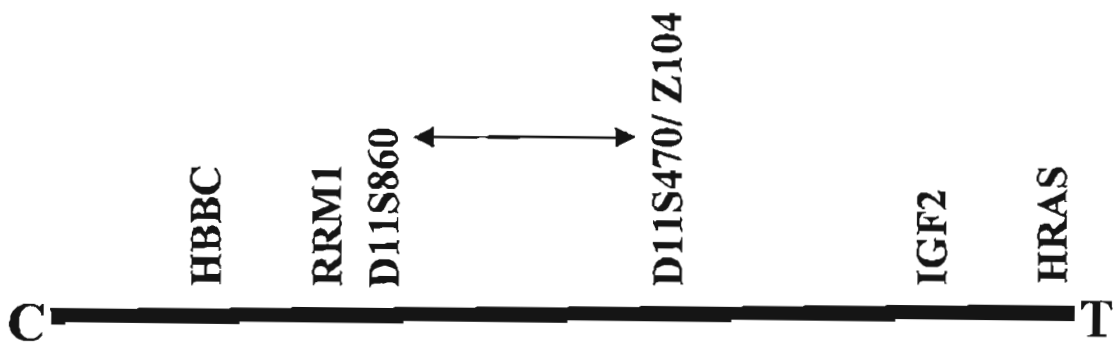
(a) The karyotype from a specimen of peripheral blood collected at presentation. The karyotype is 46,XY,t(4;11)(q21;p14-15),+2mar and was observed in all 15 metaphases scored (Hardingham *et al.*, 1991). The der(4) and der(11) chromosomes are indicated with arrows. Two marker chromosomes, of uncertain derivation, are denoted as M1 and M2.

(b) Position of the chromosome 11 breakpoint with respect to the 11p15.5 markers used for FISH and PCR mapping. The candidate breakpoint region is indicated by the arrowed line. The beta chain of hemoglobin (*HBBC*) and the H-ras oncogene (*HRAS*) are at the extremities of 11p15.5. C and T denote centromeric and telomeric respectively.

a)



b)





made it possible to test for involvement of *MLL* in t(4;11)(q21;p14-15). Southern blotting with an *MLL* probe excluded its involvement (Janet Rowley, unpublished data), thereby proving that the t(4;11)(q21;p15) was not a cryptic variant of the common (4;11)(q21;q23) translocation. For simplicity, the (4;11)(q21;p14-15) translocation will from now on be referred to as the (4;11) translocation. The more well known translocation will be named in full as the (4;11)(q21;q23) translocation.

Similar translocations to t(4;11) were observed in 3 other patients with T-cell ALL (Bloomfield *et al.*, 1986; Inoue *et al.*, 1985; Pui *et al.*, 1991). In two of these cases the translocation was the primary cytogenetic change at presentation. It was apparent that the translocation was recurrent and therefore that identification of the genes disrupted in t(4;11) would potentially reveal new oncogenes involved in leukemogenesis.

#### **1.4.2 Preliminary molecular analysis of t(4;11)(q21;p14-15)**

A positional cloning approach was initially adopted in order to narrow the breakpoint region, as it was not clear from cytogenetics if the 11p region disrupted was 11p14 or 11p15. The two derivative chromosomes were separated into different somatic cell hybrids and a range of PCR markers were mapped on the der(11) chromosome (Kalatzis *et al.*, 1993). Initial PCR analysis showed der(4) to be positive for *HRAS* but negative for *HBBC*, thereby narrowing the breakpoint to the 11p15.5 region between *HBBC* and *HRAS* (Fig. 1.1b). Subsequent PCR results revealed that der(4) was positive for *IGF2* but negative for *RRM1*, thus further refining the breakpoint region to between *RRM1* and *IGF2*. At that time there were no known markers that definitely mapped between *RRM1* and *IGF2*, therefore further localisation of the breakpoint was not possible. However, it was known that the *RBTN1* gene, disrupted in t(11;14)(p15;q11) T-cell leukaemias (Boehm *et al.*, 1991), mapped outside the candidate region and this excluded the involvement of *RBTN1*.

### 1.4.3 Narrowing the chromosome 11 breakpoint region

Chromosome region 11p15.5 is an intensely studied tumour-suppressor gene region, as it shows loss of heterozygosity in Wilms' tumour, rhabdomyosarcoma, adrenocortical carcinoma, breast, lung, hepatocellular, bladder, ovarian, and testicular cancer (reviewed in Hu *et al.*, 1997). These studies have resulted in an abundance of mapped genetic markers in 11p15.5. To further narrow the 11p15.5 breakpoint region in t(4;11), cosmid probes, mapped in other laboratories, were used in FISH (fluorescence *in situ* hybridisation) analysis on patient metaphase spreads. The two most proximal cosmids tested, *D11S470* and Z104, mapped to the der(4) chromosome and were therefore distal to the breakpoint (Dobrovic *et al.*, 1994); Finch, Peters and Dobrovic, unpublished). As the relative order of the *D11S470* and Z104 markers was unknown, it was unclear which cosmid was closest to the breakpoint. The breakpoint was therefore located to the interval between *RRM1* and *D11S470/Z104*.

After the FISH analysis of the cosmids was performed, a new chromosome 11p map was published (James *et al.*, 1994). This map showed that the *D11S860* marker was located between *D11S470* and *RRM1*. PCR of t(4;11) somatic cell hybrid DNA showed that *D11S860* was absent from the der(4) chromosome (Finch and Dobrovic, unpublished). This result suggested that the t(4;11) breakpoint was located somewhere between *D11S860* and *D11S470/Z104* (Fig. 1.1b).

The study described in this thesis commenced with further investigation of the t(4;11) 11p15.5 breakpoint region. At the beginning of the study there were no mapped genes or markers between *D11S860* and *D11S470/Z104*. Therefore, the first goal was to identify candidate genes within the breakpoint region and test these genes for disruption in t(4;11).

## **Chapter 2**

# **Materials and Methods**

# Chapter 2

## Materials and Methods

### 2.1 Materials

General reagents: General chemicals, media requirements and supplements were of analytical grade and were purchased from various suppliers.

Enzymes: Enzymes for general DNA manipulations were purchased from Roche (formerly Boehringer-Mannheim GmbH), Progen Industries, Fermentas or New England Biolabs. *Taq* polymerase was purchased from Perkin-Elmer. Proteinase K was purchased from Merck and T4 DNA ligase from Promega.

#### 2.1.1 Buffers and Solutions:

General solutions and buffers were made up with MilliQ water to the specified final concentration and sterilised by either autoclaving (103 kPa, 121°C for 20 min) or by filtering through a 0.22 micron membrane according to the methods given in Sambrook *et al.* (1989).

Solutions and additives for PCR were made up with Ultra Pure Water (Biotech) which is certified DNase and RNase free.

dNTP stock: a 40 mM stock solution of dATP, dGTP, dCTP and dTTP (10 mM each) was prepared in ultra pure water and the pH adjusted to between pH 7.0-pH 8.0 with 1 M Tris.

Deionised Formamide: Formamide was deionised by adding 5 g of Molecular Biology Grade AG<sup>®</sup> 501- $\times$ 8 (D) Resin (Bio-Rad) to 100 ml of formamide, magnetically stirred for 1 hr, filtered and stored at -20°C.

6 × Loading Buffer for DNA: 50 % glycerol, 0.2 M EDTA (pH 8.3) and 0.05 % bromophenol blue.

2 × Loading Buffer for RNA: Aliquots consisting of 500 µl of de-ionised formamide, 100 µl of 10 × MOPS, 167 µl 37% formaldehyde, 133 µl DEPC-treated water, 100 µl glycerol, 3 µl of 10 mg/ ml ethidium bromide, 0.025 % bromophenol blue and 0.025 % xylene were prepared and stored at -20°C.

Salmon Sperm DNA: 100 mg of Salmon sperm DNA (Sigma type III, sodium salt) was dissolved in 10 ml of dH<sub>2</sub>O and incubated at 37°C overnight. The salmon sperm DNA was sheared by sonication until it reached low viscosity.

20 % Sodium Dodecyl Sulphate (SDS): 20 g of sodium dodecyl sulfate (Sigma) per 100 ml of dH<sub>2</sub>O

SM buffer: 5.8 g NaCl, 2 g MgSO<sub>4</sub>·7H<sub>2</sub>O, 50 ml 1M Tris.Cl (pH 7.5), and 5 ml of 2% gelatin solution was brought to a final volume of 1 liter with H<sub>2</sub>O. The solution was stored at room temperature and used for the storage and dilution of bacteriophage lambda stocks.

20 × SSC: 3 M NaCl and 0.3 M sodium citrate pH 7.2

20 × SSPE: 3.6 M NaCl, 0.2 M Sodium phosphate, 20 mM EDTA pH 7.7

1 × TAE: 40 mM Tris base, 20 mM NaAc, 2 mM EDTA, pH 7.8 with glacial acetic acid

5 × TBE: 1 M Tris base, 0.9 M of boric acid and 0.2 M EDTA pH 8.3.

1 × TE: 10 mM Tris.HCl pH 7.5, 1 mM EDTA

1 × TES: 10 mM Tris.HCl pH 8.0, 1 mM EDTA, 0.1 M NaCl

## 2.1.2 Media

Luria Broth (LB) medium: 10 g Bacto-tryptone (Difco), 5 g Bacto-yeast extract (Difco) and 10 g NaCl was added per litre of water and the pH adjusted to 7.0 using NaOH. LB was sterilised by autoclaving.

LB Agar plates with selective antibiotic(s) and IPTG/X-gal colour selection: Bacto-agar (Difco, 15 g) was added to 1 litre of LB medium and autoclaved. After cooling to 50°C, the relevant antibiotic(s) were added. Final concentrations recommended by the manufacturer were Ampicillin (100 µg/ ml), Kanamycin (30 µg/ ml) and Tetracycline (50 µg/ ml). If appropriate, colour selection using the IPTG/ X-gal system was used. IPTG was added to a final concentration of 0.5 mM and X-gal to 80 µg/ ml (from a stock of 50 mg/ ml of dimethyl formamide). Approximately 35 ml of LB agar medium was poured into 85 mm Petri dishes and allowed to set. The plates were used immediately or stored at 4°C for up to 1 month.

Dulbecco's Modified Eagle Medium (DMEM): DMEM (Gibco-BRL, Life Technologies) was made according to the manufacturer's instructions with the addition of 60 µg/ ml benzylpenicillin (Sigma) and 50 µg/ ml dihydrostreptomycin (Glaxo Wellcome).

## 2.2 Methods

### 2.2.1 Basic nucleic acid isolation and manipulation procedures

#### 2.2.1.1 Isolation of Lymphocytes/ mononuclear cells from peripheral blood

Blood samples were centrifuged at 570 x g for 10 min in a bench top Beckman TJ6. After removal of the platelet rich plasma, the red cells and buffy coat were diluted to twice the original volume with PBS in a 50 ml centrifuge tube. Approximately 12 ml of Ficoll-Hypaque (Ficoll-Paque from Pharmacia) or Lymphoprep (Nyegaard) was carefully layered under the

diluted cells. The tube was then centrifuged at 570 x g for 25 min. Peripheral blood mononuclear cells (PBMNC) were skimmed from the interphase with a pasteur pipette in no more than 4 - 5 ml and placed in a 10 ml centrifuge tube made up to 10 ml with PBS. The number of cells present was determined and then they were pelleted at 570 x g for 10 min. The supernatant was discarded and the pellet resuspended in the residual PBS. The suspension was transferred to an Eppendorf tube using 1-1.5 ml of PBS and pelleted for 1 min. The supernatant was removed prior to nucleic acid extraction.

#### **2.2.1.2 Thawing of frozen samples**

Bone marrow or PBMNC samples frozen in nitrogen were thawed by warming the ampoule in a beaker of warm water (37°C) with shaking, once thawed the sample was transferred to an Eppendorf tube and pelleted briefly before discarding the supernatant. After this samples were treated the same as fresh samples.

#### **2.2.1.3 DNA isolation from cell suspensions**

Approximately  $20 \times 10^6$  cells were resuspended in 500  $\mu$ l of TES buffer supplemented with 30  $\mu$ l of Proteinase K. The tube was mixed by inversion before the addition of 30  $\mu$ l of 20 % SDS and incubation at 37°C overnight. 500  $\mu$ l of buffered phenol (Sigma) was added and the mixture agitated by spinning on a rotating platform for 15 min. Samples were centrifuged at 12,000 x g for 5 min and the aqueous layer removed to a clean Eppendorf tube. 500  $\mu$ l phenol/chloroform was added and the mixture returned to the rotating platform for 15 min. After centrifugation at 12,000 x g for 5 min, the aqueous phase was removed and the DNA precipitated by the addition of 2 x the volume of absolute ethanol. The pellet was washed once with 70% ethanol and then dried under vacuum. DNA was redissolved in either TE buffer or ultra pure water. The initial phenol step was sometimes omitted.

#### **2.2.1.4 RNA isolation procedure**

RNA was isolated by the addition of 1 ml of Trizol (Life Technologies) or Tri-Reagent (Sigma) per  $10^7$  cells. RNA was extracted using the manufacturer's protocol except that prior to the addition of chloroform, the cell homogenate was incubated at 37°C for 15 mins and then vortexed for 10 s. Furthermore, the RNA pellet was dissolved in ultra pure water instead of DEPC treated water.

#### **2.2.1.5 Plasmid DNA isolation**

Large scale plasmid DNA of high quality sufficient for sequencing and transfection analysis was prepared using either the Qiagen spin mini-prep kit (sequencing grade) or the Qiafilter plasmid midi kit (transfection grade). In both cases the manufacturers protocol was followed exactly.

Small scale plasmid DNA preparation in sufficient quality for assessing insert status and orientation by restriction enzyme digestion was carried out on 1.5 ml of bacterial culture. The culture was centrifuged at 12,000 x g for 1 min and the supernatant aspirated. The pellet was resuspended by vortexing in 100  $\mu$ l of ice cold lysis buffer (25 mM Tris-HCl, pH 8.0, 10 mM EDTA, 50 mM glucose). After incubation for 5 min at room temperature, 200  $\mu$ l of 0.2 N NaOH/1% SDS was added and mixed with the cells by inversion. After incubation on ice for 5 min, 150  $\mu$ l of ice cold potassium acetate solution was added, mixed by inversion, and again incubated on ice for 5 min. Bacterial cell debris and bacterial DNA was sedimented by centrifugation at 12,000 x g for 5 min and the supernatant containing the plasmid DNA was aspirated and placed in a new tube. RNA was then degraded by the addition of DNase-free RNase A (Roche), at a final concentration of 20  $\mu$ g/ml and incubated at 37°C for 30 min. The plasmid DNA was purified by phenol/chloroform extraction as described previously in Section 2.2.1.3. After chloroform extraction the plasmid DNA was precipitated, washed, dried



under vacuum and redissolved in 30  $\mu$ l of ultra pure water. 5  $\mu$ l of plasmid DNA was used for each restriction digest.

#### **2.2.1.6 DNA ligation**

Restriction fragments or PCR products were ligated into the appropriate vector DNA using T4 DNA ligase (Promega). The reaction mix was prepared according to the manufacturers instructions and ligation was performed at 15°C for 8 hours then 4°C overnight using a PTC100 thermal cycler (MJ Research). The plasmids used as cloning vectors in this thesis were pUC19, pGEM-T, pTarget (all from Promega) and pEGFP-C2 (Clontech).

#### **2.2.1.7 Preparation of competent *E. coli* and transformation with plasmid DNA**

For experiments requiring a high efficiency of transformation (such as those for transforming bacteria with ligated PCR product), competent *E. coli* JM109 were purchased (Promega) and transformed according to the manufacturers instructions.

For subcloning experiments, competent cells were prepared using the following protocol. A freshly grown single colony of *E. coli* TG-1 cells was used to inoculate 10 ml culture of LB broth and grown overnight at 37°C, with shaking (250 rpm). 250 $\mu$ l of this overnight culture was used to inoculate 50 ml of pre-warmed LB broth and grown to a density of OD<sub>600</sub> 0.3-0.6. This took approximately 3 hours. The cells were pelleted by centrifugation at 3,000 x g at 4°C. After decanting the supernatant the cells were resuspended by vortexing in 2.5 ml of ice cold 50mM CaCl<sub>2</sub>/ 20mM MgCl<sub>2</sub> and incubated on ice for 1 hour. 200  $\mu$ l aliquots of cells were prepared using pre-chilled pipette tips and 1.5 ml snap lock tubes. 2.5  $\mu$ l (1/4) of the ligation mix or 1 ng of control plasmid was added to an aliquot of cells, gently mixed by flicking the tube and incubated on ice for 40 mins. The cell/DNA mixture was heated to 42°C for precisely 2 mins and 15 s then placed on wet ice for 2 mins. 900  $\mu$ l of LB containing

20mM glucose was added to the cell/DNA mixture and then incubated at 37°C for 30-40 min. The supernatant was discarded after centrifugation at 12,000 x g for 1 min and the cells resuspended by pipetting and spread onto pre-warmed LB agar plates. containing the appropriate antibiotics and IPTG and X-gal if colour selection was possible. Plates were incubated at 37°C in a humid environment for 12-16 hours.

### **2.2.1.8 Sequencing**

The ABI Prism cycle sequencing kit was used according to the manufacturer's instructions except that a PTC-100 with hot bonnet thermal cycler was used. Cycling conditions were 96°C for 30 s, 50°C for 30 s, 60°C for 4 min for 25 cycles. Sequencing products were precipitated with sodium acetate and delivered to the Dept. of Molecular Pathology, IMVS, Adelaide for electrophoresis. DNA sequence analysis was carried out using the Sequencher (Genecodes) analysis package.

## **2.2.2 Cell culture**

### **2.2.2.1 Thawing of cells for cell culture**

The ampoule of frozen cells was thawed rapidly by shaking in a beaker of warm water. Once the cells were thawed they were transferred to a 10 ml tube. 10 ml of DMEM/ 10 % FCS (CSL) was added dropwise, first two drops at a time, then four drops at a time and so on in a logarithmic manner. An aliquot of the cell suspension was used to quantitate the cell density and the rest was pelleted at 400 x g. The supernatant was discarded and the cells resuspended in the residual solution. An appropriate amount of cells were seeded into 25 ml culture flasks and 5 ml of DMEM/ 10 % FCS was added. Cells were cultured at 37°C in air with 5 % CO<sub>2</sub> and the media was aspirated and replaced every 3 days or as required.

#### **2.2.2.2 Subculturing of NIH-3T3 mouse fibroblast cells**

NIH3T3 cells (a gift from Dr Leonie Ashman, Hanson Centre, Adelaide) were cultured in DMEM/ 10 % FCS in a 5 % CO<sub>2</sub> atmosphere at 37°C. Culture flasks were viewed using an inverted microscope to determine the density of the cells. Cells were passaged at intervals of approximately 3 days and were not allowed to exceed 80 % confluence. Cell cultures were split by washing in 1 × PBS and detached using trypsin (1.25 g trypsin (Difco), 4 g NaCl, 0.2 g KCl, 30 mg KH<sub>2</sub>PO<sub>4</sub>, 0.1 g Na<sub>2</sub>EDTA, 0.5 g glucose were dissolved in 4 ml dH<sub>2</sub>O. A few grains of phenol red indicator were added and the volume was made to 5 ml. The solution was filter sterilised through a 0.22 µm filter and stored at -20°C). 10 ml of DMEM/ 10 % FCS was added to the flask and pipetted up and down to disaggregate the cells. An aliquot was removed to determine the cell density and the rest transferred to a 10 ml tube and pelleted. The supernatant was discarded and the cell resuspended in the residue. An appropriate amount of cells were transferred to clean flasks and recultured.

#### **2.2.2.3 Transfection of NIH3T3 cells with plasmid DNA**

The day before transfection, cells were seeded in 6 well plates at a density of  $2.5 \times 10^5$ / well. The cells were transfected using either 2 µl of Superfect™ (Qiagen) or 3 µl of Fugene6™ (Roche) and 2 µg plasmid DNA according to the manufacturer's protocol.

To select for stably transfected cells, Geneticin (Life Technologies) was added to the medium at a concentration of 400 µg/ml, 48 h after the initial transfection procedure. This was followed by a 2 week selection period in medium containing 400 µg/ml of Geneticin. After selection was complete the cells were maintained in medium containing 200 µg/ml of Geneticin.

Twenty four hours prior to confocal microscopy, the cells were transferred into Permax chamber slides (Labtek®). Cells were viewed on the chamber slides using a Biorad MRC1000 confocal microscope. Technical assistance with confocal microscopy was kindly provided by Dr Peter Kolesik (Plant Research Institute, University of Adelaide, Waite Campus).

### **2.2.3 Screening the $\lambda$ gt11 bacteriophage library**

The titre of the amplified  $\lambda$ gt11 bacteriophage library (a gift from Dr Francis Shannon, Hanson Centre for Cancer Research, Adelaide) was determined by serial dilution. This involved growing *E. coli* Y1090 cells to saturation in LB broth containing 0.2% maltose and 10 mM MgSO<sub>4</sub> and 50  $\mu$ g/ml tetracycline. At saturation the cells were divided into 100  $\mu$ l aliquots in 10ml tubes. Top agar was melted and allowed to equilibrate at 45-50°C in a water bath. While the agar was equilibrating, serial 10 fold dilutions (starting from 1/100 through to 1/100,000) of the  $\lambda$ gt11 bacteriophage library were prepared using SM buffer. 100  $\mu$ l dilution aliquots were added to the *E. coli* aliquots. The bacteria and phage mixture was incubated for 20 min at room temperature, to allow the phage to adsorb to the *E. coli* and then in a water bath at 37°C for 10 min. The mixture was added to 2.5 ml of top agar, mixed lightly by vortexing, and poured onto pre-warmed 85 mm LB agar plates. The agar was spread by tilting and once the top agar had solidified they were incubated at 37°C. Phage plaques were easily visible after 4-5 hr of incubation and were counted at this time. The titre of the library was calculated from the plate that contained as many plaques as possible but not so many as to prevent accurate counting.

To screen for phage containing cDNA inserts of interest, no greater than 50,000 phage particles from the amplified library were mixed with 300  $\mu$ l of *E. coli* Y1090 in a 10 ml tube and incubated at 37°C for 20 min. 7 ml of 0.7% top agarose was equilibrated to 45-50°C then

added to each tube, one at a time, mixed by gentle vortexing, and plated onto 150 mm LB agar plates (LB agar plates had been prepared at least 2 days prior to plating to allow them to dry and prevent condensation during incubation). After incubation at 37°C for 4-5 hr the plates were incubated at 4°C for at least 1 hr to allow the top agarose to harden. Hybond N+ (Amersham) membrane was cut to the size of each plate and labelled then applied to the cold LB plates bearing the bacteriophage plaques. This was left on the plates for 10 min to allow the transfer of phage particles to the membrane. During this time the orientation of the filter to the plates was recorded by stabbing a needle through the filter into the agar at several asymmetric points around the edge of the plate. Membranes were then slowly removed from the plates using blunt forceps and placed on Whatman 3MM filter paper phage side up. To bind the plaques to the membranes, the membranes were dried at room temperature for at least 10 min. A large sheet of Whatman 3MM paper was saturated with 0.2 M NaOH/1.5 M NaCl and the membranes were placed on the filter paper, phage side up, and left for 1-2 min. The membranes were then placed phage side up on 3MM paper saturated with 0.4 M Tris.HCl, pH 7.6/ 2 × SSC for 1-2 min and then onto 3MM paper saturated with 2 × SSC for 1-2 min. DNA was fixed to the moist membranes using a UV crosslinker (Stratalinker 1800, Stratagene) with 120,000 microjoules/ cm<sup>2</sup> of UV irradiation according to the manufacturer's instructions. The membranes were stored at room temperature until hybridisation. Hybridisation with radiolabelled DNA probe and autoradiography was performed as described for Southern analysis.

#### **2.2.4 Southern Analysis**

Genomic DNA was digested with 10 units of restriction enzyme per µg of DNA in a total volume of 200 µl. PCR products were digested by removing a 5-10µl aliquot and digesting in a total volume of 20 µl. One × BSA (New England Biolabs) was always added unless already present in the restriction enzyme buffer.

The restriction enzyme digested DNA products were ethanol precipitated and heated to 56°C for 5 min to disrupt base pairing between cohesive termini before loading on an agarose gel (0.7% - 3% with 0.5 × TBE buffer supplemented with 0.25 µg/ml ethidium bromide). Agarose gels containing genomic DNA or cosmid DNA samples were depurinated with 0.25 M HCl for 15 min and rinsed in dH<sub>2</sub>O. Gels containing PCR product DNA were not depurinated. DNA was transferred to a Hybond-N<sup>+</sup> nylon membrane (Amersham) using 0.4 M sodium hydroxide and a capillary transfer protocol as per the manufacturer's instructions. The transfer was left to proceed overnight and the membrane was washed with 2 × SSC for 5 min. The membrane was either used immediately or sealed in plastic wrap and stored at -20°C.

Membranes were prehybridised for at least two hours at 42°C in 20 ml of a solution containing 50% deionised formamide, 1 M NaCl, 10% dextran sulphate, 1% SDS, and 0.2 mg/ml of sonicated and denatured salmon sperm DNA. Radiolabelling of PCR products or restriction enzyme digested DNA fragments was carried out using the GIGAprime DNA labelling kit (Geneworks), according to the manufacturer's instructions. Template DNA (25 ng) was labelled with 50 µCi of radioactive  $\alpha$ -<sup>32</sup>P dCTP (Geneworks) in a total volume of 25 µl. The labelled probe was heated to 100°C for 5 min and added to the prehybridisation mixture. Hybridisation was carried out at 42°C overnight.

Membranes were washed to a final stringency of 0.1 × SSC, 0.1% SDS at 65°C. The membrane was wrapped in plastic and exposed to X-ray film (Hyperfilm, Amersham) using Amersham Hyper cassettes with one intensifying screen and placed at -80°C for exposure. When necessary membranes were stripped of radiolabelled probe by pouring a boiling solution of 0.5 % SDS directly onto the membrane and rocking until the solution reaches

room temperature. The blot was then wrapped in glad wrap and placed at  $-20^{\circ}\text{C}$  until the next hybridisation.

End-labelling of oligonucleotides was carried out using a Terminal Kinasing Kit (Geneworks) according to the manufacturer's instructions and  $\gamma\text{-}^{32}\text{P}$  ATP (Geneworks). Hybridisation of Southern blots of PCR products or cosmid DNA fragments to end-labelled oligonucleotide probes was as described except that the hybridisation solution was  $4 \times \text{SPE}$ , 1% SDS, 5 % blotto (5% nonfat dried milk powder, 0.02% sodium azide) and 0.1 mg/ml denatured salmon sperm DNA. After washing at  $42^{\circ}\text{C}$  in  $2 \times \text{SSC}$ , 0.1% SDS, the membranes were autoradiographed at  $-80^{\circ}\text{C}$ .

### **2.2.5 Northern Analysis**

Five to 10  $\mu\text{g}$  of RNA was electrophoresed on a denaturing 1% agarose/ 1.2M formaldehyde gel, blotted onto Brightstar plus membrane (Ambion, Austin, TX) according to the manufacturer's protocol and fixed to the membrane by UV crosslinking as described in Section 2.2.3. Radiolabelled probes were prepared as described in Section 2.2.4 and hybridised to membranes at  $42^{\circ}\text{C}$  in a 20 ml solution of 1 M NaCl, 10 % dextran sulphate, 1 % SDS, 50 % deionised formamide and 0.2 mg/ml denatured salmon sperm DNA. Membranes were washed to a final stringency of  $0.2 \times \text{SSPE}$ , 0.1% SDS at  $65^{\circ}\text{C}$  and autoradiographed at  $-80^{\circ}\text{C}$ .

Multiple tissue Northern blots were purchased from Clontech (Palo Alto, CA) and hybridisation was as described above except that Express-Hyb (Clontech, Palo Alto, CA) solution was used.

### **2.2.6 Polymerase chain reaction (PCR) protocols**

#### **2.2.6.1 Standard PCR conditions**

Standard PCR reactions were carried out in 0.5 ml tubes as follows. Each reaction contained a final of 100 ng of each oligonucleotide primer, 200  $\mu\text{M}$  each dNTP, 0.5 units of Ampli Taq®

Gold polymerase (Perkin Elmer), 1 × PCR buffer (10 mM Tris-HCl, 50 mM KCl), 2.0 mM MgCl<sub>2</sub> and 100 ng of genomic DNA or other template. In earlier experiments, Ampli Taq® polymerase was used with the same reaction mix except that the final concentration of MgCl<sub>2</sub> was 1.5 mM. The reaction was made up to a final volume of 50 µl using ultra pure water and placed into a PTC-100™ Programmable Thermal Cycler (MJ Research Inc). After an initial denaturation step of 94°C for 9 min reactions were cycled for 35-45 cycles of 94-96°C for 30 sec, annealing at 60-65°C for 1 min and extension at 72°C for 1 min. The annealing temperature for PCR was determined empirically by calculating the melting temperature of each primer, assigning 4°C to each G or C and 2°C to each A or T, then subtracting 5°C from this temperature. If two primers were used whose annealing temperature were very different a touchdown PCR protocol was used. Touchdown PCR was performed after the initial denaturation and consisted of 10 cycles of 95°C for 30 sec, 70°C minus 1°C per cycle for 1 min and 72°C for 1 min, followed by 25 cycles of 95°C for 30 sec, 60°C for 1 min and 72°C for 1 min plus 5 sec per cycle.

#### **2.2.6.2 PCR conditions using Expand™ Long template and Expand™ Hi-Fi PCR System**

The manufacturer's protocol (Roche) was followed using a concentration of 2.5 mM MgCl<sub>2</sub> in a final volume of 50 µl. Cycling conditions were chosen in accordance with the manufacturer's directions.

#### **2.2.6.3 Sequence of oligonucleotides for hybridisation, PCR and sequencing applications**

Primers were designed for PCR amplification using the Amplify Primer Design program (Shareware Program) and their sequence is shown in Table 2.1. Custom Primers were purchased from Geneworks. Random Hexamers (Pharmacia) and nonomers (Geneworks) were resuspended to give a final concentration of 250 ng/ µl. The 500 µl aliquots were stored at -20.



**Table 2.1 Oligonucleotides used in this study**

The oligonucleotides shown in this Table were designed by this author for this study except where due reference is made. Names beginning with "Zf" denote sequence from *ZNF195* (Genbank AF003540). Shown in the comments column for the *ZNF195* oligonucleotides is the position of the 5' base of the oligonucleotide relative to Genbank entry AF003540, except for Zf3'UTRR where the position is given relative to an 11p15.5 P1 artificial chromosome (Genbank AC000378). Names beginning with an "N" denote a primer corresponding to sequences within *NUP98* (Genbank U41815) and those beginning with an "R" denote sequence from *RAP1GDS1* (Genbank X63465). For the *NUP98* and *RAP1GDS* sequences the number in the primer name is equal to the nucleotide position of the cDNA sequence from Genbank that the 5' base of the oligonucleotide corresponds to.

| Name        | Sequence 5'-3'                        | Comments/ References            |
|-------------|---------------------------------------|---------------------------------|
| D11S470F    | ATTCCACAGAGCCTGGCAAAGG                | (Miwa <i>et al.</i> , 1993)     |
| D11S470R    | GTTAAGGAAGCCCAGACAGAAAGC              | (Miwa <i>et al.</i> , 1993)     |
| D11S860F    | GCAACACGTACACACTGAGACA                | (McNoe <i>et al.</i> , 1992)    |
| D11S860R    | TAGTATTGCCATAGAAGAAGC                 | (McNoe <i>et al.</i> , 1992)    |
| pUC/M13F    | GTAACAACGACGGCCAGTG                   | Universal Sequencing Primer     |
| pUC/M13R    | GTCGATACTGGTACTAATGC                  | Universal Sequencing Primer     |
| gt11F       | TATCGACGGTTTCCATATGGGG                |                                 |
| gt11R       | GAAATACGGGCAGACATGGCCT                |                                 |
| AP          | GGCCACGCGTCGACTAGTACT <sub>(17)</sub> | Gibco BRL Life Technologies     |
| AUAP        | GGCCACGCGTCGACTAGTAC                  | Gibco BRL Life Technologies     |
| pTargetF    | GACGTTGTAAAACGACGGCCAGAG              |                                 |
| pTargetR    | TTACGCCAAGTTATTTAGGTGACA              | Promega Cat. No. Q4461          |
| Zf3'UTRF*   | ACCCAACAAATCTCATAAATGTGGA             | 2172-2196                       |
| Zf3'UTRR*   | GAGAATAGAATGACATCAATATTCGC            | 89190-89165 of Genbank AC000378 |
| Zf5'UTRF    | CGGGAGATCCAGAAGTGAAAC                 | 1-21                            |
| ZfEcoF      | TATACACAGTCCTCACACCTCAGTGAACA         | 1384-1412                       |
| ZfEcoR      | GGTCTGATAACTGCTTAAAGATGTTGCCAC        | 1541-1570                       |
| ZfKRABBF    | GTCTGTAAGCCAGGCCTGATCACC              | 184-207                         |
| Zfseq out3' | ACTTTACCCAGTCCTCCA                    | 1634-1651                       |
| Zfseq out5' | GGTGTGAGGACTGTGTAT                    | 1402-1385                       |
| ZfspacerR   | TGCTCTGGCAGAAGGTCTTGGGT               | 527-505                         |
| N81F        | GGACTCCTGACACTTCCCCTTC                |                                 |
| N145F       | ATGTTTAACAAATCATTGGAACACCC            |                                 |
| N185R       | GTGCCACCCCCAAAGGGTGTTTC               |                                 |
| N301F       | GGAAATTCACAGACTAAACCAGGAG             |                                 |

| Name    | Sequence 5'-3'                      | Comments/ References |
|---------|-------------------------------------|----------------------|
| N388R   | CAAACCCAAAGCCAGTGCTTGTGG            |                      |
| N585F   | CTTTGGGCCAAGTAGTTTTACAGC            |                      |
| N716R   | GCAGTAATACACTGGTGCTTGGTACT          |                      |
| N988F   | ACCAGCCTCTTCAGCAAACCATTG            |                      |
| N1252R  | GCTTGTATTGCCAAACAGGGTCCG            |                      |
| N1265F  | GAAGCAGCACACCAGTGCACCT              |                      |
| N1384R  | TTCCAAGAGTCCCAGGTGCTGGT             |                      |
| N1428F  | GGCATCTTTGTTTGGGAACAACC             |                      |
| N1459F  | ATTGGAGGGCCTCTTGGTACAGGAG           |                      |
| N1491F  | GGCCCCTGGATTTAATACTACG              |                      |
| N1503F  | TAATACTACGACAGCCACTTTGGG            |                      |
| N1511F  | CGACAGCCACTTTGGGCTTTGGAGC           |                      |
| N1531R  | CAAAGCCCAAAGTGGCTGTCCG              |                      |
| N1585F  | CAGGCTGTCTCCAGCAGCACA               |                      |
| N1681R  | CCTTCTTCTTAGGGTCTGACATC             |                      |
| N1742F  | ATAAACTGACACCCCGCCCTGCC             |                      |
| N1848R  | GGATGGTTCATCGTCATCCAGCC             |                      |
| N2334F  | GGATGACCTTGCTAAAATTACCAATG          |                      |
| N2481R  | CCTCCGGATATGCACAATATCATCC           |                      |
| N2721F  | GGTCTCCATTTTTCTAAGTATGGC            |                      |
| N2907R  | TCACTGTCCTTTTTTCTCTACCTG            |                      |
| NCy9IR  | AAGTTCCCCGGGTCATAAAACAGCTTTGTATTAGC |                      |
| NtnR    | CTATACTGGGGCCTGGGGGGCT              |                      |
| NXho IF | CTCATCTCGAGTATGTTTAACAAATCATTTGGAAC |                      |
| R5'UTRF | GGTTCCTCACCCTCGGGGAGC               |                      |
| R1F     | ATGGATAATCTCAGTGATACCTTGAA          |                      |
| R11F    | TCAGTGATACCTTGAAGAAGCTG             |                      |
| R108R   | TTGAGCCAGGGCTTGAAGCAGAC             |                      |
| R183R   | CTGTGGAGTCAACAGACTTGCAAAC           |                      |
| R340F   | GGAAACATATGTTACGATAGCC              |                      |
| R655F   | GAGTGTCTACTAGAGATTGTTTCAG           |                      |
| R688R   | CCACTTTTTGCTGAACAATCTC              |                      |
| R1175F  | GAAAGAATGTAAAGTTAGTGGAGC            |                      |
| R1195R  | CTCCACTAACTTAACATTCTTTCCC           |                      |
| R1677R  | TCAGCTTTCCACAGTAAGTCTCTGC           |                      |
| Rint2R  | GTTATGTGCTTAACACATACCAG             |                      |

\* 2.5% DMSO (dimethylsulfoxide, Sigma) was added to the PCR reaction used amplifications of these primers.

#### **2.2.6.4 Restriction endonuclease digestion of PCR products**

PCR products were digested with 5 - 10 units of enzyme in a total volume of 20  $\mu$ l with 1  $\times$  buffer, 1  $\times$  BSA (New England Biolabs) and ultra pure water. PCR products were digested at the appropriate temperature for 3 hours or overnight.

#### **2.2.6.5 PCR product purification**

**A) Direct purification:** PCR product was purified directly from the reaction mix using the Wizard™ PCR Preps DNA Purification System (Promega). The manufacturers instructions for direct purification were followed except that the step utilising direct purification buffer was omitted. For purification of fragments larger than 3 kilobases, fragments were eluted from the column at 65-80°C as specified by the manufacturer (Promega Protocols and Applications Guide, 3<sup>rd</sup> Edition).

**B) Purification from a low melting point agarose gel:** A sufficient quantity (at least 3.5 times the desired final yield) of PCR product or digested plasmid was electrophoresed through a low melting point agarose gel. The gel was made using the desired amount of low melting point agarose (Promega) in 1  $\times$  TAE and electrophoresis was performed in 1  $\times$  TAE buffer at 70 V. Ethidium bromide was added to the gel and buffer at a final concentration of 0.25  $\mu$ g/ml. DNA was loaded into as few lanes as possible to minimise the amount of agarose carried into the purification mixture. The band of interest was visualised using 300 nm UV light and excised with a sterile scalpel blade. The agarose slice containing the desired DNA was transferred to a 1.5 ml microcentrifuge tube and placed at 65°C for 5-10 min to melt the agarose. Once the agarose slice had melted, Wizard purification was performed immediately, according to the manufacturers instructions for the purification of PCR products from agarose gels.

#### **2.2.6.6 Reverse transcription (RT)- PCR conditions using MMLV**

For standard RT-PCR a primer mix consisting of 1 µg of RNA, 2 µl of random hexamers (250 ng/µl) and enough ultra pure water to make a final volume of 20 µl was added to a 0.5 ml tube. This was denatured at 70°C for 10 min and then placed on ice. After 3 min, an RT mix containing 8 µl of 5 × first strand cDNA synthesis buffer (Gibco, Life Technologies), 4 µl of 0.1 M DTT, 2 µl of 40 mM dNTPs, 5 µl of ultra pure water and 1 µl of 200 U/µl MMLV-Reverse transcriptase (Gibco, Life Technologies) was added (on ice). A negative RT mix without the MMLV enzyme was always included as a negative control. The mixtures were gently flicked and briefly spun before incubation at 25°C for 10 min. Reverse transcription was allowed to proceed by incubation at 37°C for 1-2 h and the enzyme inactivated by incubation at 70°C for 10 min. From each newly synthesised cDNA mixture, 2 µl was seeded into each subsequent PCR reaction. cDNA samples were covered with parafilm and kept at 4°C.

#### **2.2.6.7 Reverse transcription (RT)- PCR conditions using Superscript II**

RT using Superscript II (Gibco, Life Technologies), used for generating cDNA for 3' RACE and amplifying open reading frames, was performed according to the manufacturer's "First Strand cDNA Synthesis of Transcripts with High GC Content" protocol contained within the 3' RACE System for Rapid Amplification of cDNA Ends manual (Gibco, Life Technologies). All reagents, with the exception of ultra pure water, were supplied by the manufacturer. All incubation steps were performed using a PTC100 thermal cycler. Briefly, 1 µg of total RNA was brought to a final volume of 24 µl in a 0.5 ml tube using ultra pure water. After the addition of 1 µl of adapter primer (AP), the solution was mixed gently and collected by brief centrifugation. Each sample was then heated to 70°C for ten minutes then transferred immediately to 50°C. A reaction mix consisting of the following was then prepared for each sample; 7.5 µl of ultra pure water, 5 µl of 10 × PCR buffer, 5 µl of 25 mM MgCl<sub>2</sub>, 2.5 µl of

10 mM dNTP mix and 5  $\mu$ l of 0.1 M DTT. This was prepared as a master mix and prewarmed to 42°C. To each RNA sample, 25  $\mu$ l of the above mix was added and this was immediately followed by the addition of 1  $\mu$ l (200 units) of Superscript II. Incubation at 50°C was continued for 50 min. The reaction was terminated by incubating the tubes at 70°C for 15 min. Finally, 1  $\mu$ l of RNase H was added and the tubes were incubated for 20 min at 37°C. One twentieth of the reverse transcription was used for PCR.

#### **2.2.6.8 DNase I treatment of RNA**

Where DNA contamination could lead to misleading results the RNA was treated with RNase free DNase I (Gibco, Life Technologies) prior to reverse transcription. For every 1  $\mu$ g of RNA, 1  $\times$  DNase I reaction buffer (Gibco, Life Technologies), 1U DNase I Amplification Grade (Gibco, Life Technologies) enzyme and ultra pure water to a final volume of 10  $\mu$ l was added. The mixture was incubated at 37°C for 12 min before the addition of 1  $\mu$ l of 25 mM EDTA and incubation at 65°C for 10 min. The 10  $\mu$ l mix was added directly to the primer mix for RT.

#### **2.2.7 Procedure for 3' RACE**

One  $\mu$ g aliquots of PBMNC total RNA were reverse transcribed using Superscript II and the Adapter Primer (AP) from the 3' RACE kit (Gibco, Life Technologies, Gaithersburg, MD). The Expand Long Template PCR System (Roche, Mannheim, Germany) was used in all subsequent PCR amplifications. The reverse transcription product was amplified with a *NUP98* exon B primer, N1459F and the Abridged Universal Amplification Primer (AUAP) (Gibco, Life Technologies). Touchdown PCR was performed with an initial step at 94°C for 2 min, followed by 10 cycles of 95°C for 30 sec, 70°C minus 1°C per cycle for 30 sec and 68°C for 8 min, followed by 25 cycles of 95°C for 30 sec, 60°C for 30 sec and 68°C for 8 min plus

20 sec per cycle. A biotinylated *NUP98* exon B oligonucleotide, N1491F, internal to the oligonucleotide used in the first round, was used to enrich for *NUP98* containing sequences using streptavidin coated magnetic beads (Promega, Madison, WI) according to the protocol of Lankiewicz *et al.* (1997). Second round PCR of the enriched product was performed using a *NUP98* exon B primer, N1511F (internal to the previous two sense oligonucleotides), and the AUAP with cycling conditions identical to the first round PCR. Second round PCR products were electrophoresed through low melting point agarose gels, purified using Wizard PCR Purification kits (Promega), cloned into pGEM-T (Promega) and sequenced.

## **Chapter 3**

### **Investigation of the candidate gene, *ZNF195***

## Chapter 3

### Investigation of the candidate gene, *ZNF195*

As described in chapter 1, the chromosome 11 breakpoint in the t(4;11) patient had been mapped proximal to *D11S470* by FISH (Dobrovic *et al.*, 1994). Preliminary PCR results suggested that the breakpoint was distal to *D11S860* (Finch and Dobrovic, unpublished data). At that time, there were no sequenced genes between these two markers, which made it difficult to test candidate genes for their involvement in the translocation.

Z104 was one of a number of cosmids containing *Krüppel* related zinc finger coding sequences identified by screening a human chromosome 11 library, with a degenerate oligonucleotide probe based on the 7 amino acid *Krüppel* H/C linker region (Hoovers *et al.*, 1992); PFR Little, unpublished data). Z104 was mapped to 11p15.5 on a 140 kilobase (kb) *Bss* HII fragment between *RRM1* and *IGF2* (Redeker *et al.*, 1994). The *D11S470* and *D11S459* markers also mapped to this fragment, but the relative order of the three markers was undetermined (Redeker *et al.*, 1994). Since the chromosome 11 breakpoint in the t(4;11) patient was known to lie between *D11S860* and *D11S470*, and the relative order of *D11S470* and Z104 was unknown, any zinc finger gene in Z104 was a candidate breakpoint gene.

Zinc finger proteins are an important class of eukaryotic DNA binding proteins. The zinc finger motif contains spatially conserved cysteines and histidines, which bind a zinc ion. The most common zinc finger motif is the C<sub>2</sub>H<sub>2</sub> type, which was initially found in the TFIIIA transcription factor of *Xenopus* (Miller *et al.*, 1985), and subsequently in the *Drosophila* segmentation gene, *Krüppel* (Rosenberg *et al.*, 1996). The human genome has several hundred *Krüppel* related zinc finger genes in which the consensus zinc finger sequences CX<sub>2</sub>CX<sub>3</sub>FX<sub>5</sub>LX<sub>2</sub>HX<sub>3</sub>H are connected by the H/C linker, a characteristic 7 amino acid



domain, TGEKPYK (Bellefroid *et al.*, 1989). About one-third of these human *Krüppel* type genes also code for a highly conserved region, the Krüppel-associated box (KRAB domain), comprising approximately 75 amino acids found at the N-terminal end (Bellefroid *et al.*, 1991). The KRAB domain is divided into A and B boxes, usually coded for by separate exons, (Bellefroid *et al.*, 1993; Bellefroid *et al.*, 1991) and constitutes a potent DNA binding dependent repression domain (Margolin *et al.*, 1994; Witzgall *et al.*, 1994).

The expression of some *Krüppel* family zinc finger genes is disrupted, either through gene fusion or deregulated expression, in chromosome translocations associated with haematological malignancy. For example, the t(11;17) associated with acute promyelocytic leukaemia, fuses the retinoic acid receptor gene (RARA) on chromosome 17 with the *Krüppel* family *PLZF* (Promyelocytic Leukaemia Zinc Finger) gene on chromosome 11 and results in a hybrid protein containing zinc finger repeats (Chen *et al.*, 1993). Expression of the *Krüppel* family *BCL6* gene is deregulated in t(3;14)(q27;q32) associated non-Hodgkins lymphoma, because this translocation removes *BCL6* from its normal regulatory sequences and juxtaposes it to the IgH regulatory elements on chromosome 14 (Ye *et al.*, 1993). The established role of *Krüppel* like zinc finger genes in haematological malignancy determined the need to investigate the gene(s) within Z104 as a candidate for disruption in the (4;11) translocation.

In addition, Z104 was in the candidate region for the putative *WT2* (Wilms' tumour 2) gene involved in the development of Wilms' tumour (Besnard-Guerin *et al.*, 1996). The *WT1* gene, which is mutated in hereditary Wilms' tumour, is a *Krüppel* type zinc finger gene (Call *et al.*, 1990). These two points taken together gave reason for studying the gene(s) within Z104 as a candidate for *WT2*. Such studies were conducted during the course of this PhD but they are outside the scope of this thesis and will only be discussed briefly.

The Z104 cosmid had been shown to hybridise only to the der(4) chromosome of the t(4;11) patient and not to the der(11) chromosome, suggesting it was distal to the breakpoint (Finch, Peters and Dobrovic, unpublished). However, this FISH result was not sufficient evidence to exclude the possibility of disruption of any zinc finger gene(s) within Z104 by the (4;11) translocation. It was possible that a portion of the zinc finger gene was absent from Z104 and that this part was retained on der(11). Furthermore, even if Z104 contained the entire gene, it was possible that the breakpoint was near the end of the cosmid insert. If this were the case, FISH using Z104 would not detect the small amount of cosmid sequence remaining on der(11).

This chapter describes the exploration of the zinc finger gene(s) within Z104 as a candidate breakpoint gene in the (4;11) translocation. Exploration of Z104 involved characterising the gene(s) within the cosmid and determining the relative position of this gene(s) in the cosmid insert.

### **3.1 Confirmation of the 11p15.5 breakpoint region**

Before beginning the study on Z104, it was decided that the previous FISH and PCR results on the 11p15 breakpoint region should be confirmed by PCR of der(4) and der(11) DNA from t(4;11). Previous FISH results mapped the *D11S470* cosmid distal to the 11p15 breakpoint (Dobrovic *et al.*, 1994). PCR of somatic cell hybrid DNA showed that the der(4) hybrid DNA was negative for *D11S860* (Finch and Dobrovic, unpublished), but it was necessary to show that *D11S860* was present on the der(11) chromosome in order to confirm that it was proximal to the breakpoint.

DNA from somatic cell hybrids segregating the translocation chromosomes was available in the laboratory (Kalatzis *et al.*, 1993). Primers were made to amplify the *D11S860* and

*D11S470* loci (see Table 2.1 for the sequences of all oligonucleotides mentioned in this thesis) in order to test der(4) and der(11) DNA for the retention of these markers. PCR amplification (Section 2.2.6.1) of hybrid DNA revealed that only the der(4) chromosome was positive for *D11S470* (Fig. 3.1a). In contrast, only the der(11) DNA was positive for *D11S860* (Fig. 3.1b). These results confirmed that the 11p15 breakpoint in t(4;11) was between *D11S860* and *D11S470* and validated Z104 as a candidate gene marker.

### **3.2 Subcloning of zinc finger containing sequences from cosmid Z104**

The first step in isolating the sequence of the zinc finger gene(s) from Z104 was to subclone and sequence restriction fragments that contained DNA sequences coding for zinc fingers. In order to identify zinc finger coding sequence within Z104, cosmid DNA was digested with a range of restriction enzymes (Fig. 3.2a). A Southern blot of the restricted cosmid DNA samples was hybridised to an end-labelled oligonucleotide probe (Section 2.2.4) identical to the H/C linker degenerate oligonucleotide probe used to identify Z104 (Fig. 3.2b). Restriction fragments were chosen for sequencing on the basis of the following criteria; they had to be fairly small (preferably less than 2 kb) for ease of subcloning into pUC19 and sequencing, and they had to hybridise relatively strongly to the H/C linker probe. On the basis of these criteria three restriction fragments (a 336 base pair (bp) *Eco* RI fragment, a 556 bp *Eco* RI-*Xba* I fragment and a 1,503 bp *Eco* RI-*Xba* I fragment) were cloned into appropriately digested pUC19 and sequenced using the pUC/M13F and pUC/M13R universal primers (Sections 2.2.1.5 - 2.2.1.8).

The sequence of these restriction fragments was analysed using BLAST N and BLAST X searches and all three fragments had highly significant homology to zinc finger regions from

**Figure 3.1 PCR analysis of 11p15.5 markers on the der(4) and der(11) containing somatic cell hybrid DNA**

PCR amplification of (a) *D11S470*, (b) *D11S860* and (c) *Z104* markers.

**Lane 1** pUC19/*Hpa* II DNA size marker

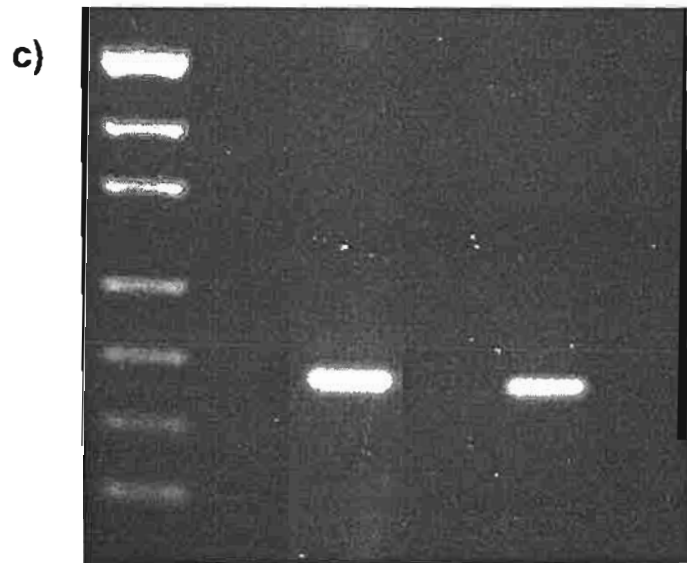
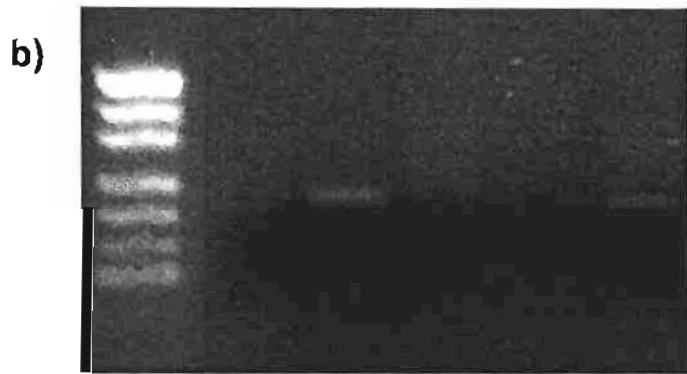
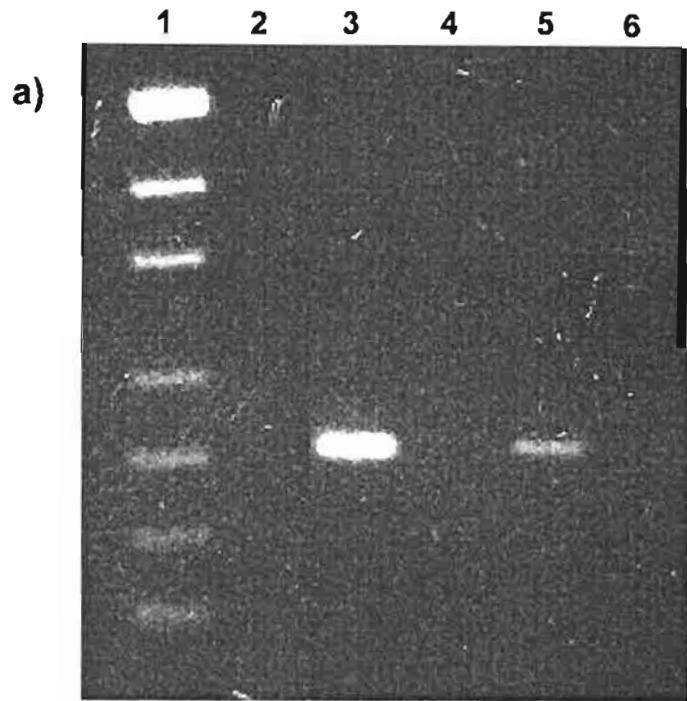
**Lane 2** H<sub>2</sub>O control

**Lane 3** Human genomic DNA control

**Lane 4** Mouse genomic DNA control

**Lane 5** der(4) hybrid DNA

**Lane 6** der(11) hybrid DNA



### Figure 3.2 Selection of Z104 fragments for cloning and sequencing

(a) Agarose gel showing Z104 restriction fragments prior to Southern blotting

(b) Autoradiograph resulting from hybridisation of the H/C linker oligonucleotide probe to the Southern blot of Z104 restriction fragments. Fragments that were subcloned and sequenced are boxed. A partially digested *Eco* RI fragment, that consists of the two fragments below it, is indicated with an arrow.

**M1, M2 and M3** are  $\lambda$ /*Hind* III, pUC19/*Hpa* II and SPPI/*Eco* RI DNA size markers

Lanes 1-14 contain Z104 DNA digested with -

**Lane 1** *Xba* I

**Lane 2** *Eco* RI

**Lane 3** *Hind* III

**Lane 4** *Pst* I

**Lane 5** *Xba* I-*Eco* RI

**Lane 6** *Xba* I-*Hind* III

**Lane 7** *Xba* I-*Pst* I

**Lane 8** *Eco* RI-*Hind* III

**Lane 9** *Eco* RI-*Pst* I

**Lane 10** *Hind* III-*Pst* I

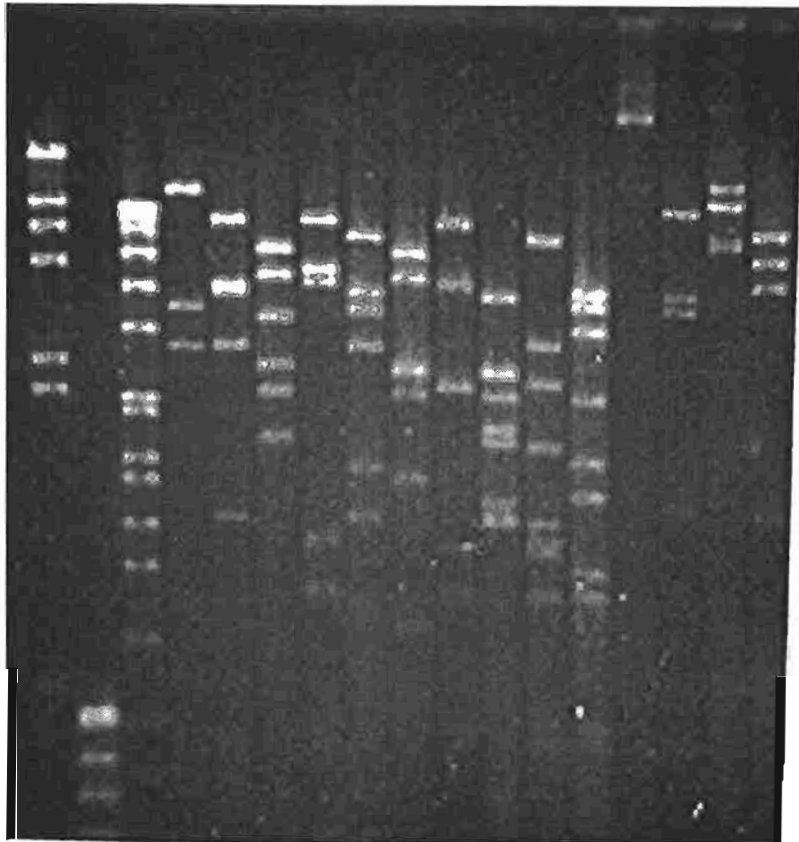
**Lane 11** *Sal* I

**Lane 12** *Sma* I

**Lane 13** *Nsi* I

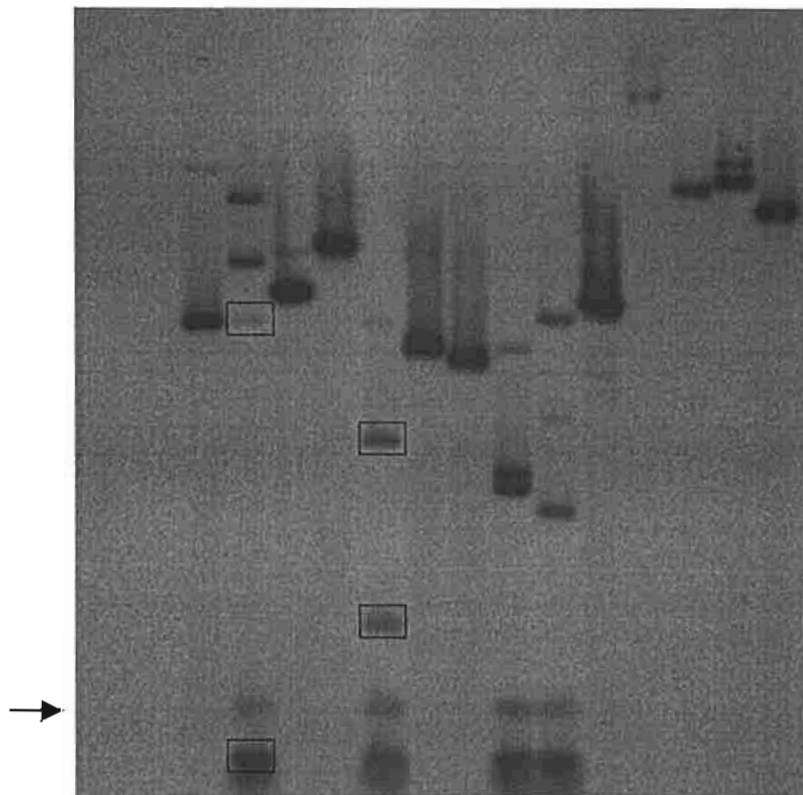
**Lane 14** *Nar* I

a)



M1 M2 M3 1 2 3 4 5 6 7 8 9 10 11 12 13 14

b)



many zinc finger genes. The restriction fragments were thus part of the zinc finger gene(s) present within Z104.

It was not known how the sequence obtained from the three restriction fragments was positioned in relation to the open reading frame of the gene in Z104. A BLAST X analysis showed that the sequence at one end of the 1503 bp *Xba* I-*Eco* RI fragment did not encode zinc fingers or any protein with significant similarity to other known proteins. This suggested that the 1503 *Xba* I-*Eco* RI fragment contained some intronic sequence.

### **3.3 Are the zinc finger coding sequences present within Z104 transcribed?**

It was unknown whether the zinc finger gene(s) within Z104 was transcribed. It was important to determine the transcriptional status because if Z104 contained a pseudo-gene with no coding potential this would reduce the likelihood of it being the breakpoint gene in the (4;11) translocation.

Many zinc finger genes are expressed at high levels in T-cell lines (Bellefroid *et al.*, 1993). Therefore, expression was examined in several human T-cell lines, using the 336 bp *Eco* RI fragment as a probe on Northern blots. Total RNA from HUT-78, JM, Jurkat and Molt4 T-cell lines was prepared (Section 2.2.1.4) and Northern blotted according to the protocol described in Section 2.2.5. The 336 bp *Eco* RI fragment was radiolabelled (Section 2.4) and hybridised to the Northern blot. Strong expression of a 4.3 kb transcript was seen in all four cell lines examined (Figs. 3.3a-b). Smaller transcripts were also visible and HUT-78 strongly expressed a transcript of approximately 2.5-2.7 kb, visible as a broad band on the autoradiograph (Fig. 3.3b). This result was taken as a preliminary confirmation that the gene present in Z104 was actively transcribed in the T-cell lines examined.



**Figure 3.3 Northern analysis of T-cell lines and t(4;11) patient samples using the Z104 336 bp *Eco* RI fragment as a probe**

(a) Expression of Z104 related transcripts in T-cell lines. Each lane contains 10 µg of total RNA as follows-

**Lane 1** Molt4

**Lane 2** JM

**Lane 3** Jurkat

(b) Elevated expression of a Z104 related transcript in t(4;11) leukaemic samples. Each lane contains 10 µg of total RNA as follows-

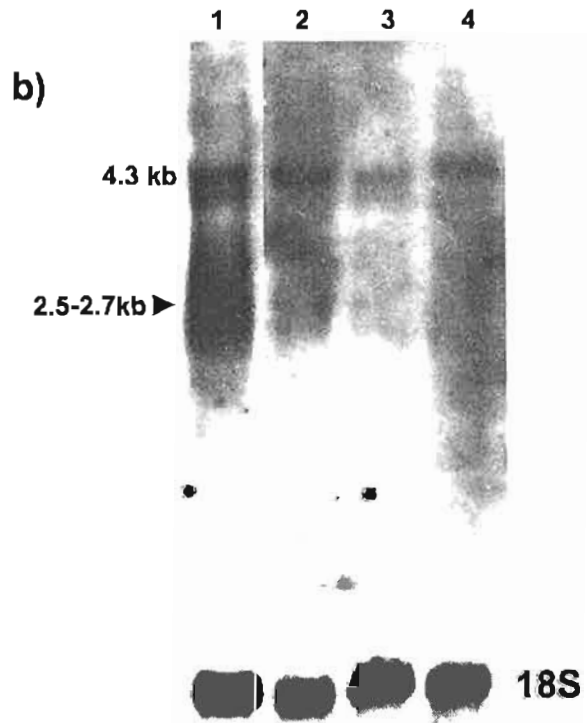
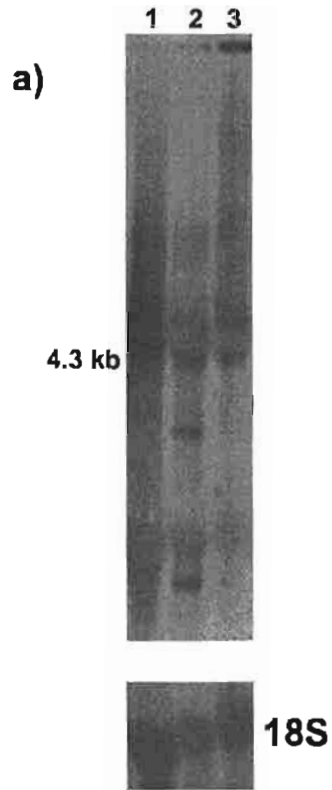
**Lane 1** HUT-78

**Lane 2** t(4;11) presentation PBMNC

**Lane 3** t(4;11) remission bone marrow

**Lane 4** t(4;11) relapse PBMNC

The same membranes probed with an 18S rRNA probe are shown as an RNA loading control.



### **3.3.1 Northern analysis of adult and foetal tissues using the 336 bp *Eco* RI fragment as a probe**

In conjunction with the Northern analysis of T-cell lines, Northern analysis of adult and foetal tissues was also performed using the 336 bp *Eco* RI fragment from Z104 (Fig. 3.4a-c). The probe detected moderate to strong expression in adult heart, brain, placenta, skeletal muscle and pancreas with a predominant transcript size of 4.3 kb. There was little detectable expression in adult lung, liver or kidney (Fig. 3.4b). In the remainder of the adult tissues tested the predominant transcript size was 2.5 kb, except in PBMNC the most prominent transcript was 4.3 kb (Fig. 3.4c). In foetal lung, liver, kidney and brain, the predominant transcript was 2.5 kb (Fig. 3.4a). Foetal brain also expressed a 2.7 kb transcript (The sizes of the transcripts in foetal tissues have been revised since Hussey *et al.* (1997)).

### **3.3.2 Assessing the specificity of the 336 bp *Eco* RI fragment for Z104**

In order to assess the specificity of the 336 bp *Eco* RI fragment as a probe for detecting only the gene within Z104, Southern analysis was performed using this probe. A probe that is specific for a single gene is expected to hybridise to one or a limited number of restriction fragments. Genomic DNA was prepared from presentation, remission and relapse specimens of the t(4;11) patient according to the method described in Section 2.2.1.3. This DNA was digested with *Eco* RI, *Xba* I and *Hind* III and Southern blotted. These samples were chosen in case the probe was in the breakpoint region and was potentially able to detect a rearrangement caused by the (4;11) translocation. The remission sample was known to have a normal karyotype and therefore it was assumed that it would result in a normal pattern of hybridisation. The radiolabelled (Section 2.2.4) 336 bp *Eco* RI fragment was hybridised to the Southern blot and then washed at the same high stringency used for the Northern analysis described above.

**Figure 3.4 Expression of Z104 related and ZNF195 transcripts in adult and foetal tissues**

**(a)/(d)** Foetal tissue Northern blot (Clontech, #7756-1) hybridised to **(a)** the Z104 336 bp *Eco* RI fragment or **(d)** a 367 bp PCR product from the *ZNF195* 3'UTR. Each lane contains 2µg of polyA RNA as follows: **Lane 1** brain, **Lane 2** lung, **Lane 3** liver, **Lane 4** kidney.

**(b)/(e)** Adult tissue Northern blot (Clontech, #7760-1) hybridised to **(b)** the Z104 336 bp *Eco* RI fragment or **(e)** a 367 bp PCR product from the *ZNF195* 3'UTR. Each lane contains 2µg of polyA RNA as follows: **Lane 1** heart, **Lane 2** brain, **Lane 3** placenta, **Lane 4** lung, **Lane 5** liver, **Lane 6** skeletal muscle, **Lane 7** kidney, **Lane 8** pancreas.

**(c)/(f)** Adult tissue Northern blot (Clontech, #7759-1) hybridised to **(c)** the Z104 336 bp *Eco* RI fragment or **(f)** a 367 bp PCR product from the *ZNF195* 3'UTR. Each lane contains 2µg of polyA RNA as follows: **Lane 1** spleen, **Lane 2** thymus, **Lane 3** prostate, **Lane 4** testis, **Lane 5** ovary, **Lane 6** small intestine, **Lane 7** colon (mucosal lining), **Lane 8** PBMNC.

The same membranes probed with *β-actin* (Clontech) are shown as an RNA loading control.



The 336 bp *Eco* RI fragment hybridised most strongly to a single band for all three enzymes used (*Xba* I and *Hind* III digests are shown in Fig. 3.5). A few minor bands were also detected, but they hybridised poorly in comparison to the predominant band. The results of Southern analysis suggested that the strong band seen on Northern blots corresponded to the expression of *ZNF195* rather than a related gene. The presence of weaker transcripts was thought to represent either cross hybridisation to other zinc finger transcripts or rare alternatively or partially spliced products.

No novel *Eco* RI, *Xba* I or *Hind* III restriction fragments were detected in the presentation or relapse DNA compared to the remission DNA. The largest restriction fragment that hybridised to the 336 bp *Eco* RI fragment was a 3 kb *Hind* III band. The 3 kb *Hind* III band was seen in presentation, remission and relapse DNA from the t(4;11) patient. This indicated that the t(4;11) breakpoint was not within 3 kb of the 336 bp *Eco* RI fragment from Z104.

To further test the specificity of the 336 bp *Eco* RI fragment for the gene present within Z104, oligonucleotide primers were designed (Section 2.2.6.3) to PCR amplify a sequence within this fragment. Oligonucleotide primers, called *ZfEcoF* and *ZfEcoR*, were designed so that they spanned the least conserved regions of the sequence (with respect to the corresponding protein sequence). This was done to minimise cross amplification of other zinc finger genes. PCR using these primers resulted in amplification of a 187 bp product from der(4) DNA from the t(4;11) patient and not from der(11) DNA (Fig. 3.1c). The sequence of this product was identical to the corresponding sequence in Z104 (data not shown). This was in agreement with the FISH result which mapped Z104 to the der(4) chromosome (Finch, Peters and Dobrovic, unpublished), and indicated that the sequence of the 336 bp *Eco* RI fragment was specific to Z104.

**Figure 3.5 Southern analysis using the Z104 336 bp *Eco* RI fragment as a probe**

Each lane contains genomic DNA that was isolated from PBMNC, except lanes 2 and 5 which contain genomic DNA isolated from bone marrow. Approximately 10 µg of restricted genomic DNA is loaded in each lane except lane 4 which is overloaded.

**Lane 1** t(4;11) presentation/*Hind* III

**Lane 2** t(4;11) remission/*Hind* III

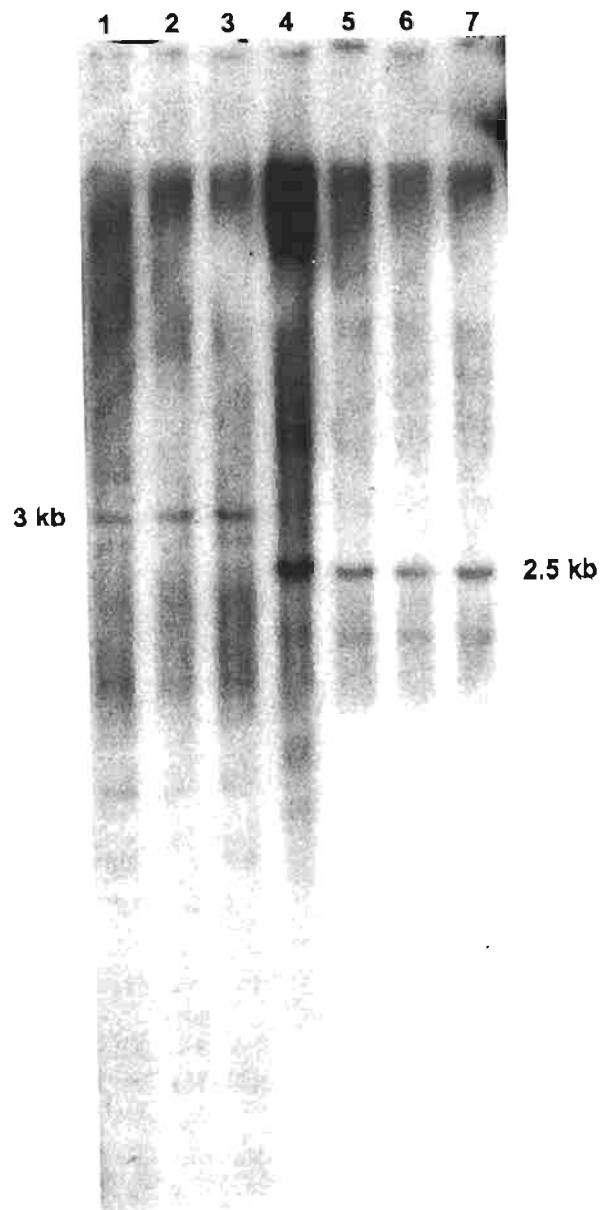
**Lane 3** t(4;11) relapse/*Hind* III

**Lane 4** normal donor DNA/*Xba* I

**Lane 5** t(4;11) presentation/*Xba* I

**Lane 6** t(4;11) remission/*Xba* I

**Lane 7** t(4;11) relapse/*Xba* I





### **3.3.3 Elevated expression of a 4.3 kb transcript in the t(4;11) patient**

Northern analysis of presentation, remission and relapse RNA from the t(4;11) patient was performed using the 336 bp *Eco* RI restriction fragment from Z104. In all three specimens a transcript of 4.3 kb was observed. However the expression of the 4.3 kb transcript was dramatically elevated in the presentation and relapse samples compared to the remission sample (Fig. 3.3b). This result gave preliminary evidence that expression of the gene in Z104 corresponding to this fragment was up-regulated by the (4;11) translocation. Up-regulation of this gene in the leukaemic specimens from the t(4;11) patient suggested that this gene might be affected by the translocation, either because the translocation deregulated expression of the gene by removing it from its normal 5' regulatory elements or because the translocation placed the gene in a more transcriptionally active chromatin environment. Therefore further investigation of the gene within Z104 was undertaken.

### **3.4 Isolation of lambda phage clones from a HUT78 cDNA library**

The 336 bp *Eco* RI restriction fragment from Z104 was used to screen a cDNA library in order to identify homologous cDNA clones. Since Northern analysis using the 336 bp *Eco* RI probe detected expression in HUT-78 cells, a  $\lambda$ gt11 HUT-78 cDNA library was chosen for screening (a gift from F. Shannon, Hanson Centre for Cancer Research, Adelaide). A total of 150,000 phage plaques were screened using the protocol described in Section 2.2.3.

Five phage plaques which hybridised strongly to the 336 bp *Eco* RI fragment were picked from plates, using a sterile Pasteur pipette, and eluted into SM buffer. The insert sequence of these clones was obtained without clone purification or DNA preparation. This was achieved using a PCR strategy with outward primers near the ends of the 336 bp *Eco* RI fragment (Zfseq out5' and Zfseq out3') and inward primers near the ends of the vector arms (gt11f and gt11r). This strategy also eliminated the unnecessary event of continuously sequencing over

the region containing the 336 bp *Eco* RI fragment. Since the orientation of the inserts was unknown, various PCRs were performed on the phage eluates of each clone using four different combinations of Zfseq with *gt11* primers (Fig. 3.6a). By restriction mapping the PCR products from each clone, it was possible to determine that two sets of two clones were identical. Furthermore, analysis of the restriction fragment sizes of the PCR products showed that one set of duplicate phage clones contained only a portion of the sequence present in the other set (data not shown). Therefore only 2 non-redundant clones had been isolated and the PCR products from these clones were sequenced in both directions using the appropriate PCR primers. Using the Sequencher program (Genecodes), the clones were placed relative to the 336 bp *Eco* RI fragment by aligning the sequence at the ends of the PCR products with sequence of the 336 bp *Eco* RI fragment (Fig. 3.6b).

### **3.5 Assembly of restriction fragment and phage sequence into a contiguous sequence**

Using Sequencher, the cDNA sequence from the HUT78 clones and the genomic sequence from the Z104 restriction fragments was assembled into a continuous stretch of 1800 nucleotides (nt). Figure 3.6b shows the position of these fragments relative to the cDNA sequence. Analysis of the open reading frame encoding the zinc fingers revealed a stop codon near the end of the 1800 bp contig. BLAST N analysis clearly showed that the contiguous sequence encoded a *Krüppel* type zinc finger gene with a Krüppel associated box. However the sequence corresponding to the 5' end of the gene did not encode a complete KRAB A box so it was concluded that the 5' end of the open reading frame was not contained in the 1800 bp contig.

A search of the expressed sequence tag (EST) database, at the time prior to submission of the manuscript (Hussey *et al.*, 1997), found 10 ESTs (Genbank AA092626, H14414, AA214696,

Z21107, AA076589, T61543, H14367, AA256440, Z41948, AA256315) with identity to portions of the 1800 bp contig (Fig. 3.6b). Two of these overlapped with the then furthestmost 5' sequence of the gene and extended the sequence 5' past the putative start codon. This sequence was confirmed by RT-PCR of HUT-78 RNA using one primer from the ESTs and another primer from the spacer domain. The partial cDNA sequence, including the complete open reading frame of the gene, is presented in Figure 3.7.

The sequence of the 1,503 bp *Eco* RI-*Xba* I genomic fragment allowed the determination of three intron/exon boundaries (Fig. 3.6b). The remaining boundaries were subsequently determined by comparison with the sequence of a P1 artificial chromosome (PAC) containing the gene from Z104 (see Section 3.6.1).

### **3.6 Description of the *ZNF195* gene present within Z104**

The sequence of the open reading frame and partial 5' and 3' untranslated regions (UTR) of the *Krüppel* zinc finger gene within the Z104 was deposited into Genbank (Accession no. AF003540). After correspondence with the Human Genome Nomenclature Committee, the gene was named *ZNF195* because it was the 195<sup>th</sup> sequenced human *Krüppel* zinc finger gene. The 1890 nt open reading frame of *ZNF195* comprises 6 exons (1, 2, 3, 4a, 4b, and 5) and is theoretically translated to give a protein of 629 amino acids. Exons 4a and 4b, located between the KRAB B box and the spacer region, are alternatively spliced (see Section 3.7).

The putative protein contains all of the motifs common to KRAB containing zinc finger genes (Bellefroid *et al.*, 1991) including a KRAB A domain (exon 2) located immediately after a methionine at the end of exon 1, a KRAB B domain (exon 3), a spacer region and a zinc finger domain (exon 5). A BLAST X analysis of the spacer region shows that *ZNF195* is most closely related to the *ZNF91* family of zinc finger genes (72% amino acid similarity)

### Figure 3.6a Method used PCR amplify insert DNA from $\lambda$ gt11 HUT-78 cDNA clones

Shown is a schematic representation of the method used to PCR amplify cDNA sequence flanking the 336 bp *Eco* RI fragment from eluates of  $\lambda$ gt11 HUT-78 cDNA plaques. Primers are represented by half arrows, with their names listed above the arrowhead. The  $\lambda$ gt11 vector is represented by a blue line and the insert is represented by a thick red line.

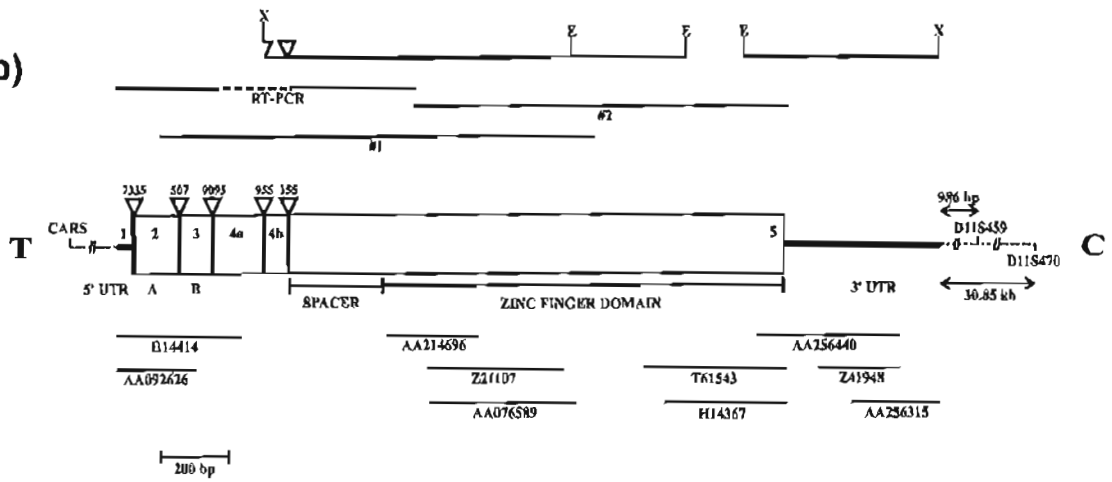
### Figure 3.6b *ZNF195* Sequencing strategy

The positions of the genomic restriction fragments (*Eco* RI and *Xba* I restriction endonuclease sites are denoted by E and X respectively) and the two HUT-78 cDNA clones (marked #1 and #2) are shown above the cDNA. The coding region is represented by rectangles with the positions of introns marked by inverted triangles and numerals indicating the intron size in base pairs above the triangles. The RT-PCR product sequenced is also shown. The dashes indicate sequence absent from the RT-PCR product. The 5' and 3' UTR sequence is not complete. The positions of the ten ESTs and their Genbank accession numbers are shown. The distance between *ZNF195* and proximal markers is shown but the distance between *ZNF195* and *CARS* is not known at present. C and T represent the centromeric and telomeric ends of the fragment respectively.

a)



b)



described elsewhere (Bellefroid *et al.*, 1993). Although the methionine before the KRAB A domain is not within a consensus Kozak sequence GCCA/GCCATGG (reviewed in Kozak, 1991), it must be the start codon as it is preceded by an in frame nonsense codon (Fig. 3.7).

Znf195 contains 14 Krüppel related zinc finger motifs. The first consensus zinc finger in Znf195 is separated from the other eight by five degenerate fingers in which at least one of the cysteine or histidine residues is absent (Fig. 3.7). A similar arrangement was reported for *ZNF43*, a *ZNF91* family zinc finger gene, where three degenerate fingers separate the first consensus motif from the others (Lovering and Trowsdale, 1991). The second and fourth degenerate fingers of Znf195 are twenty rather than twenty one amino acids long.

The *ZNF195* termination codon occurs within the H/C linker immediately following the ninth consensus zinc finger. Another 4 highly degenerate motifs with numerous deletions follow the termination codon. This indicates that *ZNF195* evolved from a gene with a greater number of zinc finger repeats.

A nearly perfect (6/7) immediate response box (IRB) is present within the 3' UTR. The IRB is a putative *cis*-acting control element which mediates induction of the immediate-early gene set and has also been found in other zinc finger genes (Freter *et al.*, 1992).

### **3.6.1 Determination of the intron/exon boundaries of *ZNF195***

Some time after the characterisation of the *ZNF195* open reading frame, the sequence for a PAC containing 11p15.5 genomic sequence was deposited into Genbank (Accession no. AC000378). BLAST N analysis of the *ZNF195* cDNA sequence with this PAC showed that the PAC contained all of the *ZNF195* cDNA sequence. Alignment of the sequence from the

**Figure 3.7 The mRNA and putative amino acid sequence of ZNF195 (Genbank Accession AF003540)**

Nucleotide numbers are given above the sequence. Amino acid numbers are given on the right. This sequence includes the variably spliced exons 4a (nucleotides 272-418) and 4b (nucleotides 419-486). The KRAB A domain is underlined with a single dashed line, and the KRAB B domain is underlined a double line. Arrowheads denote the positions of introns. The first cysteine of each of the 9 consensus zinc finger motifs is circled and numbered. The 9 consensus fingers generally fit the Krüppel consensus CX<sub>2</sub>CX<sub>3</sub>FX<sub>5</sub>LX<sub>2</sub>HX<sub>3</sub>H with well conserved H/C linker regions. Finger 3 substitutes a tyrosine for the phenylalanine normally at the eighth amino acid and finger 7 substitutes an isoleucine for the leucine normally at the fourteenth amino acid of the Krüppel zinc finger consensus. The 52 bp sequence not present in the PAC sequence is underlined in bold. The positions of the *Eco* RI restriction sites that give rise to the 336 bp probe are shown. The sequence similar to an immediate response box (IRB) is underlined with a cross (†).

CGGGAGATCCAGAAAGTGAACCCAGGCTCTCTGAGGCCAGGAGATGACTCTTTGACGTTGAGGATGCGCCATAGAAATTCCTCCGAGGAGTGGAAATGCCTGGACCTCGCTCAG 25  
M T L L T F R D V A I E F S L E E W K C L D L A Q  
-----  
CAGAAATTTGTACAGGATGTGATGTTGGAGAACTACAGAACTTGTCTCCGTTGTTCTCACTGTCTOTAAAGCCAGGCTGTATCACCTGCTGAGCAACGAAAGAGCCCTGGAATGTTG 65  
Q W L Y R D V M L E N Y R M L F S V G L T V C K P G L I T C L E Q R K E P W N V  
-----  
AAGAGACAGGAGGACAGACCGACATCCAGAGATGSGGATTCACCATGTCTACTCAGGCTTGTCTTGAACCTCTGGGCTCAAGCAATCTGCTGCTCAGCCTCCCAAAAGTCTGGGAT 105  
K R Q E A A D G H P E M G F H H A T Q A C L E L L G S S D L P A S A S Q S A G I  
-----  
ACAGGTGTAAACCACCTGTCGCCAGCCAGCCCTCAATGTTTCTGTGGCAAAATCACTGTGCCTCACCTGGGCTGCTGCAACTGTGAAATGTTTTAGAGTTTCAAGTGCATCTTC 145  
T G V M B R A Q P G L N V S V D K F T A L C S P G V L Q T V R W F L E F R C I Y  
-----  
TCCTAGCTATGCTCTCTCATTTTACCCAAAGACCTTCTGCCAGAGCAAGGCTACAAAGATGCAATCCCAAAAGAAATACTGAGAGGATATGGAAATTTGGCCCTTGAATAATTTATATTA 185  
S L A M S S B F T Q D L L P E Q G I Q D A F P K R I L R G Y G N C G L D M L Y L  
-----  
AGAAAGACTGGGAAAGTTTAGATGAGTGTAAATGCAAAAGATTAATATGACTTAAACCAATGTTTCACTCACTCCCATAGCAAAATCTTTCAATATAAATATGTTTAAATCTTT 225  
R K D W E S L D E C R L Q R D Y M G L H Q C S S T T H S K I F Q Y N K Y V K I F  
-----  
GATAACTTTTCAAAATTTACATAGACGTAAATAAAGTAAATCTGAGAGAGAAACCTTCAAAATGCAAGAAATGCGCAAAATCTTTCAAAATGCTCTCATCTTCAACTGAACATCAGAAAT 265  
D N F S E L E R R N I S M T G E K P F K Q E C G K S F Q M L S F L T E H Q K I  
-----  
CACACTGGAAAGAAATTTCCAAAAATGTTGGAGAAATGTTGGCAAAACCTTTTACAGTGTACACTTTCAGAACCTTGAACATTTGACACTGGAGAGAAACCTTCAAAATGTTCAAGAAAT 305  
H T G K K P Q K C G E C G K T F I Q C S R F T E P E N I D T G E K P Y K C Q E C  
-----  
AACAACTCATTAATAACTTGTCTCAGTCTTACTTAAATAAATTAATGCGGAGGGGAAACATTACAGATGTAAGAAATTTGGCAAAAGTATTTAAACAGTGTCTCCACCTTACTGAACT 345  
N N V I K T C S V L T K N R I Y A D G K E Y R C E E F G K V F N Q C S H L T E H  
-----  
GAGCATGGTACTGAGGAAACCCCTGCAAAATTAAGAGAGTGCAGCAGTGTCTTATCTCTTCTGCTCAAGCCCTTCTAAATCAACAGATGATCTTGTCTGGAGAGAAAGCTTCCAAATGTGAA 385  
E H O T Y E E E K P C K C S S V F I S C S B L S N Q Q M I L A G E K L S K C E  
-----  
ACATGGTACAAAGGTTTAAACCACAGCCCAAAATCTTCCAAAACCCAGAGAAATGAGATTTGAGGGAAACCTTTCAAAATGTTGAGGAAATGACAGCATCTTCAAGTGTCTTCAAGCTT 425  
T W Y K G P N H S P N P G K R Q R N E I G G K P F E E E C D S I F K W F S D L  
-----  
ACTAAACATAAGAGAAATTTACACTGTGAGAGAAACCTACAAATGTTGAGGAAATGTTGGGAAAGCCCTTATACAGTGTCTTCAACCTCAGTGAACACAGGAGGATTCACACCGGAGAGAAACCC 465  
T K B K R I H T G E K P Y K D E C G R A Y T Q S S B L S E E R R I H T G E K P  
-----  
TACCAATGTGAAGAAAGTGGGAAAGTCTTCAAGCTTCTTAAACCTAAGAGAACTCATTCTGAAGAAACCCCTACACGTTGAAGAAATGTTGCAACATCTTTAAGCAG 505  
Y Q E E C G K V F R T C S S L E H H K R T E S E E K P Y T E E C G M I P K Q  
-----  
TATCAGACCTCACTAAGCAATAAGAAACCCATCTGAGAGAGAAAGCCCTCAAAATGTTGAGGAAATGTTGGGAAACCTTTAACCAGTCTTATGTTACATAAAGAAATTTCACT 545  
L S D L T K H K K T H T G E K P Y K D E C G R N F T Q S S N L I V H K R I B T  
-----  
GGAGAGAAACCCCTACAAATGTTGAAGAAATGTTGCAAGGCTTCAATGTTCTCAGACATTTACAAACATAAGCAAAACCCATCTGAGAGAGAAACCCCTCAAAATGTTGAGGAAATGTTGGCAA 585  
G E K P Y K E E C G R A F M W F S D I T K E K Q T R T G E K P Y K D E C G K  
-----  
AACTTACCCAGTCTCAAACTTATTTACATAAGAGAAATTTACTGAGAGAAACCCCTACAAATGTTGAAGAAATGTTGGCAAGGCTTCAACCAGTCTTCAACCTGACTGTACATGAA 625  
N P T Q S S N L I V H K R I H T G E K P Y K E K C Q K A F T Q F S H L T V H E  
-----  
AGCATTCACTTGAAGAAAGAAATAAATAAATAAAGCAAAAGCCCTTAAATATCTGCTCGCATCCCACTTTTACATCAGAGTTTCAAGTCTTAAATTAAGTATTCATAAAATTTACT 665  
S I B T \*  
-----  
TATTGAAGACCTTTTCATGAAATTAAGTCTCCAGAGCACACAAGAGTATTTCTTCTGAAAAAATGTTACAAATTAACATAGATGGAAGACTTGCATCAGTTGCTTAAACTTTTCAAGAAA 705  
-----  
TTTTAGAGAAACCCCAAACTCTCAAAATGTTGAAATTAACATTTGTTCAAAATGATAGCTTAAAGAAACACTAGAGTTTCACTTAAAGATATTTTGCAATACACTAAATGGAAGAAAT 745  
+++++  
AAGCAAAATCCAAATTTAAGTAAACAATGAGGATTTGAAGGAGAAAGGAAATGAGACCGGACAGCTTCAAGAAATGACACTGAAATCCCTGTGTTGAGTCTACAGAGACTCTAGA 785



PAC with the *ZNF195* cDNA sequence enabled the determination of the positions of the remaining introns within the cDNA sequence of *ZNF195* (Fig. 3.6b)

### 3.6.2 Positioning of *ZNF195* within the Z104 cosmid

Previous FISH of Z104 suggested that *ZNF195* was distal to the chromosome 11 breakpoint in t(4;11) (Finch, Peters and Dobrovic, unpublished). However, it was possible that the 5' end of *ZNF195* was either very close to the end of the Z104 insert, or absent from the Z104 insert. If this were the case then FISH of Z104 may not have detected a potential 5' disruption of *ZNF195* caused by t(4;11). The cloning and sequencing of 3 Z104 restriction fragments (Section 3.2) demonstrated that Z104 contains the *ZNF195* 3'UTR and the zinc finger coding sequence, but did not demonstrate the presence of any exons 5' of exon 4b. It was still possible that the majority of *ZNF195*, including the zinc finger coding region, was removed from the 5' part of the gene by the (4;11) translocation. It was therefore important to determine the relative position of *ZNF195* within the Z104 insert.

During the subcloning of zinc finger containing restriction fragments from Z104, a 2.5 kb *Eco*RI fragment which hybridised weakly to the H/C oligonucleotide probe was also subcloned into pUC19 and sequenced. This fragment was found to contain a small amount of the cosmid vector from Z104 at one end, and therefore contained the junction between the vector and insert sequence of Z104. No *ZNF195* coding sequence was found in the non vector part of the fragment. Presumably the weak hybridisation of the H/C linker oligonucleotide to this fragment was due to the relatively high GC content of both this fragment and the H/C linker.

Despite the lack of *ZNF195* coding sequences in this fragment, its sequence proved useful as the relative position of *ZNF195* within Z104 was able to be deduced. By aligning the non-

vector sequence of the 2.5 kb *Eco* RI fragment from Z104 with the sequence of the 132 kb 11p15.5 PAC containing *ZNF195*, it was possible to determine that one end of the Z104 cosmid insert corresponded to nt 51,000 of the PAC. This end of the cosmid insert was 17,624 nt 5' of the beginning of the *ZNF195* cDNA sequence, ie the *ZNF195* 5'UTR sequence begins nearly 18 kb from the 5' end of the Z104 insert. Therefore Z104 definitely contains all of the *ZNF195* coding sequence, from the 5'UTR (and perhaps the more 5' regulatory sequences) to the 3'UTR. This result, in combination with the FISH results which mapped Z104 distal to the t(4;11) breakpoint (Finch, Peters and Dobrovic, unpublished), indicates that the 5' end of *ZNF195* is not disrupted by t(4;11), and that the entire gene is translocated intact onto the der(4) chromosome.

### **3.7 Alternative splicing of exons 4a and 4b**

The *ZNF195* cDNA sequence was derived from the HUT-78 cell line, which like the other T-cell lines tested, gave a transcript of 4.3 kb on Northern blots. RT-PCR was performed on HUT-78 and Jurkat RNA, using primers from the 5' UTR (*Zf5'UTRF*) and exon 5 (*ZfspacerR*). The PCR resulted in amplification of multiple products which indicated that the exons between the primers were alternatively spliced (Fig. 3.8a). Sequencing of the most abundant RT-PCR product, representing the major transcript, showed that it corresponded to an mRNA that retained the KRAB A and B boxes but lacked exon 4a and 4b. Minor RT-PCR products were also seen, these corresponded in size to alternatively spliced isoforms of *ZNF195*, one of which contained exon 4a, one that contained exon 4b, and one that contained both exons. The exons contained in these minor products were verified by digestion with restriction enzymes that had unique sites within the exons (data not shown). Similar RT-PCR results were obtained from PBMNC from four normal individuals, indicating that the alternative splicing is real and not simply a cell line artefact (Fig. 3.8b).

**Figure 3.8 Alternative splicing of exons 4a and 4b in T-cell lines and normal donors**

RT-PCR products from (a) T-cell lines and (b) PBMNC of normal donors using a forward primer from the *ZNF195* 5'UTR and a reverse primer from the spacer region of *ZNF195* (see text for details). The products containing exons 4a and/or 4b are indicated with a line. The major product, indicated by an arrowhead, does not contain exons 4a and 4b. Lane details are as follows-

**(a) Lane 1** SPPI/*Eco* RI and pUC19/*Hpa* II DNA size markers

**Lane 2** H<sub>2</sub>O control

**Lane 3** Jurkat

**Lane 4** HUT-78

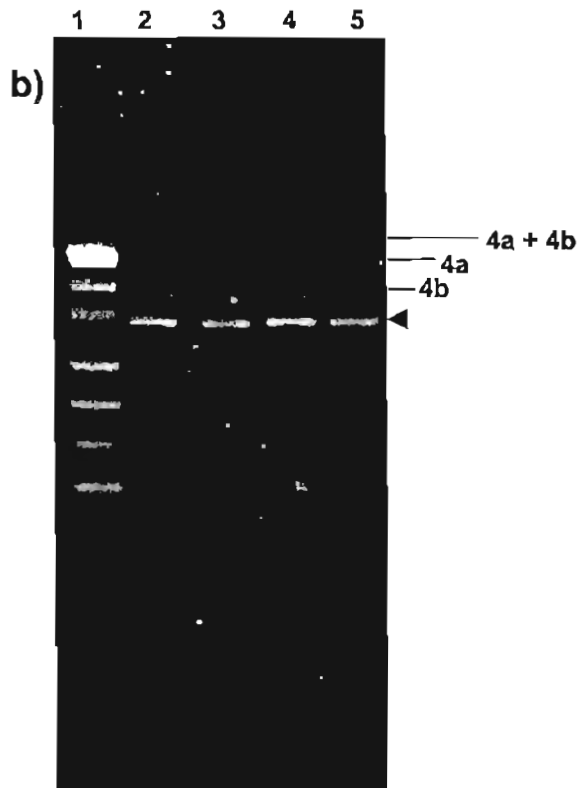
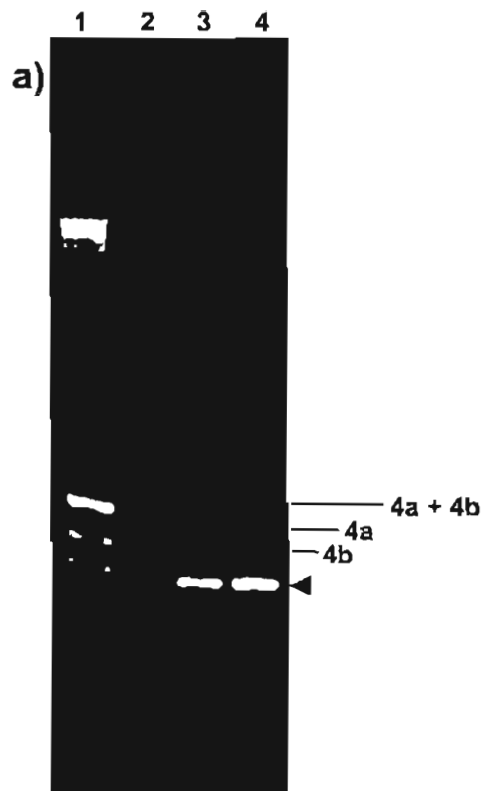
**(b) Lane 1** pUC19/*Hpa* II DNA size marker

**Lane 2** donor 1

**Lane 3** donor 2

**Lane 4** donor 3

**Lane 5** donor 4



Exon 4a consists of the last 115 bp of an inverted Alu repeat sequence and 32 bp of adjacent genomic sequence. Exon 4a was present in a HUT-78 cDNA clone (Fig. 3.6b) and also in an EST from infant brain (Genbank H14414). Several other genes have been reported to have alternatively spliced exons that contain part of an inverted Alu element. These include the proto-oncogenes *c-rel* and *c-myb*, and like exon 4a of *ZNF195* the Alu exons do not disrupt the open reading frame (Brownell *et al.*, 1989); Genbank U22376)

### **3.8 Revised Northern analysis with a probe from the 3'UTR of *ZNF195***

After completing the characterisation of the *ZNF195* open reading frame, it was possible to design a new probe for Northern analysis based on sequence from the *ZNF195* 3'UTR. A probe from the 3'UTR should be more specific for *ZNF195* than a probe from the zinc finger coding sequence (ie the 336 bp *Eco* RI fragment) because the sequence of the 3'UTR is less likely to resemble other zinc finger genes. The rationale for this assumption was the fact that the zinc finger coding sequence of all zinc finger genes is under a high degree of evolutionary selection pressure for DNA sequence conservation. In contrast, the 3'UTR, which consists of non-coding sequence, is not under the same degree of selection pressure and is free to diverge so that it does not resemble other sequences. In support of this, a BLAST N analysis of the *ZNF195* 3'UTR sequence resulted in significant matches only to *ZNF195* (see below).

The Unigene database, maintained by the National Centre of Biotechnology Information (<http://www.ncbi.nlm.nih.gov/UniGene/>), assembles all homologous sequences in Genbank into clusters and then identifies which gene in Genbank is represented by a particular cluster. To obtain 3'UTR sequence, the Unigene cluster for *ZNF195* was assembled into a virtual consensus sequence using Sequencher. This extended the sequence from the Genbank entry for *ZNF195* by 45 bp at the 5'end and 330 bp at the 3'end. Based on this sequence,

oligonucleotide primers (Zf3'UTRF and Zf3'UTRR) were designed to PCR amplify a 367 bp product whose 5' end occurred 239 bp from the *ZNF195* stop codon. BLAST N analysis of the 3'UTR sequence amplified by the primers detected no significant homology to any genomic sequences other than that of the PAC containing *ZNF195*. Furthermore, there was no significant homology to ESTs other than those that were an exact match, suggesting that the probe would detect only the *ZNF195* transcript.

Northern analysis using the *ZNF195* 3'UTR probe on the same adult and foetal blots described previously did not detect a 4.3 kb transcript in any of the tissues examined (Fig. 3.4d-f). The analysis was performed under the same high stringency wash conditions used for Northern analysis with the Z104 336 bp *Eco* RI fragment. This indicated that the 4.3 kb transcript detected by the 336 bp *Eco* RI fragment did not represent *ZNF195*. The 4.3 kb transcript must therefore represent expression of a gene related to *ZNF195*. Since *ZNF195* is most closely related to the *ZNF91* family of zinc finger genes, the 4.3 kb transcript may represent one or more members of this family. In support of this, some members of the *ZNF91* family are known to be expressed with a transcript size of 4.3 kb (Bellefroid *et al.*, 1993; Bellefroid *et al.*, 1991).

The *ZNF195* 3'UTR probe detected a 2.5 kb transcript in all adult and foetal tissues examined (Fig. 3.4). The virtual *ZNF195* mRNA, predicted using the Unigene cluster for *ZNF195*, is 2.7 kb, which is around 200 bp larger than the transcript seen on Northern blots. However, the virtual *ZNF195* mRNA is not representative of the major *ZNF195* transcript because it contains exons 4a and 4b which are absent from the major *ZNF195* transcript (see Section 3.7). Removing the sequence corresponding to exons 4a and 4b from the virtual consensus sequence predicts a major transcript of 2.5 kb, which is in excellent agreement with the 2.5 kb mRNA seen with Northern analysis using the 3'UTR probe.

In addition to the 2.5 kb transcript, foetal brain expresses a 2.7 kb transcript (Fig. 3.4d). This larger mRNA is the same the size as the virtual consensus sequence containing exons 4a and 4b, suggesting that exons 4a and 4b are present in the 2.7 kb transcript. Additional evidence that the 2.7 kb transcript contains at least exon 4a is that a foetal brain EST (Genbank H14414) from the *ZNF195* Unigene cluster contains exon 4a (Fig. 3.6b). The broad band seen detected in HUT-78 RNA using the 336 bp *Eco* RI fragment is in the 2.5-2.7 kb range (Fig. 3.3b) and presumably reflects the presence of transcripts with and without exon 4a (and perhaps exon 4b). The isolation of a HUT-78 cDNA clone containing exons 4a and 4b supports this hypothesis.

The expression of *ZNF195* appears to be significantly lower in PBMNC compared to the other tissues (Fig. 3.4f). As described in Section 3.3.3, the 336 bp *Eco* RI fragment detected significantly elevated expression of a 4.3 kb transcript in total RNA from presentation and relapse samples compared to the remission sample. In light of the fact that this 4.3 kb transcript does not represent *ZNF195*, and that no other transcripts were seen in the t(4;11) patient, there is no evidence to suggest that *ZNF195* expression is affected by the (4;11) translocation. Furthermore, Northern analysis of the t(4;11) PBMNC total RNA samples with the *ZNF195* 3'UTR probe did not detect any transcripts (data not shown). This result is consistent with the relatively low *ZNF195* expression observed in mRNA from PBMNC. It is likely that the 4.3 kb transcript with elevated expression in the t(4;11) patient represents a gene from the *ZNF91* family which has stronger expression in leukaemic blast cells than in normal PBMNC. Such members of this family are known to exist (Bellefroid *et al.*, 1993).

## 3.9 Discovery of a deletion/insertion polymorphism in the *ZNF195*

### 3'UTR

The 11p15.5 PAC described above contained all of the *ZNF195* cDNA sequence except that 52 nt were missing from the PAC over a sequence corresponding to the *ZNF195* 3'UTR. The number of missing nt was mistakenly reported as 47 bp in Hussey *et al.* (1997) due to a misinterpretation of the BLAST alignment of the PAC sequence with the *ZNF195* cDNA sequence. When BLAST encounters a significant gap in homology, it terminates the predicted homology and recommences after the gap. The DNA sequence just before the point of deletion is similar to that just after the deletion and this led to an incorrect interpretation of the BLAST result. A subsequent alignment of the sequence from the PAC and the *ZNF195* cDNA sequence, using the Sequencher large gap function, clearly showed that the gap is 52 nt long (Fig. 3.7). Another difference between the PAC and the *ZNF195* cDNA was the presence of an additional G in the PAC sequence, inserted in the region corresponding to the *ZNF195* 3'UTR sequence.

When the 3'UTR region of *ZNF195* was amplified by PCR in order to generate a probe for Northern analysis, a number of DNA specimens from normal donors were tested to ensure that only a single specific band was amplified consistently. Agarose gel electrophoresis detected only a single PCR product in most of the samples, however some of the samples had an additional PCR product approximately 50 bp lower than expected (Fig. 3.9). DNA sequence analysis of the two PCR products revealed that the upper band was identical to the 3'UTR sequence predicted from the 556 bp *Eco* RI-*Xba* I fragment of the Z104 cosmid. The lower product also corresponded to the expected sequence except that it lacked 52 nt between nt 2286 and 2339 of *ZNF195* and contained a single G insertion between nt 2572 and 2573 of *ZNF195*. A Sequencher alignment of the *ZNF195* cDNA sequence with the smaller PCR



### Figure 3.9 *ZNF195* imprinting analysis

(a) A region of the *ZNF195* 3'UTR was amplified by RT-PCR from the PBMNC of 1 homozygous (donor 1) and 5 heterozygous (donors 2-6) normal donors, as described in section 3.8. The products from two independent RT-PCR reactions from each normal donor are loaded next to each other on the agarose gel shown in this figure. Control reactions containing no reverse transcriptase enzyme (minus RT) were also performed independently for each individual and are loaded next to each other. PCR products from PBMNC genomic DNA from 3 of the 5 heterozygous donors are shown for comparison to the RT-PCR products. Five  $\mu$ l of product was loaded into each lane as follows-

|  |   |
|--|---|
| <b>Lane 1</b> pUC19/ <i>Hpa</i> II DNA size marker | <b>Lanes 15 and 16</b> donor 4 RT-PCR   |
| <b>Lane 2</b> H <sub>2</sub> O control reaction    | <b>Lanes 17 and 18</b> donor 4 minus RT |
| <b>Lanes 3 and 4</b> donor 1 RT-PCR                | <b>Lanes 19 and 20</b> donor 5 RT-PCR   |
| <b>Lanes 5 and 6</b> donor 1 minus RT              | <b>Lanes 21 and 22</b> donor 5 minus RT |
| <b>Lanes 7 and 8</b> donor 2 RT-PCR                | <b>Lanes 23 and 24</b> donor 6 RT-PCR   |
| <b>Lanes 9 and 10</b> donor 2 minus RT             | <b>Lanes 25 and 26</b> donor 6 minus RT |
| <b>Lanes 11 and 12</b> donor 3 RT-PCR              | <b>Lane 27</b> donor 3 genomic DNA PCR  |
| <b>Lanes 13 and 14</b> donor 3 minus RT            | <b>Lane 28</b> donor 4 genomic DNA PCR  |
|  | <b>Lane 29</b> donor 5 genomic DNA PCR  |

(b) A region of the *ZNF195* 3'UTR was amplified by RT-PCR from the tissues from a normal heterozygous individual, as described in section 3.8. The products from three independent PCR reactions for each tissue sample are loaded next to each other on the gel shown here. Control reactions containing no reverse transcriptase enzyme (minus RT) were also performed in triplicate for each individual and are loaded next to each other. 10  $\mu$ l of product was loaded into each lane as follows-

|  |                                    |
|--|------------------------------------|
| <b>Lane 1</b> pUC19/ <i>Hpa</i> II DNA size marker | <b>Lanes 12-14</b> kidney minus RT |
| <b>Lane 2</b> H <sub>2</sub> O control reaction    | <b>Lanes 15-17</b> lung RT-PCR     |
| <b>Lanes 3-5</b> liver RT-PCR                      | <b>Lanes 18-20</b> lung minus RT   |
| <b>Lanes 6-8</b> liver minus RT                    | <b>Lanes 21-23</b> spleen RT-PCR   |
| <b>Lanes 9-11</b> kidney RT-PCR                    | <b>Lanes 24-26</b> spleen minus RT |



product and the 11p15.5 PAC confirmed that the smaller PCR product and the PAC were identical in sequence.

A total of 50 individual DNA specimens were analysed to determine the frequency of each *ZNF195* allele with respect to the insertion/deletion polymorphism. Of the 50 individuals, 34 were homozygous for the upper allele, 16 were heterozygous and none were homozygous for the lower allele. This corresponds to frequencies of 84 % and 16 % for the upper and lower alleles respectively.

### **3.10 Assessing the imprinting status of *ZNF195***

The discovery of a number of individuals who were heterozygous for the deletion allele in the 3'UTR of *ZNF195* indicated that the deletion represented a naturally occurring transcribed sequence variation. In the context of the t(4;11) study this result was of no obvious significance. However, the location of a second Wilms' tumour suppressor gene (*WT2*) in the 11p15.5 region containing *ZNF195* has been indicated by loss of heterozygosity studies (reviewed in Besnard-Guerin *et al.*, 1996). The recurrent loss of maternal alleles in the *WT2* region in Wilms' tumours indicates that the *WT2* gene is paternally imprinted, ie that the maternally derived *WT2* allele is transcribed and the paternally derived allele is not transcribed or transcribed at very low levels (Schroeder *et al.*, 1987). The 11p15.5 region contains a cluster of both paternally and maternally imprinted genes including *H19*, *IGF2* and *CDKN1C* (*p57<sup>KIP2</sup>*) (Chung *et al.*, 1996). It is possible that this cluster might extend to *ZNF195*.

The discovery of the transcribed sequence variation in *ZNF195* made it possible to determine whether *ZNF195* was imprinted. The imprinting analysis was performed using PBMNC since

access to human tissues was restricted and PBMNC were easily obtained from willing normal donors.

In order to determine if there was any imprinting present, it was necessary to determine the standard ratio of amplification of both alleles in the case where the alleles were present in equal number. This was done by PCR amplification of the polymorphic region using genomic DNA from 5 normal individuals. The ratio of the intensity of ethidium bromide staining of the upper allele versus the lower allele was estimated to be approximately 1:3 and was consistent within and across all individuals tested (shown for 3 individuals in Fig. 3.9). The abundance of the lower allele over the upper allele can be explained by the fact that its smaller size gave it an amplification advantage over the upper allele. If *ZNF195* was imprinted in PBMNC one would expect to see a deviation of the 1:3 ratio in the RT-PCR product. This deviation would either be seen as a skew in amplification toward the upper allele or the lower allele, depending on which allele was inherited through the maternal or the paternal germline.

Total RNA was prepared from PBMNC from the 5 normal individuals who were heterozygous for the insertion/deletion polymorphism. Prior to reverse transcription, contaminating DNA was removed from the RNA by digestion with DNase I (Section 2.2.6.8). This was necessary because the Zf3'UTRR and Zf3'UTRR primers were in the same exon (because of the size of this exon, it was difficult to design the primers otherwise) and therefore the DNA and cDNA PCR products could not be differentiated because they were the same size. RT-PCR was performed in duplicate for each sample with MMLV reverse transcriptase (Section 2.2.6.6). After agarose gel electrophoresis, the ratio of ethidium bromide staining of the upper and lower allele was estimated (Fig. 3.9). The ratio of RT-PCR amplification of the upper and lower allele was very similar to that obtained in PCR of genomic DNA. This was the case for all 5 individuals tested, and duplicate PCR results were

consistent. Since both alleles were present in equal number in heterozygous genomic DNA, the RT-PCR results indicated that there was no significant imprinting of *ZNF195* in PBMNC.

An attempt was made to determine the imprinting status of *ZNF195* in other tissues. DNA samples (gift from Dr B. Tycko, College of Physicians and Surgeons, Columbia University) from 10 individual autopsies were assessed to determine if they were informative for the *ZNF195* polymorphism. Nine of the 10 individuals were homozygous for the upper allele and only one individual was heterozygous and therefore informative. cDNA (also from Dr B. Tycko, College of Physicians and Surgeons, Columbia University) from liver, kidney, lung and spleen tissue of this adult was available for testing.

The RNA samples that these cDNA samples were made from had not been treated with DNase I prior to cDNA synthesis, and therefore the samples contained contaminating genomic DNA. This meant that the usual PCR, with primers Zf3'UTRF and Zf3'UTRR, could not be used because both of these primers are in exon 5 and result in a product of the same size from genomic DNA and cDNA. A new primer was designed (ZfKRABBF), based on the least conserved sequence from the KRAB B box of *ZNF195*, which is in exon 3, approximately 12 kb 3' of the Zf3'UTRR primer. PCR using ZfKRABBF and Zf3'UTRR would not result in significant amplification of contaminating genomic DNA.

The cDNA samples from the heterozygous individual were PCR amplified using ZfKRABBF and Zf3'UTRR. Unfortunately this PCR resulted in several unexpected bands under the main product (data not shown), which made it difficult to determine the exact ratios of upper and lower alleles. However, it was obvious that both alleles were expressed.

RNA samples (gift from Dr B. Tycko, College of Physicians and Surgeons, Columbia University) from tissues of the same heterozygous individual were sent to our laboratory so that they could be DNase I treated to remove contaminating genomic DNA prior to cDNA synthesis. The concentration of RNA was measured by spectrophotometry and cDNA was made using MMLV reverse transcriptase. PCR was performed using primers Zf3'UTRF and Zf3'UTRR because contaminating genomic DNA was not expected to be a problem, and these primers had been shown to amplify the upper and lower alleles at a consistent ratio (Fig. 3.9). Unfortunately, PCRs performed in triplicate did not yield consistent results with respect to the amplification of both alleles (Fig. 3.9b). For example, triplicate RT-PCR of lung RNA, resulted in amplification of the lower allele only in one reaction, amplification of both alleles in the second reaction, and amplification of the upper allele only in the third reaction.

Some of the samples which had reverse transcriptase omitted for the cDNA synthesis step resulted in a PCR product, indicating that contaminating genomic DNA had not been completely removed prior to cDNA synthesis. It was also noted that a visible PCR product from the cDNA samples was only obtained after 45 cycles of PCR. This was in contrast to RT-PCR of PBMNC RNA which required only 35 PCR cycles for visible product. Taken together, these results suggested that the amount of RNA used for reverse transcription had been overestimated because the RNA samples contained a high concentration of contaminating genomic DNA. Due to time constraints these experiments could not be pursued further.

Despite problems in obtaining reproducible results, it was possible to conclude that expression of *ZNF195* was biallelic in all of the tissues examined. Therefore, if *ZNF195* is imprinted in any of these tissues, the imprinting would only be partial and would not involve

mono-allelic expression. Partial imprinting has been reported for the  $p57^{KIP2}$  locus on 11p15.5 (Chung *et al.*, 1996).

### 3.11 Discussion

The aims of this study were to characterise the zinc finger gene(s) within Z104 and to assess it as a candidate breakpoint gene in the (4;11) translocation. Analysis of the results led to the conclusions that Z104 contained only one zinc finger gene, *ZNF195*, and that this gene was not disrupted by the (4;11) translocation (see below). The following discussion concerns the organisation and expression of *ZNF195* and also the position of *ZNF195* with respect to the (4;11) translocation breakpoint.

#### 3.11.1 Northern analysis of *ZNF195* expression

It is clear from the results described in this chapter that the 336 bp *Eco* RI fragment cross-hybridised to a gene related to *ZNF195* and probably belonging to the *ZNF91* family. The related gene was detected in Northern analysis as a 4.3 kb transcript that was present at far greater abundance than the actual 2.5 kb *ZNF195* transcript. Southern analysis using the 336 bp *Eco* RI fragment indicated that this sequence was quite specific to *ZNF195*. This fragment hybridised most strongly to a 336 bp *Eco* RI fragment, a 2.8 kb *Eco* RI fragment and a 3 kb *Hind* III fragment. The sizes of the restriction fragments are in agreement with the *ZNF195* restriction fragments predicted, from the PAC containing the *ZNF195* gene, to contain the 336 bp *Eco* RI fragment. It therefore seemed reasonable to assume that the predominant 4.3 kb transcript seen on Northern analysis using the 336 bp *Eco* RI fragment represented expression of *ZNF195*. Subsequent Northern analysis with a 3'UTR probe specific for *ZNF195* indicated that this assumption was incorrect. This apparent discrepancy between the Southern and Northern analysis can only be explained if the 4.3 kb transcript is expressed at a much higher

level than *ZNF195*. As mentioned in the results Section, some members of the *ZNF91* family with 4.3 kb transcripts are up-regulated in T-cell leukaemias (Bellefroid *et al.*, 1993).

### 3.11.2 Alternative splicing of *ZNF195*

*ZNF195* contains two alternatively spliced exons, 4a and 4b, that are absent from the major transcript. Northern analysis suggested that isoforms containing at least exon 4a and possibly 4b were expressed abundantly in the HUT-78 leukaemic T-cell line and in foetal brain. It is a matter of speculation whether the alternative splicing of the Alu related exon 4a has functional significance. Exon 4a has 74% amino acid identity to exon 13A of *c-myb*. It has been shown for *c-myb* that the Alu sequence present is translated (supporting detail in Genbank U22376). The presence of inverted Alu sequences within proteins is intriguing and may allow the evolution of new properties (Miller and Zeller, 1997). Exon 4b is unrelated to anything else in the database and its functional significance also remains to be determined.

### 3.11.3 What are the functions of *ZNF195*?

The results presented in this chapter show that expression of *ZNF195* is biallelic in all of the tissues examined, including adult kidney. In the tumour suppressor model for *WT2*, only one allele of *WT2* is expressed in normal kidney and tumorigenesis arises when this allele is deleted and results in absence of *WT2* expression. The biallelic expression of *ZNF195* in normal adult kidney is not in accordance with the proposed tumour suppressor model for *WT2*. In addition, it was recently shown by Northern analysis that *ZNF195* is expressed at normal levels in Wilms' tumours (Dao *et al.*, 1999). Therefore, it seems unlikely that *ZNF195* is the *WT2* gene.

An exploration of the other functions of *ZNF195* was outside the scope of this study. Recently, differential display was used to study genes that are regulated in the placenta in



response to physiologically or pathologically induced low levels of oxygen (Pak *et al.*, 1998). A sequence corresponding to the 3' end of the *ZNF195* cDNA was found to be down-regulated in response to exposure to a decrease in oxygen levels. It was suggested by the authors that *znf195* may play a role in placental maintenance during pregnancy.

### 3.11.4 Genomic position of *ZNF195*

Most *Krüppel* type zinc finger genes in the *ZNF91* family (of which *ZNF195* is a member) occur in a large cluster, at the interface of chromosomal bands 19p12-p13.1, consisting of at least 40 zinc finger genes (Bellefroid *et al.*, 1993). Parts of this cluster contain up to 4 zinc finger genes in a 300 kb region. It was therefore of interest to determine if *ZNF195* was clustered on 11p15.5 with other members of the *ZNF91* family. In order to assess this, a BLAST analysis of the entire 132 kb PAC sequence, containing *ZNF195*, was performed. This PAC contained no zinc finger coding sequence other than that of *ZNF195*, indicating that *ZNF195* was not within close proximity to other zinc finger genes.

Some parts of the 11p15.5 region are quite well covered by contiguous genomic clones (see <http://www.ncbi.nlm.nih.gov/genome/seq/>). However BLAST N searches, using the sequences at the ends of the PAC containing *ZNF195* detected no homologous sequences, indicating that there were gaps of unknown length on both sides of PAC. The closest PAC centromeric of the PAC containing *ZNF195* is 162 kb long (Genbank accession AC018803) and a BLAST search of the sequence of this PAC revealed that it contains no zinc finger genes. Similarly, the closest PAC telomeric of *ZNF195* is 244 kb long (Genbank accession AC001228) and does not contain any zinc finger genes. These results, in addition to the fact that no other zinc finger genes have been mapped to 11p15.5, make it likely that 11p15.5 does not contain a clustered region of zinc finger genes. Therefore *ZNF195* appears to be the sole member of the *ZNF91* family on 11p15.5.

The absence of known zinc finger genes on 11p15.5 raises the question of how *ZNF195* came to be located on 11p15.5. The 3' end of *ZNF195* is immediately followed by a retroviral DNA sequence called Herv-Fb, for human endogenous retrovirus-Fb (Kjellman *et al.*, 1999a; Kjellman *et al.*, 1999b). Homologous recombination between retroviral sequences is an important mechanism in primate genomic organisation and evolution (reviewed in Sverdlov, 2000). It is possible that *ZNF195* was incorporated into 11p15.5 from the *ZNF91* cluster on 19p12-p13.1 by homologous recombination between Herv-Fb like retroviral DNA sequences common to 11p15.5 and 19p12-p13.1. In favour of this hypothesis is the presence of retroviral sequences on 19p12-p13.1 which are highly related to Herv-Fb and are in close proximity to *ZNF91* related zinc finger genes (supporting evidence Genbank Accession nos. AC007204 and AC004510).

### **3.11.5 Is *ZNF195* the 11p15.5 breakpoint gene in t(4;11)?**

Redeker *et al.* (1994) established an ordered restriction map of 11p15.5 in order to facilitate study of this region. They identified a *Bss* HII restriction fragment containing the cosmids D11S459, D11S470 and Z104 and determined that this fragment was telomeric of *RRM1*. However, they were not able to determine the relative order of these three markers. Subsequent to their study, *D11S459* was found to be telomeric of *D11S470*, but the position of the Z104 cosmid with respect to the two markers remained unknown (Shows *et al.*, 1996).

The 11p15.5 PAC described in Section 3.6.1 contains the entire *ZNF195* gene between nt 68,624 and 89,434. A BLAST N search using the sequence of the PAC revealed that it also contains the markers *D11S470* between nt 119,901 and 120,094 and *D11S459* between nt 90,006 and 90,183. Combining this data with the known order of *D11S470* and *D11S459*, it follows that the order of these markers is centromere-*D11S470*-*D11S459*-3' *ZNF195* 5'-telomere.

The size of the D11S470 cosmid is stated as somewhere between 36-153 kb (Redeker *et al.*, 1994). Given that cosmid D11S470 maps distal to the t(4;11) breakpoint (Dobrovic *et al.*, 1994), the estimated minimum distance between the 3'end of *ZNF195* and the breakpoint is between 66 and 183 kb (66 to 183 kb equals 30 kb (the distance between the 3'end of *ZNF195* and *D11S470*) plus 36 to 153 kb (the size of the D11S470 cosmid). The placement of *ZNF195* 60-183 kb distal to the breakpoint definitively excluded *ZNF195* as the chromosome 11 breakpoint gene in the (4;11) translocation. It was thus necessary to identify other candidate genes on 11p15.5.

## **Chapter 4**

### **Identification of the t(4;11)(q21;p15) breakpoint genes**

## Chapter 4

### Identification of the t(4;11)(q21;p15) breakpoint genes

The nup98 protein (for nucleoporin of 98kDa) was initially identified in *Rattus norvegicus* as a 98kDa protein component of the nuclear pore complex (NPC) (Radu *et al.*, 1995b). The partial amino acid sequence of this protein was used to create a probe for screening a cDNA library, and the complete cDNA sequence for rat *Nup98* was obtained (Radu *et al.*, 1995b). Nup98 is a vertebrate member of a family of nucleoporins that were first identified in yeast (Wente *et al.*, 1992; Wimmer *et al.*, 1992). Members of this nucleoporin family contain numerous phenylalanine-glycine (FG) repeats, many of which are flanked by glycine(G) and leucine(L), and are referred to as GLFG nucleoporins. The FG repeats function as docking sites for karyopherin  $\beta$  and other molecules in transport across the NPC (Moroianu *et al.*, 1995; Radu *et al.*, 1995b).

Subsequent to its discovery in *Rattus norvegicus*, *NUP98* was identified in *Homo sapiens* as the chromosome 11 translocation partner of *HOXA9* in a recurrent (7;11)(p15;p15) translocation associated with acute myeloid leukaemia (AML) (Nakamura *et al.*, 1996a; Borrow *et al.*, 1996). Borrow *et al.* (1996) reported a 3630 bp cDNA sequence for human *NUP98* (Genbank U41815). The *NUP98* breakpoint in t(7;11)(p15;p15) occurred in the intron between nt 1552 and 1553 of the cDNA sequence (Borrow *et al.*, 1996; Nakamura *et al.*, 1996a).

The (7;11)(p15;p15) translocation results in a fusion protein containing the amino-terminus of nup98, with 37 of the 38 FG repeats, fused to the carboxy-terminus of the homeobox protein *hoxa9* which retains the DNA binding homeobox sequence (Borrow *et al.*, 1996; Nakamura *et*

*al.*, 1996a). The *nup98-hoxa9* fusion does not retain the transcriptional repression domain present in the amino-terminal portion of *hoxa9* (Kasper *et al.*, 1999).

The nucleoporin 98 gene is located within 11p15.5 proximal to the region recognised by the cosmid Z104 (Nakamura *et al.*, 1996a). *NUP98* is present on two *Bss* HII fragments of 340 and 170 kb, the latter fragment also contains Z104 (Nakamura *et al.*, 1996a). FISH of Z104 mapped it to the der(4) chromosome of the t(4;11) patient, indicating that Z104 is distal to the breakpoint (Finch, Peters and Dobrovic, unpublished results). In the discussion of Chapter 3 it was deduced, using the sequence of the PAC pDJ1173a5 (Genbank AC000378), that Z104 is distal to *D11S470*. This PAC contains the sequence for Z104 and *D11S470*, but not the sequence for *NUP98*, thereby placing *NUP98* proximal to *D11S470* and within the breakpoint region (Fig. 4.1a). The positioning of *NUP98* in the candidate breakpoint region in the t(4;11) patient, and its involvement in a recurrent translocation associated with AML, made it a strong candidate for the chromosome 11 breakpoint gene in the (4;11)(q21;p15) translocation. This chapter describes the investigation of *NUP98* as a candidate breakpoint gene.

## **4.1 Investigating *NUP98* as a candidate breakpoint gene in t(4;11)(q21;p15)**

The potential disruption of *NUP98* in the t(4;11) patient was investigated simultaneously by 3 methods (Northern analysis, Southern analysis and PCR analysis of hybrids) in case one or more methods did not yield informative results.

### **4.1.1 Northern analysis**

The (7;11)(q21;p15) translocation creates a novel mRNA consisting of the 5' end of *NUP98* fused to *HOXA9* (Borrow *et al.*, 1996; Nakamura *et al.*, 1996a). Therefore, it was reasoned that if *NUP98* was disrupted in the (4;11) translocation this would most likely also result in a

**Figure 4.1 Testing *NUP98* as a candidate breakpoint gene in t(4;11)**

(a) Position of *NUP98* and *ZNF195* with respect to the 11p15.5 breakpoint region. *NUP98* lies within the candidate breakpoint region indicated by the arrowed line. The beta chain of hemoglobin (*HBBC*) and the H-ras oncogene (*HRAS*) are at the extremities of 11p15.5. C and T denote centromeric and telomeric respectively.

(b) Southern analysis of t(4;11) presentation and normal donor DNA with the *NUP98* cDNA probe described in section 4.1.1. Each lane contains 10 µg of genomic DNA extracted from PBMNC. Novel restriction fragments in the t(4;11) presentation digests are indicated with an arrow head. Lane details are as follows-

Lane 1 donor/*Bam* HI

Lane 2 t(4;11)/*Bam* HI

Lane 3 donor/*Eco* RI

Lane 4 t(4;11)/*Eco* RI

Lane 5 donor/*Pst* I

Lane 6 t(4;11)/*Pst* I

Lane 7 donor/*Taq* I

Lane 8 t(4;11)/*Taq* I

Lane 9 donor/*Xba* I

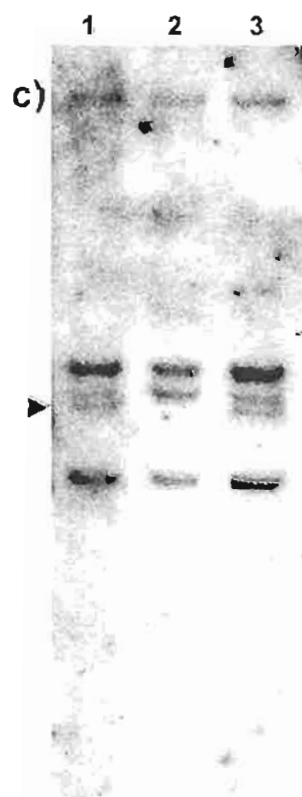
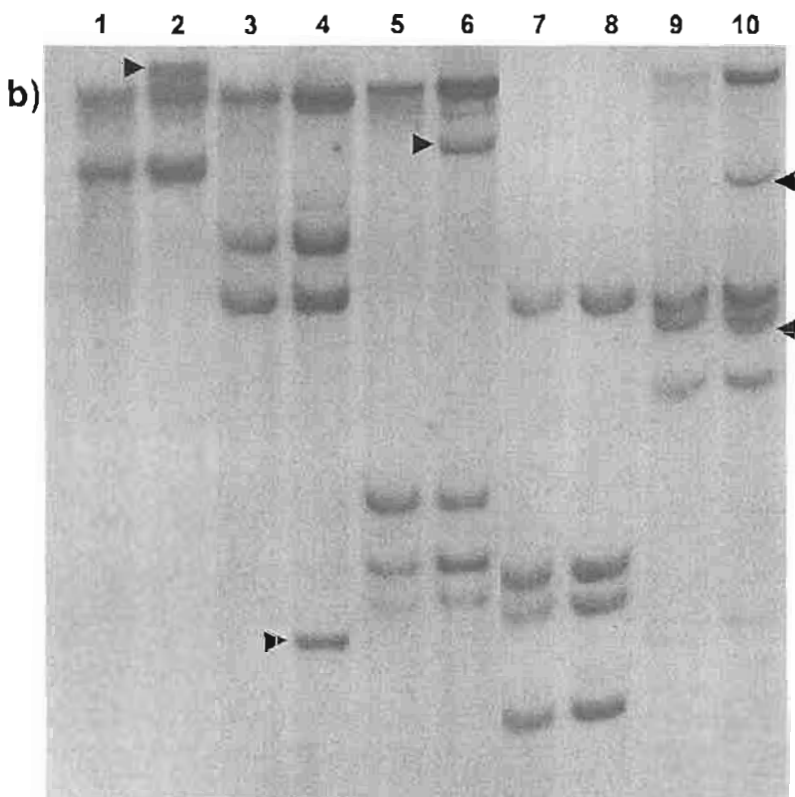
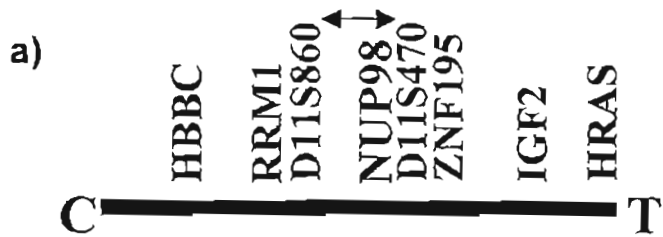
Lane 10 t(4;11)/*Xba* I

(c) Southern analysis of t(4;11) presentation, remission and relapse DNA with the *NUP98* cDNA probe described in section 4.1.1. Lanes 1 and 3 contain 10 µg of genomic DNA extracted from PBMNC, and lane 2 contains 10 µg of genomic DNA extracted from bone marrow. The novel restriction fragment in the t(4;11) presentation and relapse sample is indicated with an arrow head. Lane details are as follows-

Lane 1 t(4;11) presentation/*Xba* I

Lane 2 t(4;11) remission/*Xba* I

Lane 3 t(4;11) relapse/*Xba* I





novel hybrid message consisting of *NUP98* sequence and sequence from chromosome region 4q21. If this transcript differed in size from normal *NUP98*, and was expressed at sufficiently high levels, it would be possible to detect it by Northern analysis with a *NUP98* probe.

In order to test for the presence of a novel transcript in the t(4;11) patient, Northern analysis was performed using a *NUP98* cDNA probe. The probe was made by RT-PCR using total RNA isolated from PBMNC of a normal donor. Oligonucleotide primers were designed based on the *NUP98* cDNA sequence reported by Borrow *et al*, (1996). An explanation of primer nomenclature is given in Table 2.1. Primers were positioned approximately equidistant to nt 1552 (N98 1265F, and N98 1848R) so that the PCR product contained approximately 300 bp of *NUP98* cDNA sequence 5' and 3' of the reported breakpoint. Since this probe contains sequence flanking the reported breakpoint, it should theoretically detect any novel *NUP98* transcripts in Northern analysis and any novel restriction fragments in Southern analysis.

Initial Northern analysis using total RNA from the PBMNC of t(4;11) patient presentation, remission and relapse samples as well as three normal controls revealed two predominant transcripts of 4.05 kb and 7.25 kb in all samples (data not shown). The size of these transcripts corresponded well to those reported by Borrow *et al*, (1996). The *NUP98* bands in the presentation sample were weak but this was explained by the relative underloading of this sample and that fact that the RNA was quite degraded. There was no convincing evidence of novel transcripts of distinctly different sizes to those of *NUP98*.

#### **4.1.2 Southern analysis**

To test for disruption of *NUP98* in t(4;11)(q21;p15), Southern analysis of t(4;11) patient genomic DNA, digested with either one of *Bam* HI, *Eco* RI, *Pst* I, *Taq* I or *Xba* I, was

performed using the 586 bp *NUP98* cDNA probe described in Section 4.1.1. A normal control DNA was included so that any novel restriction fragments in the patient could be detected.

The probe clearly detected novel restriction fragments in the *Bam* HI, *Eco* RI, *Pst* I and *Xba* I digested leukaemic DNA specimens (Fig. 4.1b). The presence of these novel bands indicated a rearrangement of *NUP98* in the (4;11) translocation.

There was a remote possibility that these novel restriction fragments were due to the presence of *NUP98* restriction fragment length polymorphisms in the patient, rather than due to a rearrangement of the *NUP98* gene. To exclude this possibility, t(4;11) leukaemic and remission DNA specimens were digested with *Eco* RI and *Xba* I and subjected to Southern analysis using the 586 bp probe (*Xba* I result shown in Fig. 4.1c). The novel 1.7 kb *Eco* RI and 4.4 kb *Xba* I restriction fragments in the presentation and relapse DNA samples were absent from the remission DNA sample. This confirmed that the novel *Eco* RI and *Xba* I fragments in the t(4;11) leukaemic DNA specimens were due to a rearrangement of the *NUP98* gene.

#### **4.1.3 Testing der(4) and der(11) for retention of *NUP98* sequences flanking the t(7;11)(p15;p15) breakpoint.**

The Southern analysis described above had clearly demonstrated a disruption of the *NUP98* gene in the t(4;11) patient. However, Northern analysis had failed to detect a novel *NUP98* fusion transcript. It was possible that disruption of *NUP98* may not create a hybrid message but may simply inactivate the *NUP98* gene. Alternatively, a hybrid transcript may be very similar in size to one of the normal *NUP98* transcripts and therefore co-migrate with normal *NUP98*. This would make it difficult to distinguish a hybrid transcript from the normal *NUP98* transcripts. Therefore, as a complement to Northern and Southern analysis, genomic

DNA from the somatic cell hybrids segregating the der(4) and der(11) chromosomes was tested for the retention of *NUP98* sequences on either side of the reported *NUP98* breakpoint.

#### **4.1.3.1 Primer pairs 5' of the published *NUP98* breakpoint**

The *NUP98* exon immediately 5' of the published breakpoint was known to be 141 bp long (Nakamura *et al.*, 1996a). This exon corresponds to that later defined as exon B in a subsequent study of the *NUP98* breakpoint region (Arai *et al.*, 1997). The primer pair N98 1428F + N98 1531R was used to PCR amplify genomic DNA sequence within exon B.

Four primer pairs were used to PCR amplify small regions of *NUP98* that occurred 5' of the reported t(7;11)(p15;p15) breakpoint, ie 5' of nucleotide 1552. The sequence of the primers was derived from the *NUP98* cDNA sequence published by Borrow *et al.*, (1996) (Genbank U41815). Primer pairs N98 81F + N98 185R, N98 301F + N98 388R and N98 1265F + N98 1384R were positioned away from the reported t(7;11) breakpoint in case the potential *NUP98* breakpoint in the t(4;11) patient occurred at a different position. Prior to testing these primers on genomic DNA, it was impossible to know if they would amplify a product of the size predicted from the cDNA sequence. This was due to the fact that the corresponding *NUP98* exon-exon boundaries had not been determined, so it was not known if one or more introns occurred between the primers.

Before testing genomic DNA from the somatic cell hybrids, each primer pair was tested using PCR on cDNA and genomic DNA to determine if the primers would amplify genomic DNA. In summary all primer pairs amplified a product of the expected size from cDNA. All primer pairs, except 1428F + 1531R, failed to amplify a product from genomic DNA even when the time of the PCR extension step was increased, indicating that there was at least one intron present between these primers.

Primer pair N98 1428F + N98 1531R amplified the expected product of 104 bp from genomic DNA. When these primers were used to amplify der(4) and der(11) DNA from the somatic cell hybrids, a 104 bp product was amplified from der(4) DNA, der(11) DNA as well as mouse genomic DNA and human genomic DNA. No sequence data for mouse *Nup98* was available, however the human *NUP98* and rat *Nup98* exon B sequences were known to be 95% homologous (Nakamura *et al.*, 1996a). Therefore the human and mouse exon B sequences were also likely to be highly homologous. The exon B primers were able to bind to and amplify the corresponding sequence in mouse *Nup98*.

A comparison of the human and rat sequences of exon B showed that there was a *Taq* I restriction endonuclease site that was present in the rat sequence but absent from the human sequence. This *Taq* I site was shown to be present in the mouse PCR product (Fig. 4.2a). Thus to distinguish between the human and mouse PCR products, the PCR products were restriction digested with *Taq* I (Fig. 4.2a). The product from der(4) DNA was fully digested but the product from der(11) DNA was only partially digested. This indicated that human *NUP98* and mouse *Nup98* exon B had been amplified from the der(11) somatic cell hybrid DNA, but only mouse *Nup98* exon B had been amplified from the der(4) DNA. Therefore *NUP98* exon B was present on the der(11) chromosome.

#### **4.1.3.2 Primer pairs 3' of the published *NUP98* breakpoint**

Two primer pairs were designed 3' of the published *NUP98* breakpoint, namely N98 1585F + N98 1681R and N98 1742F + N98 1848R.

PCR of genomic DNA and cDNA with N98 1742F + N98 1848R resulted in a product of the size predicted from the cDNA sequence, indicating that these two primers were within one exon of *NUP98* (data not shown). The exon in which these primers are situated was later

**Figure 4.2 PCR analysis of the der(4) and der(11) containing somatic cell hybrids**

**(a)** *NUP98* exon B PCR product digested with *Taq* I.

Mouse and human *NUP98* cDNA sequences are highly conserved and the exon B PCR also amplified mouse *NUP98*. The mouse and human exon B PCR products were distinguished by a *Taq* I restriction site, which is present in the mouse product but absent in the human product.

**(b)** *NUP98* exon C PCR product.

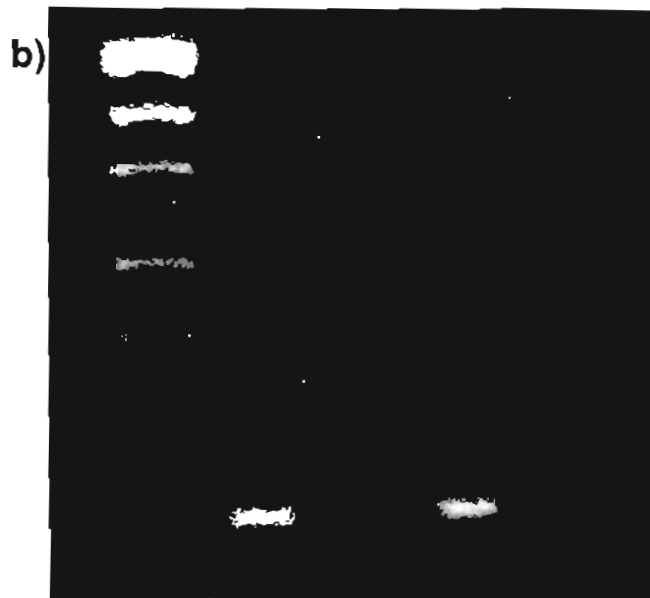
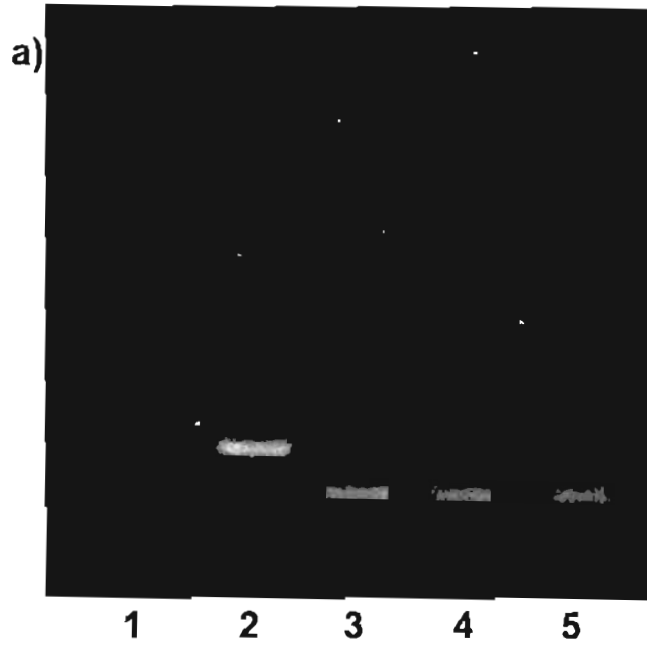
**Lane 1** pUC19/*Hpa* II DNA size marker

**Lane 2** normal human donor

**Lane 3** mouse

**Lane 4** der(4) hybrid

**Lane 5** der(11) hybrid



defined as exon D (Arai *et al.*, 1997). Similarly, PCR of genomic DNA and cDNA with N98 1585F + N98 1681R indicated that these two primers were within one exon of *NUP98* (Fig. 4.2b). The exon containing primer pair N98 1585F + N98 1681R was later defined by Arai *et al.* (1997) as exon C.

PCR of der(4) and der(11) DNA with primer pair N98 1742F + N98 1848R showed that exon D of *NUP98* was present on the der(4) chromosome (data not shown). Therefore the *NUP98* breakpoint in the t(4;11) patient occurred somewhere between exons B and D. PCR of der(4) and der(11) DNA using primer pair N98 1585F + N98 1681R showed that *NUP98* exon C was present on der(4) (Fig. 4.2b). Therefore, since exons B and C are on the complementary derivative chromosomes, *NUP98* is disrupted between exons B and C in the t(4;11) patient and is the chromosome 11 breakpoint gene. This breakpoint is identical, with respect to the *NUP98* cDNA sequence, to that first identified in the (7;11)(p15;p15) translocation (Borrow *et al.*, 1996; Nakamura *et al.*, 1996a).

## 4.2 Identification of the chromosome 4 breakpoint gene

After identifying *NUP98* as the chromosome 11 breakpoint gene, there were two choices for the method of identifying the chromosome 4 translocation sequence. One choice involved a genomic approach and the other involved a cDNA based approach. Both of these approaches, and the reasons for selecting the cDNA approach, are discussed below.

### The genomic approach

Southern analysis with a *NUP98* cDNA probe spanning the breakpoint revealed a novel leukaemia specific 1.7 kb *Eco* RI fragment. It would have been possible to create an *Eco* RI digest of DNA from the t(4;11) patient and attach linkers to the *Eco* RI fragments. PCR using the N1428F primer and a reverse primer from the linker sequence could then be used to

generate a leukaemia specific PCR product. The sequence of the PCR product could be analysed using BLAST to see if it corresponded to a known gene. Alternatively the PCR product could be used as a probe for screening a chromosome 4 genomic library. Positive clones could be exon trapped and subject to BLAST analysis to see if their sequence corresponds to a known gene. If the chromosome 4 gene is novel then Northern analysis could be used to determine which tissues it is expressed in. Primers could be designed based on exon trapped sequences and used in 5' and 3' RACE of RNA from the appropriate tissue to characterise the entire cDNA. The problem with this approach is that the library screening may identify clones with sequence that is entirely intronic.

#### The cDNA approach

Initial Northern analysis of the t(4;11) presentation sample did not reveal any convincing novel *NUP98* transcripts (Section 4.1.1). However, it was still possible that any such transcripts were present but not detected because they were of similar size to *NUP98* and co-migrated with the *NUP98* transcript. Failure to detect a novel transcript may have also been caused by the degradation of the presentation sample, or the fact that the novel transcript was expressed at levels below the detection limit of Northern analysis.

For fusion transcripts, the 3' RACE procedure (3' Rapid Amplification of cDNA Ends) can be used to determine the partner gene. 3' RACE allows amplification of nucleic acid sequences from an mRNA template between a defined internal site and the 3' end of the mRNA (Frohman *et al.*, 1988). Briefly, this method involves reverse transcription of RNA using an oligo-dT adapter primer which binds to the polyA tract in mRNA. This primer has a 20 bp adapter sequence at its 5' end, usually containing one or more restriction endonuclease recognition sites for subsequent cloning. Specific cDNA is amplified from the reverse transcription reaction using a gene specific forward primer that anneals to known exonic



sequence and a reverse primer with a sequence complementary to that of the adapter sequence.

At the time that these experiments were being planned, details of two recurrent AML associated chromosomal rearrangements involving *NUP98* were available. One of these was the (7;11)(p15;p15) translocation described above, and the other was the inv11(p15 q22) which fuses the *NUP98* and *DDX10* genes. Both of these chromosomal rearrangements result in fusion mRNAs containing the 5' end of *NUP98* (Arai *et al.*, 1997; Borow *et al.*, 1996; Nakamura *et al.*, 1996a). Given that the t(4;11)(q21;p15) breakpoint is identical to that reported in t(7;11) and inv(11), it was probable that a chimaeric fusion mRNA containing *NUP98* and chromosome 4 sequences would be expressed in the leukaemic cells of the t(4;11) patient. For this reason, and the fact that this approach could potentially yield results much faster than the genomic approach, the cDNA approach was the method of choice for identifying the chromosome 4 sequence fused to *NUP98*.

#### **4.2.1 3' RACE to determine the sequence fused to *NUP98***

3' RACE (Section 2.2.7) was used to determine the chromosome 4 gene fused to *NUP98*. Experiments were done in parallel on PBMNC from the presentation sample of the t(4;11) patient and a normal individual.

The Expand Long Template PCR System (Section 2.2.6.2) was used in all subsequent PCR amplifications, because a hybrid transcript might have had a length above the amplification range of standard *Taq* polymerase. The reverse transcription product was amplified with a *NUP98* exon B primer, N1459F, in combination with the Abridged Universal Amplification Primer (AUAP).

The first round 3' RACE products were barely visible as distinct bands and appeared more like a smear on the agarose gel. This is a common problem of 3' RACE and is caused by a lack of specificity in the first round of PCR. This problem was overcome using a published protocol which involves enrichment of the specific sequences from the first PCR and a subsequent second round of PCR (Lankiewicz *et al.*, 1997). Second round PCR of the enriched product was performed using a *NUP98* exon B primer, N1511F (internal to the previous 2 oligonucleotides) and the AUAP.

The second round PCR products were compared to the first round products by running them next to each other on an agarose gel (Fig. 4.3a). It was clear that the enrichment procedure had greatly improved the specificity of the 3' RACE products. The first round PCR products barely contained products visible above the smear. In comparison, the second round products had distinctly visible bands. A second round PCR without intermediate magnetic purification was also performed on an amount of product judged to be equivalent to that used in second round with magnetically purified products. This PCR resulted in a smear very similar to the first round and demonstrated that the magnetic enrichment procedure had increased the specificity of amplification.

There appeared to be a cluster of bands in the t(4;11) patient at around the 2 kb position on the gel shown in Fig. 4.3a. In order to resolve this cluster into distinct bands the products were run on a wide lane agarose gel (Fig. 4.3b). The increased resolution separated the bands around the 2 kb position and it was apparent that one product was common to the normal individual and the t(4;11) patient (band 2). In addition to this product, another 5 bands were visible in the leukaemic presentation sample (bands 1,3,4,5 and 6). The bands were excised from the gel and purified using Wizard PCR Preps. The bands were then re-amplified with the same primers and then gel purified again. The final products were either sequenced with the

**Figure 4.3 3' RACE to identify the *NUP98* fusion partner**

(a) Comparison of 3' RACE products from a normal donor and t(4;11) presentation before and after enrichment.

**Lane 1** SPPI/*Eco* RI DNA size marker

**Lane 2** t(4;11) 1<sup>st</sup> round 3' RACE

**Lane 3** t(4;11) 2<sup>nd</sup> round 3' RACE without prior enrichment

**Lane 4** t(4;11) 2<sup>nd</sup> round 3' RACE with prior enrichment

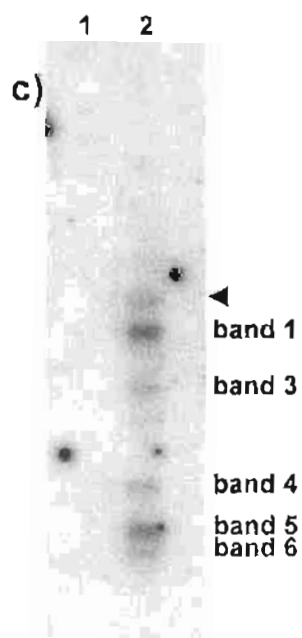
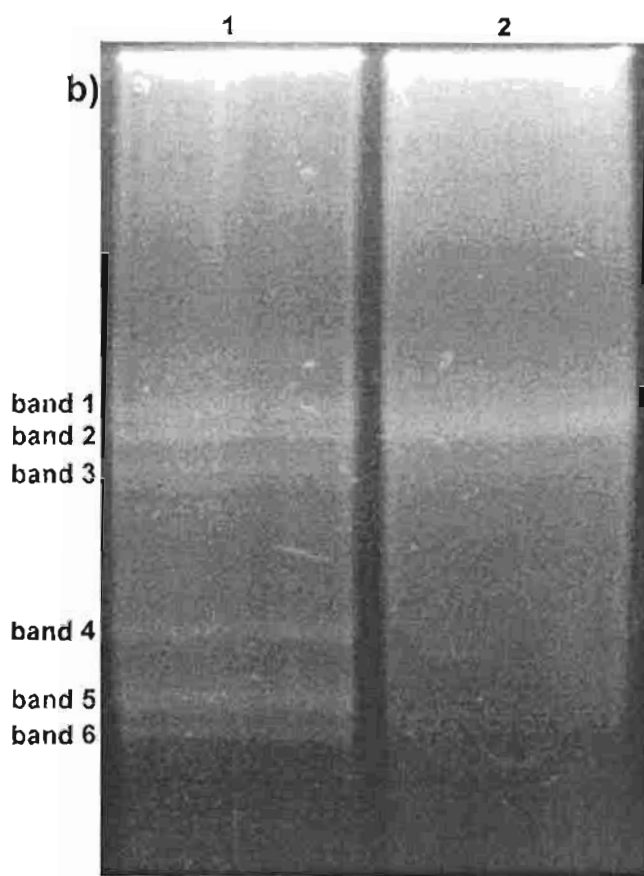
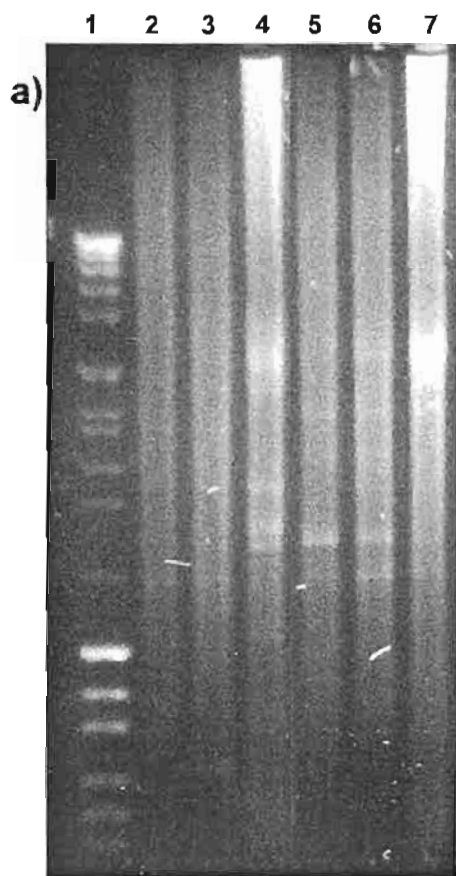
**Lane 5** donor 1<sup>st</sup> round 3' RACE

**Lane 6** donor 2<sup>nd</sup> round 3' RACE without prior enrichment

**Lane 7** donor 2<sup>nd</sup> round 3' RACE with prior enrichment

(b) Second round 3' RACE products from (**Lane 1**) t(4;11) presentation and (**Lane 2**) a normal donor electrophoresed on a wide lane agarose gel. 6 distinct bands are visible in the t(4;11) sample and are denoted bands 1-6.

(c) Hybridisation of a *RAP1GDS1* oligonucleotide probe (R108R) to a Southern blot of normal donor (**Lane 1**) and t(4;11) (**Lane 2**) 2<sup>nd</sup> round 3' RACE products. The band indicated with an arrowhead, which was not visible on the agarose gel, corresponds in size to the upper isoform of *NRG* (see Fig. 4.5).



primers used to generate them or cloned into pGEM-T (Promega) and sequenced with Universal Sequencing Primers.

The sequence of each product was analysed using the BLAST N algorithm to search the Genbank sequence database. The common product in the leukaemic and normal RT-PCR (band 2) corresponded to the normal 4.05 kb *NUP98* transcript (Genbank U41815). The other products in the leukaemic sample all proved to be chimaeric sequences consisting of the 5' end of *NUP98* up to and including exon B (see below).

A product (band 1) that was slightly larger than the normal *NUP98* band had the 5' end of *NUP98* fused with the coding region of the guanine nucleotide disassociation stimulator gene, *RAP1GDS1* (Genbank X63465). Although this product was not initially sequenced over its full length, it contained the first coding exon of *RAP1GDS1* and was shown, by hybridisation of an oligonucleotide probe (R1677R) to a Southern blot of the 3' RACE products, to contain sequence including the *RAP1GDS1* stop codon (data not shown). By virtue of its size, it was presumed to contain all the published sequence in-between. RT-PCR with a forward primer from the start codon of *NUP98* (N145F) and a reverse primer from the stop codon of *RAP1GDS1* (R1677R) was performed. This resulted in a product corresponding in size to band 1, which when sequenced was found to contain the 5' end of *NUP98* fused to the coding region of *RAP1GDS1* (see Section 5.1.1 for this result and a detailed description of the *RAP1GDS1* coding sequence). The fusion maintained the reading frame of *RAP1GDS1*. This hybrid transcript is hereafter denoted as *NRG* (for *NUP98-RAP1GDS1*). The *RAP1GDS1* sequence in *NRG* starts at nucleotide 5 of the coding sequence. The methionine and the first G of the codon for aspartic acid are lost. However, the aspartic acid is retained in the fusion protein because the last base of *NUP98* exon B is a G (Fig. 4.4a).

#### Figure 4.4 RT-PCR analysis and sequence of *NRG*

(a) Nucleotide and amino acid sequences around the junctions of the *NRG* and *NRG2* fusion transcripts (Genbank Accession AF133331 and AF133333 respectively).

(b) RT-PCR analysis of *NRG* and *RGN* fusion transcripts in three t(4;11) patients.

RT-PCR was performed on PBMNC from the presentation samples of three t(4;11) patients with ALL, and on PBMNC from two normal donors. The primers used are described in section 4.3. Control reactions containing no reverse transcriptase enzyme (minus RT) were also performed. The primers used are described in section 4.3. H<sub>2</sub>O controls are negative control RT-PCRs without target. The lane marked **M** contains both SPP1/*Eco* RI and pUC19/*Hpa* II DNA size markers. The most prominent bands in the *RGN* PCR of patient 3 are alternative splicings of *RGN* with and without exon B.

**Lane 1** patient 1 RT-PCR

**Lane 2** patient 2 RT-PCR

**Lane 3** patient 3 RT-PCR

**Lane 4** donor 1 RT-PCR

**Lane 5** donor 2 RT-PCR

**Lane 6** patient 1 minus RT

**Lane 7** patient 2 minus RT

**Lane 8** patient 3 minus RT

**Lane 9** H<sub>2</sub>O control

(c) Hybridisation of a *NUP98* exon B (N1511F) oligonucleotide probe to a Southern blot of the gel in (b) above. Hybridisation to the *RGN* products in patient 3 (Lane 3) indicates that *NRG2* is not simply an alternatively spliced form of *NRG*.



#### 4.2.1.1 Characterisation of smaller RACE products

The other RACE products (bands 3, 4, 5 and 6) that were cloned and sequenced had an identical *NUP98-RAP1GDS1* junction to *NRG* but continued into presumed *RAP1GDS1* intron/exon splice sites and terminated in either introns of *RAP1GDS1* or as yet unsequenced exons of *RAP1GDS1* (Table 4.1, bands 3-6). One of the RACE products, (band 4), was characterised in more detail. The sequence identity of this clone to *RAP1GDS1* terminated at nt 112 (*RAP1GDS1* sequence is numbered according to Genbank X63465) and was immediately followed by GTAAG which is the consensus intronic sequence present immediately after a 5' splice site in higher eukaryotes (Moore *et al.*, 1993). A stop codon occurred 83 bp after the point of divergence from the *RAP1GDS1* sequence but there was no polyadenylation signal. A reverse primer (Rint2r) was made in the region containing the stop codon and used in PCR with a forward primer (R11F) from the preceding *RAP1GDS1* sequence. The DNA product was the same length as that obtained from cDNA (data not shown), confirming that the non-*RAP1GDS1* coding sequence is an intron that is contiguous with the *RAP1GDS1* coding sequence in the genome.

In summary bands 3, 4, 5 and 6 were all smaller than band 1, which contained the full length *RAP1GDS1* sequence, and diverged from the *RAP1GDS1* sequence at positions corresponding in sequence to introns. It is possible that they are not representative of abundantly expressed transcripts, but are artefacts of 3' RACE caused by the T rich adapter primer used for reverse transcription binding to A rich sequences in introns of *RAP1GDS1* remaining in a partially spliced intermediates of *NRG*. In support of this hypothesis is the fact that subsequent Northern analysis with cDNA probes for *NUP98* and *RAP1GDS1* (Section 4.4) did not detect transcripts corresponding in size to the 3' RACE products of bands 3, 4, 5 and 6.



**Table 4.1: Sequence characterisation of 3' RACE products.**

3' RACE products were all from patient 1. All had the *NUP98* sequence up to and including exon B and thus the size of the RACE product is derived from *RAP1GDS1* cDNA and introns. *RAP1GDS1* sequence is numbered from the beginning of the coding region (Genbank accession number X63465). The *NRG* sequence (band 1) and band 4 are described in more detail in the text.

| Clone  | Size   | Details  |
|--------|--------|--|
| Band 1 | 2.4 kb | <i>NRG</i> ; <i>NUP98</i> sequence up to and including nt 1552 of U41815 immediately followed by <i>RAP1GDS1</i> sequence beginning at nt 5.   |
| Band 2 | 2.1 kb | <i>NUP98</i> ; corresponding to 4.05 kb transcript.  |
| Band 3 | 2 kb   | truncated <i>NRG</i> ; homology to <i>RAP1GDS1</i> terminates at nt 112 and is immediately followed by GTAAG. This clone has a CAG inserted between <i>NUP98</i> and <i>RAP1GDS1</i> .   |
| Band 4 | 1.1 kb | truncated <i>NRG</i> ; homology to <i>RAP1GDS1</i> sequence terminates at nt 112 and is immediately followed by GTAAG. Contains a portion of the intronic sequence present in band 3.  |
| Band 5 | 0.9 kb | truncated <i>NRG</i> ; homology to <i>RAP1GDS1</i> terminates at nt 112 as for bands 3 and 4, but this is not followed by GTAAG. The sequence 3' of nt 112 was different to bands 3 and 4, it contained a MER2 repetitive element. |
| Band 5 | 0.9 kb | truncated <i>NRG</i> ; identical to above band 5 but CAG inserted between <i>NUP98</i> and <i>RAP1GDS1</i> .   |
| Band 5 | 0.9 kb | truncated <i>NRG</i> ; homology to <i>RAP1GDS1</i> ceases at nt 361 and is not immediately followed by GTAAG. Part of the sequence 3' of nt 361 is homologous to the AluSb2 subfamily.   |
| Band 6 | 0.8 kb | truncated <i>NRG</i> ; homology to <i>RAP1GDS1</i> ceases at nt 490 and this is immediately followed by GTAAG.   |

Despite the fact that these products did not represent abundantly expressed full length transcripts, their characterisation was useful. By determining the point of their sequence

divergence from the published *RAP1GDS1* sequence, some of the exon/intron boundaries in *RAP1GDS1* were deduced. These boundaries are described in Section 5.1.1.

### 4.3 RT-PCR

RT-PCR of t(4;11) presentation RNA using primers flanking the *NUP98-RAP1GDS1* junction (N1265F + R108R) gave a product of the expected size (395 bp) confirming that a *NRG* fusion mRNA was formed (Fig. 4.4b). No bands were seen in the PBMNC from normal controls.

During the identification of the t(4;11) breakpoint genes, another two T-cell ALL patients, whose karyotype also included the (4;11) translocation, became available for testing. For the remainder of this thesis, the initial t(4;11) patient will be referred to as patient 1, and the other two patients will be referred to as patients 2 and 3 (Table 4.2). Presentation samples of PBMNC from patients 2 and 3 were kindly provided by Dr Bik To (Institute of Medical and Veterinary Sciences, Adelaide, SA) and Dr Alec Morley (Flinders Medical Centre, Adelaide, SA) respectively. RNA from these two samples was tested for expression of *NRG* mRNA by RT-PCR (Fig. 4.4b). Patient 2 was clearly positive with a RT-PCR product of identical size to patient 1. Patient 3 had a smaller RT-PCR product of 162 bp. Sequencing revealed that patient 3 had a novel in-frame fusion of *NUP98* to *RAP1GDS1* with the *NUP98* breakpoint immediately preceding exon A and an identical *RAP1GDS1* junction (nt 5 of the coding sequence) to patients 1 and 2 (Fig. 4.4a). This transcript, denoted as *NRG2*, also maintains the first aspartic acid in the *RAP1GDS1* sequence as once again the 3' *NUP98* nt is a G.

The complexity of minor bands seen with all 3 patients in the *NRG* RT-PCR is a repeatable observation. Whereas some of the faint upper bands in patient 3 appear to be the same size as the *NRG* RT-PCR products in patients 1 and 2, they do not contain *NUP98* exon B as shown by hybridisation with the N1511F oligonucleotide (Fig. 4.4c), confirming that *NRG2* is not just an alternatively spliced version of *NRG*.

**Table 4.2: Clinical features, cytogenetics, immunophenotype and gene rearrangements of patients with a t(4;11)(q21p14-15).**

Details of the three patients reported in this study (1-3) and three patients from the literature (4-6) are presented. All features except for survival are at presentation. The comparatively long survival of patient 1 is at least in part due to the patient having undergone two allogeneic transplantations. The survival time of patient 5 is unknown but is at least 25 months. Where surface markers are positive in less than 50% of cells, the values are indicated in parentheses. Rearrangements of the heavy chain of immunoglobulin (IgH) and of the T-cell receptor  $\gamma$  gene (TCR $\gamma$ ) are indicated as G (germline -no rearrangement) or R (rearranged). Reference numbers are the same as the ones in Hussey *et al.*, 1999 (see Appendix).


| Patient | Sex /Age | WCC x 10 <sup>9</sup> /L | Diagnosis | Survival (months) | Karyotype   | Immunophenotype   | Gene rearrangement          | Reference        |
|---------|----------|--------------------------|-----------|-------------------|---|---|-----------------------------|------------------|
| 1       | M/21     | 423                      | ALL L1    | 43                | 46,XY,t(4;11)(q21p15)+2mar (presentation)<br>48,XY,t(4;11),-7,-7,-9,-9,11p+, 17p+,-20,-21,-21+9 mar (final relapse) | CD2+, CD3+ (30%), CD4-, CD5+, CD7+, CD8-, CD10+, CD11b (14%), CD14-, CD19-, CD20-, CD33+ (34%), CD34+, CD71+, HLA-DR+ | IgH (R)<br>TCR $\gamma$ (R) | 7<br>This report |
| 2       | F/25     | 1.83                     | ALL L1    | 1                 | 46,XX,t(4;11)(q21p14-15), del(12)(p13),+del(13)(q12q14)   | CD2+ (13%), CD3-, CD4-, CD5+ (13%), CD7+, CD8-, CD10+ (14%), CD13+ (5%), CD14-, CD19-, CD33+ (18%), CD34+ (5%)        | IgH (G)<br>TCR $\gamma$ (R) | This report      |
| 3       | M/49     | 169                      | ALL L2    | 14                | 46,XY,t(4;11)(q21p15), del(5)(q13q31)   | CD2-, CD4-, CD5+, CD7+, CD8-, CD10+ (9%), CD19-, CD34+  | IgH (G)<br>TCR $\gamma$ (R) | This report      |
| 4       | M/14     | 1.4                      | ALL L2    | 1                 | 46,XY,t(4;11)(q21p14),12p-/46,XY,t(4;11)(q21p14)  | CD2-, CD5+, CD10-, CD15+, pan-T+, TdT+  | ND                          | 21               |
| 5       | M/40     | 127                      | ALL       | 21                | 46,XY,t(4;11)(q21p15)   | T (surface markers not reported)  | ND                          | 22               |
| 6       | F/6      | 49                       | ALL L2    | 25+               | 46,XX,t(4;11)(q21p14-15)  | CD2+, CD5-, CD7+, CD10-, CD11b+, CD13+, CD14-, CD15-, CD19-, CD20-, CD22-, CD33+, CD36-, HLA-DR+, TdT+                | ND                          | 23               |

Expression of the complementary fusion cDNA, *RAP1GDS-NUP98* (*RGN*), was analysed by RT-PCR. Primers (R5'UTRF + N1848R) that could amplify *RGN* from all three patients showed that *RGN* is only expressed in patient 3 (Fig. 4.4b-c).

Some of the RACE products of patient 1 showed an insertion of the trinucleotide CAG at the *NUP98-RAP1GDS1* junction. The variable insertion of CAG was also seen in RT-PCR products from all three patients (data not shown). This insertion is most likely due to alternative splicing of intronic sequence immediately adjacent to an exon. As there are two distinct *NUP98* breakpoint regions in our t(4;11) patients, the CAG insertion probably comes from the *RAP1GDS1* intron in which the breakpoint occurs. This hypothesis was confirmed by aligning the *RAP1GDS1* sequence from *NRG* with the sequence of a *RAP1GDS1* genomic clone deposited into Genbank during the writing of this thesis (Genbank accession AC019077). The CAG conforms to the consensus sequence YAG (Y is a pyrimidine) of the 3' end of an intron (Moore *et al.*, 1993). Alternative splicing involving a single trinucleotide has previously been reported for the *c-kit* gene (Crosier *et al.*, 1993).

#### **4.4 Northern analysis**

The 3' RACE experiments discussed above showed that the *NUP98* and *NRG* transcripts were almost identical in size. This made it highly likely that one of the reasons that a hybrid transcript was not detected in previous Northern analysis was due to comigration of these transcripts. In order to overcome this problem, the Northern analysis was repeated but the RNA was electrophoresed further through the gel to achieve a higher degree of resolution. The quality of RNA isolated from the patient 1 presentation sample was improved by culturing the cells for 2 hours after thawing prior to RNA isolation. Remission and relapse specimens from patient 1 and presentation specimens from patients 2 and 3 were also included in this Northern analysis.



Rather than using the 586 bp *NUP98* probe described previously, a 1084 bp *NUP98* cDNA probe was used for Northern analysis (Fig. 4.5a). The probe was made by RT-PCR using Superscript II (Section 2.2.6.7) the Expand™ Long template PCR system. Primers N301F + N1384R were used so that the probe only hybridised to the portion of *NUP98* present in *NRG* of patients 1 and 2, and not to any *NUP98* sequence absent from *NRG*. Therefore, the relative intensities of the *NRG* and *NUP98* transcripts are a true representation of their relative levels of expression. Only the last 66 bases at the 3' end of this probe are homologous to the sequence 3' of the *NUP98* breakpoint in patient 3. Given that only 5 µg of total RNA per lane was used for this Northern analysis, 66 bp of probe is not likely to result in any significant signal compared to the 1018 bp of probe sequence present in *NRG2*. Therefore the relative intensities of the *NRG2* and *NUP98* transcripts in patient 3 are also a true representation of their relative levels of expression. The normal controls show 4.05 kb and 7.25 kb bands. The 4.4 kb *NRG* transcript can be seen above the 4.05 kb *NUP98* transcript for the presentation samples of patients 1 and 2. In patient 3, the *NRG2* transcript cannot readily be seen as it migrates just above the normal *NUP98* band. *NRG* is not seen in the remission sample from patient 1. The relapse specimen from the same patient shows markedly increased *NRG* expression compared to the endogenous *NUP98*. The increased *NRG* expression in the relapse specimen may be related to the addition to the short arm of the previously normal chromosome 11 (Table 4.2), or it may reflect selection pressure for increased *NRG* expression.

A second new transcript of approximately 5.8 kb was seen in the presentation and relapse samples of patient 1. This band is also present in patient 2 but not discernible on Figure 4.5a. Patient 3 showed a 5.5 kb transcript. The shorter size corresponds approximately to the size difference (233 bp) between *NRG* and *NRG2*.

## Figure 4.5 Northern analysis of *NRG* expression

(a) Hybridisation using a *NUP98* cDNA probe (see text for details).

(b) Hybridisation of the same membrane with a *RAP1GDS1* cDNA probe (see text for details).

(c) 18S rRNA from the ethidium bromide stained gel prior to transfer.

RNA was isolated from two normal donors and from the three t(4;11) patients (P1-P3). Each lane contains 5 µg of total RNA from peripheral blood mononuclear cells, except that lane 4 contains 5 µg of total RNA from bone marrow. The lane marked **M** is a RNA ladder (Promega). The band in **Lane M** in panel C is marker and not 18S RNA. **N** indicates the *NUP98* 4.05 and 7.25 kb bands. The 7.25 kb band is a precursor that also contains the *NUP96* coding sequence (Fontoura *et al.*, 1999). **RG** indicates the 2.8 and 4.1 kb *RAP1GDS1* bands. *NRG* indicates the 4.4 kb *NRG* transcript. *NRG2* in patient 3 is not indicated as it is not distinguishable from the 4.05 *NUP98* and 4.1 kb *RAP1GDS1* bands. The arrowheads indicate higher molecular weight transcripts which hybridise with both the *NUP98* and *RAP1GDS1* probes.

**Lane 1** donor 1

**Lane 2** donor 2

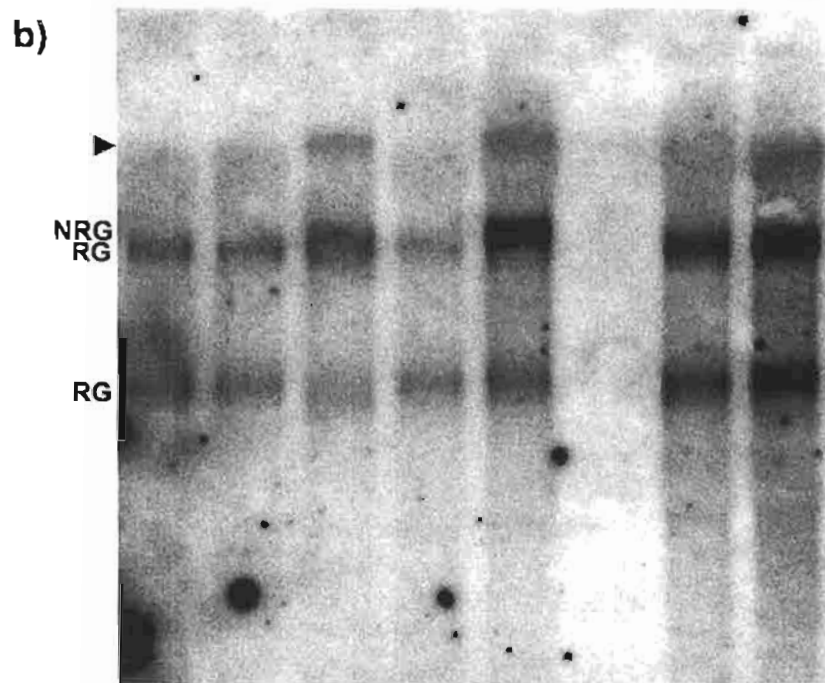
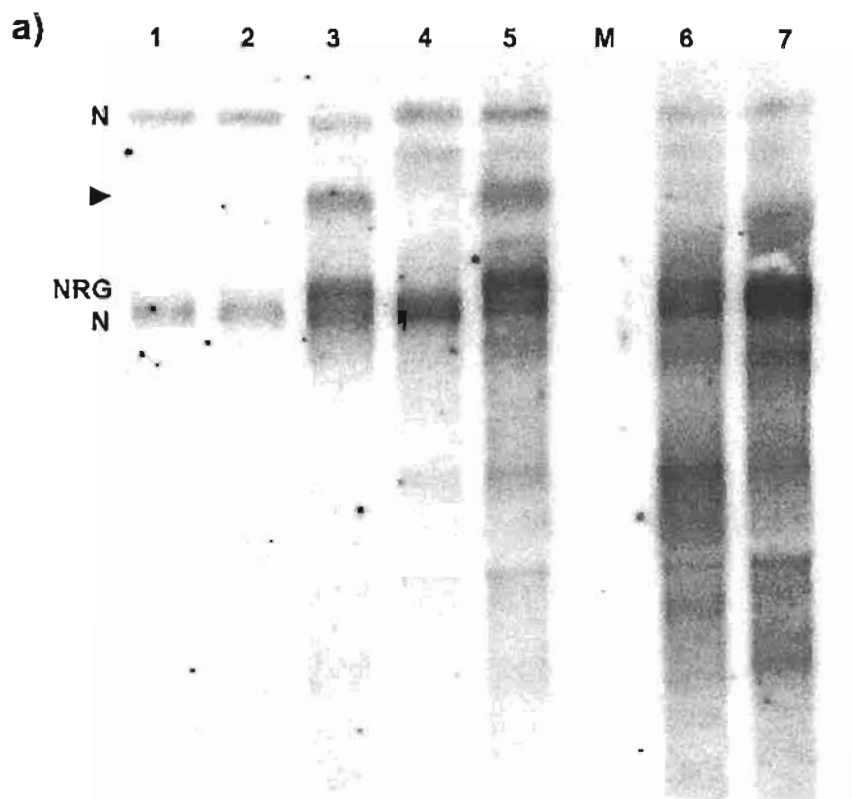
**Lane 3** t(4;11) patient 1 presentation

**Lane 4** t(4;11) patient 1 remission

**Lane 5** t(4;11) patient 1 relapse

**Lane 6** t(4;11) patient 2 presentation

**Lane 7** t(4;11) patient 3 presentation



A cDNA probe was created by RT-PCR, using Superscript II and primers RG1F + RG1677R, to assess *RAP1GDS1* expression by Northern analysis. *RAP1GDS1* shows 2.8 kb and 4.1 kb transcripts in all tissues tested (Fig. 4.6). When, the patient Northern was probed with the *RAP1GDS1* probe, the 4.1 kb transcript was visible as a distinct band slightly lower than the *NRG* transcript, although the two bands are not readily distinguishable after photo-reproduction (Fig. 4.5b). The 5.8 and 5.5 kb bands are present in the patient samples confirming that they are *NRG* transcripts. They are probably generated by the same mechanism that generates the upper 4.1 kb *RAP1GDS1* transcript. If the 5.5 kb transcript in patients 1 and 2 and the 5.8 kb transcript in patient 3 are *NRG* transcripts then they should have been visible as 3' RACE products. A RACE product corresponding in size to the 5.5 kb transcript is not visible on the agarose gels shown in Figure 4.3a and Figure 4.3b. However, hybridisation of an oligonucleotide probe for the first coding exon of *RAP1GDS1* (R108R) to a Southern blot of patient 1 *NUP98* 3' RACE products detected a product significantly higher than *NRG* (Fig. 4.3c) which corresponds in size to the 5.5 kb transcript in patient 1.

## 4.5 Discussion

Patient 1 was originally reported as a t(4;11)(q21;p14-15) translocation in a patient with T-cell ALL (Hardingham *et al.*, 1991). Molecular analysis then localised the chromosome 11 breakpoint to 11p15.5 (Kalatzis *et al.*, 1993). Subsequently, two further T-cell ALL patients, referred to patients 2 and 3 and karyotyped as t(4;11)(q21;p14-15) and t(4;11)(q21;p15) respectively, became available for analysis. Three other patients have been reported with either a t(4;11)(q21;p14-15) or a t(4;11)(q21;p15) as the primary translocation (Bloomfield *et al.*, 1986; Inoue *et al.*, 1985; Pui *et al.*, 1991) The clinical data, cytogenetics and immunophenotype of all six patients is summarised in Table 4.2.



**Figure 4.6 Multiple tissue Northern analysis of *RAP1GDS1***

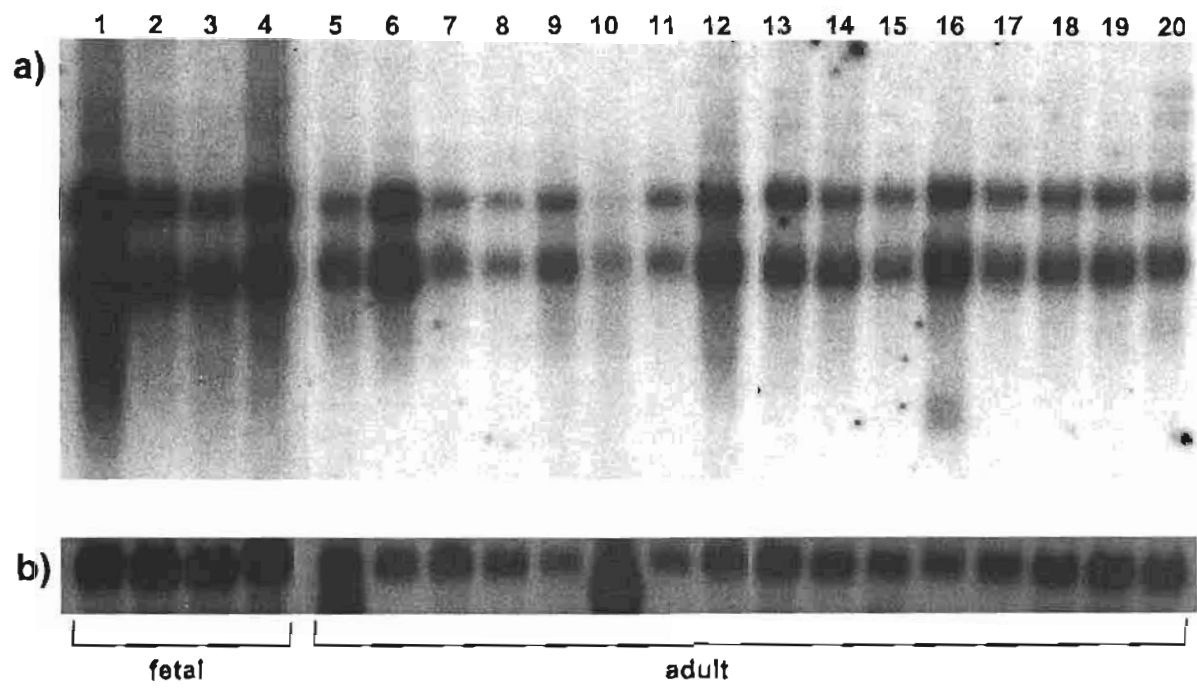
**(a)** Hybridisation with a *RAP1GDS1* cDNA probe shows two predominant bands of 4.1 kb and 2.8 kb.

**(b)** Hybridisation with a  $\beta$ -actin cDNA probe (Clontech).

Each lane contains 2  $\mu$ g of polyA RNA.

Lanes 1-4 contain RNA from the following foetal tissues: **Lane 1** brain, **Lane 2** lung, **Lane 3** liver, **Lane 4** kidney.

Lanes 5-20 contain RNA from the following adult tissues: **Lane 5** heart, **Lane 6** brain, **Lane 7** placenta, **Lane 8** lung, **Lane 9** liver, **Lane 10** skeletal muscle, **Lane 11** kidney, **Lane 12** pancreas, **Lane 13** spleen, **Lane 14** thymus, **Lane 15** prostate, **Lane 16** testis, **Lane 17** ovary, **Lane 18** small intestine, **Lane 19** colon (mucosal lining), **Lane 20** PBMNC.



#### **4.5.1 Common features of t(4;11) ALL patients**

Whereas different surface markers have been tested in each individual, the following generalisations can be drawn; (i) the cytochemistry and surface markers of all six patients are consistent with T-cell ALL, (ii) the leukaemic cells are positive for CD7 and CD5, usually positive for CD2, but negative for CD4 and CD8, (iii) CD10 is often positive in a proportion of the cells and (iv) most express one or more of the myeloid markers CD11b, CD13 and CD33 in a proportion of the cells. None of the six patients with the primary translocation were infants. They ranged in age from 6 - 53 with a preponderance of younger individuals as is typical for T-cell ALL (Catovsky, 1990). All had a fairly short survival after diagnosis. Patient 1, who showed the longest survival, underwent two matched allogeneic bone marrow transplants but relapsed with aggressive disease on both occasions.

Four of the six patients presented with additional karyotypic rearrangements (Table 4.2). This may account for some of the differences between their clinical pictures. Interestingly, the two patients who presented with a very low white cell count both had a 12p deletion.

#### **4.5.2 t(4;11)(q21;p15) fuses the *NUP98* and *RAP1GDS1* genes**

*NUP98* was identified as the chromosome 11 breakpoint gene by Southern analysis of patient 1 DNA and PCR analysis of somatic cell hybrids containing the derivative chromosomes of patient 1. Southern analysis using a *NUP98* cDNA probe flanking the reported t(7;11)(p15;p15) breakpoint detected a *NUP98* rearrangement in leukaemic DNA specimens from patient 1. Subsequently it was shown that exons B and C of *NUP98* were found on the der(11) and der(4) chromosomes respectively, thereby mapping the breakpoint to the intron between exons B and C. This confirms the previously reported orientation of *NUP98* with regard to the centromere (Borrow *et al.*, 1996).

As the principal transcript in the other reported *NUP98* translocations fuses the 5' end of the *NUP98* gene in frame to the 3' end of a second gene, 3' RACE was used and identified the *RAP1GDS1* gene as the 3' partner. *RAP1GDS1* has previously been mapped to 4q21-25 (Riess *et al.*, 1993). Its involvement in the (4;11) translocation has confirmed this mapping, and further submapped *RAP1GDS1* to the 4q21 region.

RT-PCR showed that *NUP98-RAP1GDS1* (*NRG*) fusion mRNAs were present in patients 1, 2 and 3. Sequencing showed that the same *RAP1GDS1* sequence, starting at nt 5 of the coding region, was present in all three patients (Fig. 4.4a). Patients 1 and 2 had an identical fusion mRNA containing the 5' sequence of *NUP98* up to and including exon B whereas patient 3 lacked *NUP98* exons A and B.

The absence of the reciprocal *RGN* transcript in patients 1 and 2 (Fig. 4.4b-c) indicates that *NRG* is the leukaemia associated transcript. It is unclear why the reciprocal transcript is absent as *RAP1GDS1* is universally expressed and *RGN* is under the control of the *RAP1GDS1* promoter. A similar situation has been observed for the *BCR-ABL* translocation in which the reciprocal *ABL-BCR* transcript is not expressed in all CML patients although *ABL* is also universally expressed (Melo *et al.*, 1993a).

### **4.5.3 *NUP98* breakpoints**

Prior to publication of the manuscript (Hussey *et al.*, 1999), breakpoints in *NUP98* had been reported to occur between exons A and B (Kwong and Pang, 1999), or between exons B and C (Arai *et al.*, 1997; Borrow *et al.*, 1996; Kwong and Pang, 1999; Nakamura *et al.*, 1996a; Nakamura *et al.*, 1999; Raza-Egilmez *et al.*, 1998; Wong *et al.*, 1999), or between exons D and E (Arai *et al.*, 1997), with the predominant breakpoint occurring between exons B and C.

After publication of the manuscript (Hussey *et al.*, 1999), another *NUP98* breakpoint between exons C and D was reported (Ahuja *et al.*, 1999).

The breakpoint in patient 3 (in the intron preceding exon A) is the most proximal *NUP98* breakpoint reported. The more proximal breakpoint position is consistent with the Northern results where the *NRG2* fusion band is almost identically sized to the 4.05 kb *NUP98* transcript. *NRG2* is 233 bp shorter than *NRG* on account of the missing exons A and B. After publication of the manuscript (Hussey *et al.*, 1999), a *NUP98* breakpoint in the same intron as that disrupted in patient 3, was reported in a case of the (7;11)(p15;p15) translocation (Hatano *et al.*, 1999).

#### **4.5.4 Fusion partners of *NUP98***

To date, a total of 6 fusion partners have been described in various translocations involving *NUP98* (Table 4.3). Another 3 or 4 as yet unidentified fusion partners are implicated by the involvement of *NUP98* in other translocations. All of these *NUP98* fusions are associated with myeloid malignancies except for the *NRG* fusion which is associated with ALL.

Nup98 is a component of the NPC, involved in the export of RNA and protein from the nucleus (Bachi *et al.*, 2000; Powers *et al.*, 1997; Pritchard *et al.*, 1999; Zolotukhin and Felber, 1999). The 6 known fusion partners of *NUP98* are functionally diverse. *HOXA9*, *HOXD13* and *PMX1* code for transcription factors required for normal development (Lawrence *et al.*, 1997; Martin *et al.*, 1995; Muragaki *et al.*, 1996). *DDX10* codes for a putative RNA helicase with a probable role in ribosome biogenesis (Savitsky *et al.*, 1996). *TOP1* codes for DNA topoisomerase I which creates transient single-stranded DNA nicks to facilitate DNA replication, transcription, recombination and chromosome condensation (reviewed in Wang, 1996). Finally, *RAP1GDS1* codes for a guanine nucleotide exchange factor (Mizuno *et al.*, 1991).

**Table 4.3 Translocations involving the *NUP98* gene.**

Therapy related AML is abbreviated as t-AML. Chronic myelomonocytic leukaemia is abbreviated as CMMoL. The *NUP98-PMX1* fusion, arising from t(1;11)(q23;p15), was reported after the t(1;11)(q21;q15) translocation in which the fusion partner was not determined. Although not been suggested by others, it is possible that these two translocations are the same and were reported as cytogenetically distinct because of variation in the interpretations of the same translocation.

| Translocation     | Partner Gene            | Disease                                | Primary ref(s) in which <i>NUP98</i> disruption was identified   |
|-------------------|-------------------------|--|--|
| t(7;11)(p15;p15)  | <i>HOXA9</i>            | De novo and t-AML/MDS<br>De novo-CMMoL | (Borrow <i>et al.</i> , 1996;<br>Nakamura <i>et al.</i> , 1996a) |
| inv(11)(p15;q22)  | <i>DDX10</i>            | De novo and t-AML/MDS                  | (Arai <i>et al.</i> , 1997)                                      |
| t(4;11)(q21;p15)  | <i>RAP1GDS1</i>         | De novo T-cell ALL                     | (Hussey <i>et al.</i> , 1999)                                    |
| t(2;11)(q31;p15)  | <i>HOXD13</i>           | t-AML                                  | (Raza-Egilmez <i>et al.</i> , 1998)                              |
| t(11;12)(p15;q13) | unknown                 | De novo and t-AML                      | (Kobzev and Rowley, 1999;<br>Nishiyama <i>et al.</i> , 1999)     |
| t(11;17)(p15;q21) | unknown                 | t-AML                                  | (Nishiyama <i>et al.</i> , 1999)                                 |
| t(1;11)(q23;p15)  | <i>PMX1</i>             | t-AML                                  | (Nakamura <i>et al.</i> , 1999)                                  |
| t(1;11)(q21;p15)  | unknown<br><i>?PMX1</i> | CML                                    | (Kobzev and Rowley, 1999)  |
| add(11)(p15)      | unknown                 | t-AML/MDS                              | (Nishiyama <i>et al.</i> , 1999)                                 |
| t(11;20)(p15;q11) | <i>TOP1</i>             | De novo and t-AML/MDS                  | (Ahuja <i>et al.</i> , 1999)                                     |

#### 4.5.5 Retention of the FG repeat region in fusions involving nucleoporins

Another nucleoporin gene, *NUP214*, is also involved in translocations in leukaemia. *NUP214*, also known as *CAN*, is fused to either the *DEK* gene or to the *SET* gene in cases of AML (von Lindern *et al.*, 1992a; von Lindern *et al.*, 1992b).

Both *nup98* and *nup214* contain multiple FG repeats. The FG repeats are presumed contact sites for multi-protein transport complexes that mediate bi-directional transport across the nuclear pores (Radu *et al.*, 1995b). All known *nup98* and *nup214* fusions retain the majority of the FG repeats (von Lindern *et al.*, 1992a; von Lindern *et al.*, 1992b) and references in Table 4.3). The FG repeats are also retained in the three patients reported in this thesis (Fig. 4.7). Patient 3, who has the most 5' breakpoint yet reported, still has 30 of the 37 FG repeats.

The FG repeat-containing *nup98* portion of the *nup98-hoxa9* fusion protein acts as a potent activator of *hoxa9* activity by recruiting the *cbp* and *p300* transcriptional co-activators (Kasper *et al.*, 1999). The interaction of *nup98-hoxa9* with *cbp/p300* is correlated to the transforming activity of *nup98-hoxa9*. The transforming ability is retained when the FG repeat region from *nup98* is exchanged for that of *nup214* which directly implicates the FG repeats in the transforming activity. Not all of the FG repeats are required to interact with *cbp/p300* or to transform as a *nup98-hoxa9* splice variant with twenty FG repeats still retains transforming ability (Kasper *et al.*, 1999).

#### 4.5.6 The role of smgGDS

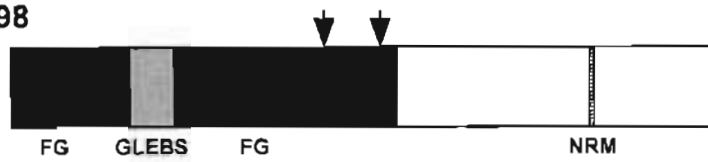
The entire coding region, except for the start codon of *RAP1GDS1*, is retained in the *NRG* and *NRG2* transcripts. The product of *RAP1GDS1*, usually referred to as smgGDS, has guanine nucleotide exchange factor (GEF) activity (Mizuno *et al.*, 1991). GEFs stimulate or inhibit

**Figure 4.7 Schematic representation of the nup98, smgGDS, nrg and nrg2 proteins**

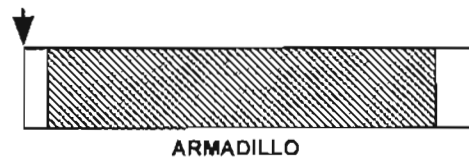
Vertical arrowheads represent breakpoints in nup98 and smgGDS. FG, FG (phenylalanine-glycine) repeat rich areas; GLEBS, GLEBS (Gle2p-binding motif) like motif (Pritchard *et al.*, 1999); NRM, nucleoporin RNA binding motif (Radu *et al.*, 1995); ARMADILLO, tandem armadillo repeats.



**NUP98**



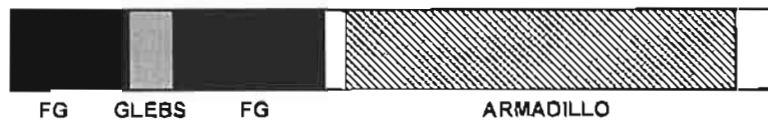
**smgGDS**



**NRG**



**NRG2**



exchange of GDP for GTP at small GTPase proteins to convert the inactive GDP bound form to the active GTP bound form. SmgGDS was first reported as a stimulator of GDP/GTP exchange for rap1a, then called smg p21a (Yamamoto *et al.*, 1990). SmgGDS also acts on rap1b as well as other small GTPases including K-ras, rac1, rac2, rhoA and ralB (Chuang *et al.*, 1994; Iouzalén *et al.*, 1998; Mizuno *et al.*, 1991). Interestingly, rap1a and K-ras are antagonistic as the protein smg p21a/rap1a was first identified as Krev-1 which has the ability to revert K-ras transformed NIH 3T3 fibroblasts (Kitayama *et al.*, 1989). However, rap1 is unlikely to be the principal target of smgGDS as smgGDS co-operates with K-ras in transformation (Fujioka *et al.*, 1992). RhoA and rac2 have been reported to be more important targets for smgGDS than rap1a (Chuang *et al.*, 1994; Strassheim *et al.*, 2000).

SmgGDS is structurally unique among the GEFs as it shows no homology to other GEFs and is composed largely of tandem repeats of the 43 amino acid armadillo motif (Fig. 4.7) (Peifer *et al.*, 1994). The armadillo motif was originally found in the *Drosophila melanogaster* armadillo gene and its vertebrate homologues  $\beta$ -catenin and plakoglobin (McCrea *et al.*, 1991; Peifer and Wieschaus, 1990). Subsequently it was identified in a number of other genes which contain tandem repeats of armadillo (Peifer *et al.*, 1994), including importin  $\alpha$  (Gorlich *et al.*, 1996). It has been suggested that armadillo repeats mediate protein-protein interactions (Peifer *et al.*, 1994).

Determining the cellular location of nrg will be critical in determining its role in malignancy (see chapter 5). SmgGDS normally interacts with membrane bound and cytoplasmic ras superfamily GTPases. If the nrg hybrid protein is cytoplasmic, its function may involve alterations of signalling via ras family small GTPases. However, by analogy to other armadillo proteins like  $\beta$ -catenin and importin  $\alpha$ , smgGDS may have an as yet undescribed cytoplasmic-nuclear shuttling capacity. The armadillo repeats of smgGDS may lead it to

mimic  $\beta$ -catenin and interact with the transcription factors involved in the wingless signalling pathway (reviewed in Dierick and Bejsovec, 1999), or possibly other pathways. Alternatively, the amino-terminal end of nup98 might re-locate smgGDS to the nuclear pore so that the fusion protein may modify nuclear transport. As nrg contains an intact smgGDS sequence, it may act as a second GEF for ran in promoting nuclear transport. Finally, nrg may be located in the nucleus where it may modify transcription as happens with other nup98 fusion proteins (Kasper *et al.*, 1999; Nakamura *et al.*, 1999). Transcription factors that interact with the armadillo repeats may become coupled to transcription factors that interact with FG repeats.

The results presented in this chapter show that *NUP98* can be involved in T-cell ALL as well as myeloid malignancies. Moreover, the identification of 3 patients with the *NRG* fusion shows that the t(4;11)(q21;p15) is a recurrent translocation in T-cell ALL.

## **Chapter 5**

# **Cellular localisation of nrg and its components**

## Chapter 5

### Cellular localisation of nrg and its components

To gain insight into the transforming mechanisms of nrg, it was necessary to investigate its cellular location in comparison to that of its components ie (i) the amino-terminal portion of nup98 consisting of 37 FG repeats (called nup98t, where t stands for truncated) and (ii) the smgGDS protein. This chapter describes experiments in which amino-terminal green fluorescent protein (gfp) tagged proteins were expressed in NIH3T3 mouse fibroblast cells to investigate the cellular localisation patterns of nup98, nup98t, smgGDS and nrg.

The green fluorescent protein from *Aequorea victoria* folds into a structure called a chromophore which fluoresces brilliant green (509nm) when excited by blue light (395nm) (Ward *et al.*, 1980). Enhanced green fluorescent protein (egfp) is an engineered mutant which fluoresces 35-fold more intensely than wild type gfp (Cormack *et al.*, 1996; Yang *et al.*, 1996). The maximal excitation peak of egfp occurs at 490nm, which is virtually identical to the light emission of a confocal microscope (488nm), therefore egfp can be viewed by confocal microscopy. When cells that express proteins tagged with egfp are viewed using a confocal microscope, the subcellular location of the tagged proteins can be precisely determined. This is achieved due to the very thin (0.5 $\mu$ m) focal plane provided by a confocal microscope. A major advantage of using egfp to tag proteins is that it eliminates the need for cell fixation, permeabilisation and antibody staining and therefore allows one to examine the cells while they are alive and undisturbed. In recent years egfp has been used extensively as a tool for exploring the cellular location and trafficking patterns of various proteins. For simplicity, egfp will be referred to simply as gfp in this thesis.

## 5.1 Creating constructs for studying the cellular location of nrg

The key feature of this study was the creation of a plasmid construct consisting of an in frame fusion of *GFP* to the 5' end of the *NRG* coding sequence. Control plasmids fusing *GFP* to *NUP98*, *NUP98t* and *RAP1GDS1* were also made.

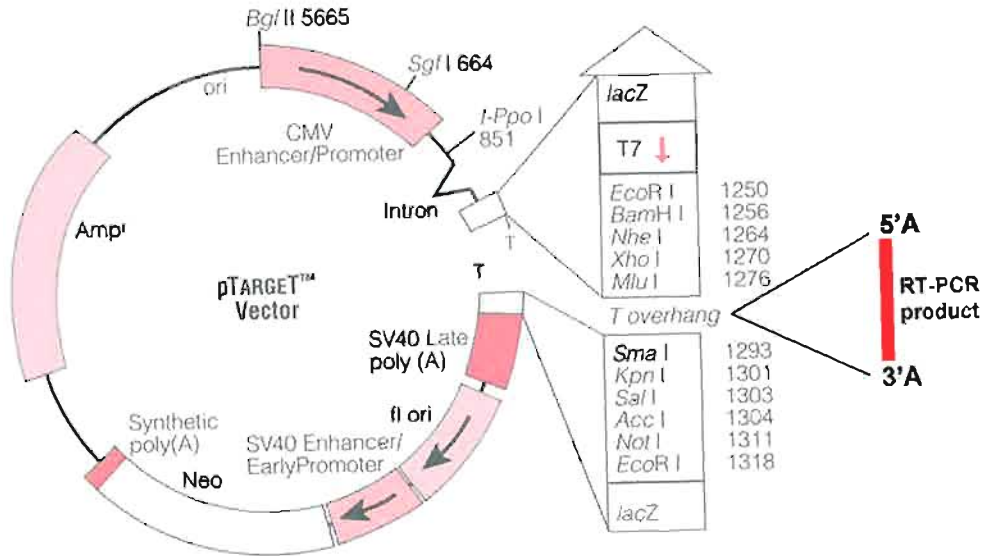
The first step in creating the expression constructs was to generate RT-PCR products encoding the relevant proteins. All RT-PCR products were made from normal donor RNA except for the *NRG* product which was made from patient 1 RNA. RNA was extracted from PBMNC and reverse transcribed with an oligo-dT primer (AP) and Superscript II to ensure that full length first strand cDNA was generated. High-Fidelity PCR (Section 2.2.6.2) was used to amplify the open reading frame (ORF) from the first strand cDNA (see Fig. 5.1 legend for the names of all primers used for ORF amplification). The PCR products were purified from a low melting point agarose gel and a portion of the products were restriction digested with *Hinf*I. The size of the *Hinf*I restriction fragments was compared to that predicted by the relevant Genbank entries. This comparison was done prior to cloning and sequencing to ensure that RT-PCR had resulted in amplification of the correct product.

RT-PCR products (except for *NUP98*-see Section 5.1.3) were then cloned into the pTarget™ mammalian expression vector which allows direct cloning of PCR products (Fig. 5.1). Cloning into pTarget™ also created a number of restriction sites flanking the insert which could be used for subcloning into a variety of tagging vectors including pEGFP-C2. Positive clones were identified and two for each different RT-PCR product were sequenced to confirm that the ORF was intact and no PCR induced errors were present. Any sequence deviation from the relevant Genbank entries was investigated by sequencing of the original RT-PCR product to determine whether the variation was a naturally occurring polymorphism or an incorrectly incorporated base. Some deviations were found to be due to misincorporation of a

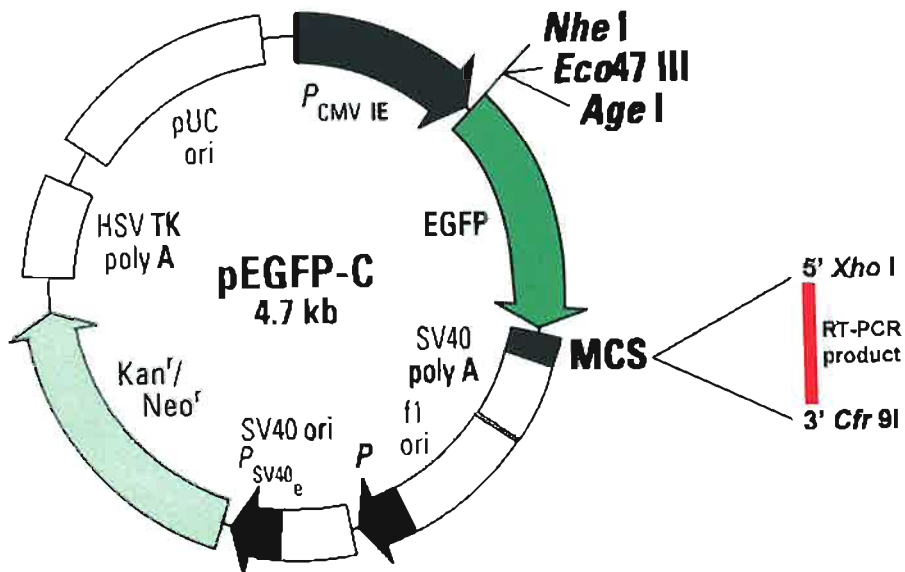
## Figure 5.1 Subcloning strategy for creating gfp tagged proteins

RT-PCR products were amplified with the following primers: *NUP98* *NXho* 1F + *NCfr* 91R, *NUP98t* N145F + *NtrnR*, *RAP1GDS1* R1F + R1677R, *NRG* N145F + R1677R. The products were cloned directly into pTarget (except *NUP98*, see text for details) using the A overhang generated by *Taq* polymerase and the T overhang on the pTarget. After sequence verification the insert was excised from pTarget with *Xho* I (near the 5' end of the insert) and *Cfr* 91I (near the 3' end of the insert). The restriction fragment was then ligated into the corresponding sites in pEGFP-C2 to create an in-frame fusion of GFP with the RT-PCR product. Pictures of the pTarget and pEGFP-C2 vectors were taken from the Promega and Clontech manuals respectively.

a)



b)





base, presumably due to polymerase error. However other variations were found to be naturally occurring and these are described below.

### **5.1.1 Discovery of a novel exon in the predominant isoform of *RAP1GDS1* in PBMNC**

The *RAP1GDS1* RT-PCR product was expected to be 1667 bp based on the *RAP1GDS1* Genbank entry X63465. Agarose gel electrophoresis of the RT-PCR products that contained *RAP1GDS1* sequence, ie *RAP1GDS1* and *NRG*, revealed a major PCR product approximately 150 bp larger than expected (Fig. 5.2a). *Hinf* I digestion of gel purified *RAP1GDS1* (and *NRG*) product resulted in two unpredicted fragments of 130 bp and 333 bp, and the absence of a predicted 316 bp fragment, however all other fragments were as predicted from Genbank entry X63465 (Fig. 5.2b). This difference in the *Hinf* I restriction fragments of the *RAP1GDS1* (and *NRG*) RT-PCR product suggested the presence of additional sequence to that of X63465. The RT-PCR product was cloned into pTarget and sequenced.

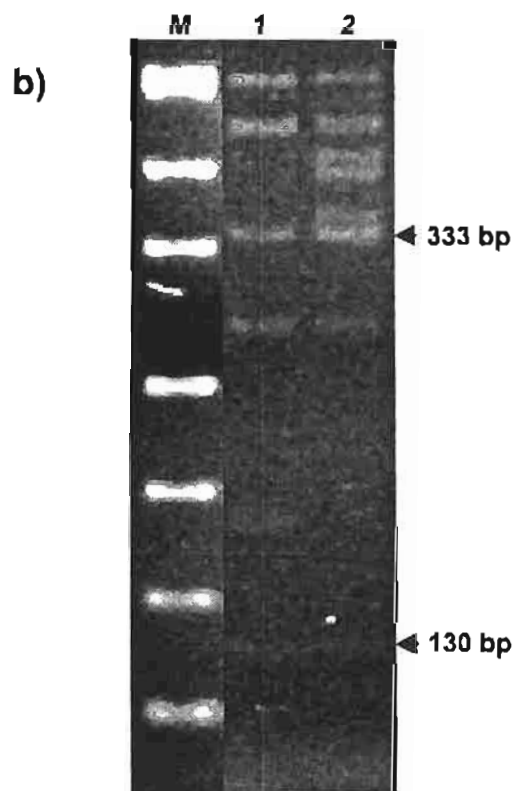
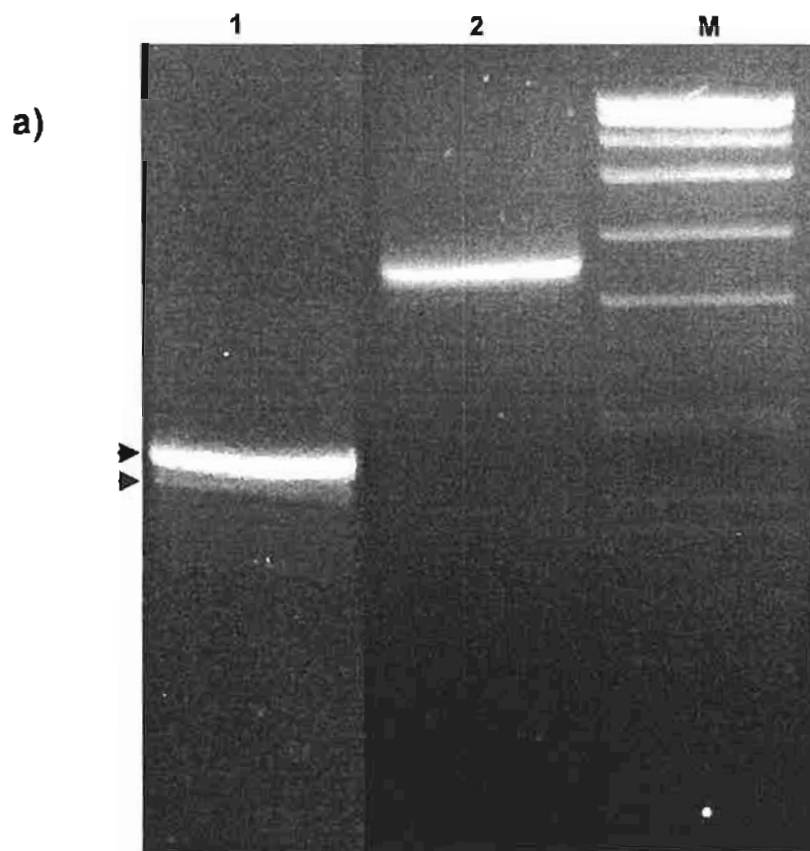
To determine if there was any evidence for the presence of naturally occurring additional (to X63465) sequence in *RAP1GDS1*, it was necessary to obtain all of the sequence data available in Genbank in the form of expressed sequence tags (ests) homologous to *RAP1GDS1*. Using Sequencher, a consensus sequence was assembled from the ests in the Unigene cluster for *RAP1GDS1*. This consensus sequence revealed the presence of an additional 147 bp between bases 361 and 362 of X63465. The predicted *Hinf* I digestion pattern of this consensus sequence resulted in the same restriction fragments as that obtained by *Hinf*I digestion of the actual *RAP1GDS1* RT-PCR product.

Initial sequencing of the cloned RT-PCR product using inward primers based on the pTarget sequence revealed that one of the two *RAP1GDS1* clones was missing the A from its start

**Figure 5.2 Agarose gel electrophoresis of *RAP1GDS1* and *NRG* RT-PCR products**

(a) RT-PCR products encoding the open reading frames of *RAP1GDS1* (**Lane 1**) and *NRG* (**Lane 2**) are shown prior to gel purification. **Lane M** contains the SPPI/*Eco* RI DNA size marker. The upper arrow in Lane 1 indicates the predominant *RAP1GDS1* product which is 147 bp larger than expected from Genbank X63465. The lower arrow indicates a minor product of 1677 bp which is the size predicted from Genbank X63465.

(b) *Hinf*I digests of the gel purified predominant RT-PCR products of (**Lane 1**) *RAP1GDS1* and (**Lane 2**) *NRG* from PBMNC. **Lane M** is the pUC19/*Hpa* II DNA size marker. The novel restriction fragments that were not predicted from the sequence of Genbank X63465 are indicated with arrow heads and sizes. In the *NRG* RT-PCR product two fragments of approximately identical size (333 bp) com-migrate, one of these fragments is derived from *NUP98* sequence and the other from the novel *RAP1GDS1* sequence.



codon. The full sequence of this clone was not determined because the absence of a start codon made it unsuitable for use in the localisation studies. However, enough sequence was obtained from this clone to cover the region containing the additional 147 bp. When the *RAP1GDS1* sequence from the two pTarget clones containing the RT-PCR product was aligned with the consensus sequence of the Unigene *RAP1GDS1* cluster, it was clear that the two clones contained an extra 147 bp in exact agreement with the consensus sequence. Sequencing also revealed that the *RAP1GDS1* portion of *NRG* contained the additional 147 bp, thereby explaining the apparent size discrepancy seen upon initial agarose gel electrophoresis of the RT-PCR products for *RAP1GDS1* and *NRG*.

The position of the additional 147 bp sequence corresponds to a predicted exon/exon junction (see Table 4.1), which suggested that it could be a self contained exon. RT-PCR with primers which flanked the extra sequence (R11F + R688R) gave a major product of 825 bp and two less abundant products of 678 bp and 550 bp (Fig. 5.3). The major 825 bp product contained all of the sequence predicted from Genbank entry X63465 as well as the extra 147 bp of sequence. The 678 bp product lacked the 147 bp sequence and was identical to Genbank entry X63465. This result confirmed that the extra 147 bp is an alternatively spliced exon that is contained in the most highly expressed isoform of *RAP1GDS1* mRNA expressed in PBMNC. The full coding sequence of the major isoform of *RAP1GDS1* in PBMNC was deposited into Genbank and assigned the accession number, AF237413.

The 550 bp band was actually a mixed population of two products, both of them lacking the 147 bp sequence present in the predominant *RAP1GDS1* isoform. One of the products was missing 126 bp of sequence corresponding to nt 236-361 of the Genbank entry X63465. The other product was missing 129 bp of sequence corresponding to nt 362-490 of Genbank entry X63465.

**Figure 5.3 The predominant isoform of *RAP1GDS1* in PBMNCs contains a novel 147 bp exon**

RT-PCR of RNA from normal donors, with primers flanking the additional 147 bp sequence, shows that the *RAP1GDS1* isoform containing the extra 147bp is predominant. The upper arrow indicates the predominant RT-PCR product of 825 bp which contains the 147bp of additional DNA. The 678 bp product lacks the 147 bp exon and corresponds to the sequence of Genbank X63465. The 550 bp band is a mixture of two alternatively spliced products as described in section 5.1.1. Control reactions containing no reverse transcriptase enzyme (minus RT) were also performed.

**Lane 1** SPPI/*Eco* RI DNA size marker

**Lane 2** pUC19/*Hpa* II DNA size marker

**Lane 3** H<sub>2</sub>O control

**Lane 4** H<sub>2</sub>O control

**Lane 5** DNA target PCR control

**Lane 6** No target RT control

**Lane 7** donor 1 RT-PCR

**Lane 8** donor 1 minus RT

**Lane 9** donor 2 RT-PCR

**Lane 10** donor 3 RT-PCR

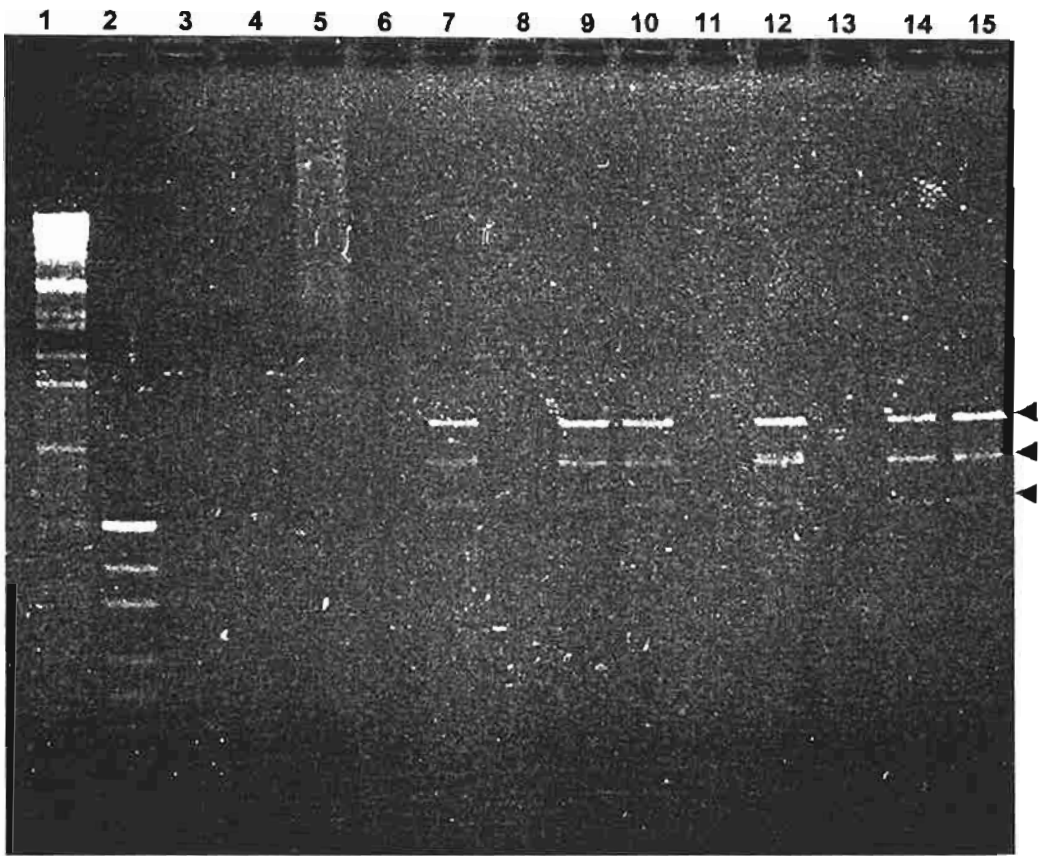
**Lane 11** donor 3 minus RT

**Lane 12** donor 4 RT-PCR

**Lane 13** donor 4 minus RT

**Lane 14** donor 5 RT-PCR

**Lane 15** donor 6 RT-PCR



The positions at which the sequence of the three RT-PCR products varied from that of Genbank X63465 implied the positions of three exon-exon boundaries in the *RAP1GDS1* cDNA, between nt 235-236, nt 361-362, and nt 490-491. The presence of an exon-exon boundary at nt 235-236 is in accordance with the boundary predicted from the sequence of a partially spliced form of *NRG* (see Section 5.1.2 below). The predicted exon-exon boundaries at nt 361-362 and 490-491 are in accordance with the boundaries predicted from the partially spliced *NRG* 3' RACE products, which also predicted a boundary at nt 112-113 (Table 4.1). The fact that the *RAP1GDS1* sequence in *NRG* begins at nt 5 of the *RAP1GDS1* cDNA sequence indicates yet another exon-exon boundary between nt 4-5. Alignment of the full length cDNA sequence of *RAP1GDS1* with that of a *RAP1GDS1* genomic clone, deposited into Genbank during the writing of this thesis (Genbank AC019077), has confirmed that these predicted exon-exon boundaries are correct.

#### **5.1.1.1 Other sequence variations in *RAP1GDS1***

Besides the alternatively spliced 147 bp exon, a number of other sequence variations were identified in the completely sequenced pTarget clone containing the *RAP1GDS1* RT-PCR product. The trinucleotide sequence CAG, which corresponds to nt 1154-1156 in Genbank entry X63465, was missing from the cloned sequence. To investigate the absence of these 3 bases, the sequence of this clone was aligned with the ests from the Unigene cluster for *RAP1GDS1*. Two of 11 ests lacked a CAG at this position whereas 9/11 contained a CAG. The result of the alignment confirmed that the absence of CAG is a naturally occurring sequence variation. This result is similar to that obtained during the initial identification of *NRG* and is due to alternative splicing of *RAP1GDS1* intronic sequence immediately adjacent to the first exon of *RAP1GDS1* in *NRG* (discussed in Section 4.3).

The cloned *RAP1GDS1* sequence also differed from Genbank X63465 at position 561 where it contained an A instead of a C. In three of three ests over this region the A was present at this position, therefore it was a naturally occurring variation rather than a PCR induced error. The sequence variation does not alter the amino acid sequence of smgGDS, as an isoleucine is coded regardless of whether an A or C is present at this position.

Yet another variation from Genbank X63465 occurred at position 1033 where an A was present instead of G. Three of three ests over this region contained the A, however the single nt substitution resulted in the coding of asparagine rather than the expected aspartic acid. This variation may represent an error in the Genbank X63465 sequence since all three ests were in agreement with the sequence of the cloned RT-PCR product.

The final sequence variation occurred at nt 1630 and 1631 of Genbank entry X63465. The bases at this position in the Genbank entry are C and G respectively whereas in the pTarget clone they are G and C. This results in the coding of alanine rather than the expected arginine predicted from Genbank X63465. Alignment of the cloned sequence with the *RAP1GDS1* Unigene cluster showed that six of six ests over this region had a G at position 1630. Four of the six ests had a C at position 1631 and the other two had a G at this position. This polymorphism at position 1631, indicated by the ests, codes for an alanine in both cases. The most common sequence represented by the *RAP1GDS1* ests corresponded to the cloned sequence.

### **5.1.2 Three fragment ligation for the creation of the gfp-nrg expression construct**

Both of the *NRG* clones that were sequenced had errors. One clone had a correct *RAP1GDS1* sequence, including the extra 147 bp, but the *NUP98* sequence of this clone contained two



single nt variations from Genbank entry U41815 for *NUP98*. These two variations were not present in the Unigene cluster for *NUP98*. Sequence analysis of the *NUP98* RT-PCR product revealed these base variations were absent and were therefore due to cloning of a product that contained PCR induced errors.

The other *NRG* clone contained a perfect *NUP98* sequence and its *RAP1GDS1* sequence was perfect until base 235. However, after base 235 the sequence diverged from that of Genbank entry X63465. The sequence immediately 5' of nt 235 was GTGAG, which is a match for the consensus sequence at the 5' end of an intron (Moore *et al.*, 1993). Further supporting evidence that this sequence is intronic was the presence of repeat sequences. The intronic sequence introduced a premature termination codon into *NRG* which would result in a protein with only one armadillo repeat and this clone was therefore unsuitable for the localisation studies.

After sequencing the two *NRG* clones described above and finding that both were unsuitable for subcloning into pEGFP-C2 there were two options available. One option was to sequence the remaining clones and hope that one was correct. The other option was to devise a subcloning strategy utilising the two sequenced clones to create a correct *NRG* clone. The latter method was chosen for the following reasons. Firstly, cloning with pTarget™ results in a higher than usual frequency of background white colonies containing vector with no insert (pTarget™ Mammalian Expression Vector System Technical Manual, Promega). In this study, the frequency of white colonies with no insert was disappointingly high (ie only seven of 20 plasmids contained inserts and one of these was in the anti-sense orientation). Secondly, the RT-PCR products were more error prone than expected, perhaps because the reverse transcription step introduced extra opportunity for the incorporation of incorrect bases. Thirdly, the identification of a variant containing a premature stop codon in one of two clones

that were sequenced indicated the risk of identifying more clones like this. Therefore, in the interest of time the correct parts of the two sequenced clones were ligated and subcloned into pEGFP-C2.

There is only one *Drd* I site present in the full length *NRG* cDNA, at the position corresponding to base 180 of *RAP1GDS1*. In the clone of *NRG* containing the intron, after base 235 in the *RAP1GDS1* sequence, there are an additional two *Drd* I sites. The subcloning strategy involved ligating the *Xho* I-*Drd* I fragment from the clone with the correct *NUP98* sequence to the *Drd* I-*Cfr* 9I fragment of the clone with the correct *RAP1GDS1* sequence (Fig. 5.4). This was done as a 3 fragment ligation directly into appropriately digested pEGFP-C2. The corrected clone was then checked by digesting with *Xho* I and *Cfr* 9I to ensure that the ligation had re-created these sites. In addition to this a portion of the *NUP98* region was sequenced to ensure that the correct *Xho* I-*Drd* I fragment had been inserted. To check the *RAP1GDS1* portion of the clone, the insert was amplified by PCR and digested with *Drd* I to confirm the presence of only one site.

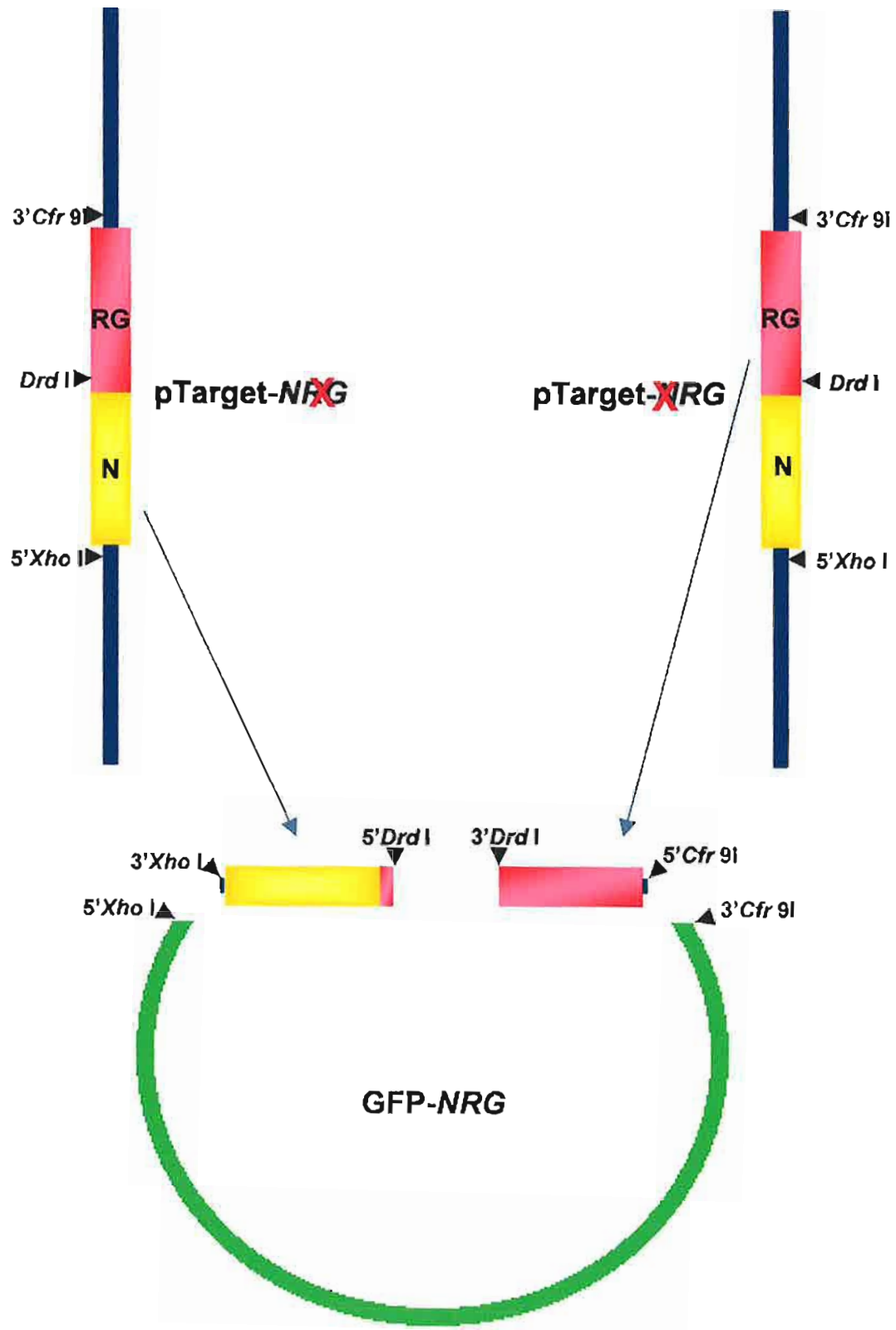
### 5.1.3 *NUP98* cloning

Full length *NUP98* was cloned directly into pEGFP-C2 rather than subcloning it from pTarget. An RT-PCR product encoding the *NUP98* ORF was amplified using a forward primer with an engineered *Xho* I site at its 5' end and a reverse primer with an engineered *Cfr* 9I site at its 5' end. The RT-PCR product was digested with *Xho* I and *Cfr* 9I prior to ligation into *Xho* I-*Cfr* 9I digested pEGFP-C2. Two plasmids containing the *NUP98* insert were sequenced and found to be identical. The sequence was a match for the *NUP98* cDNA sequence (Genbank accession U41815) except for the inclusion of 51 bp between nt 1318 and 1319. The presence of an intron between nt 1318-1319 of the *NUP98* cDNA sequence was indicated in a previous study (Arai *et al.*, 1997). Therefore, it was thought that the 51 bp

### Figure 5.4 Three fragment cloning strategy for correcting *NRG*

Diagrammatic representation of the cloning process used to create the GFP-*NRG* construct from two different pTarget-*NRG* clones each containing regions unsuitable for localisation studies (denoted by the red cross). Please see text for details.

There is only one *Drd* I site in the pTarget-*NRG* insert containing the correct *RAP1GDS1* sequence. However there are an additional two *Drd* I sites in the *RAP1GDS1* intronic sequence of the other pTarget-*NRG* insert, and for simplicity they have been omitted from the diagram. The pTarget vector also contains 6 *Drd* I sites not shown in this diagram.



insertion in the *NUP98* RT-PCR product resulted from partial splicing of intronic sequence. This was confirmed by aligning the sequence of the cloned *NUP98* RT-PCR product with that of a *NUP98* genomic clone deposited into Genbank during the writing of this thesis (Genbank AC018803). Sequence alignments revealed that the additional 51 bp in the cloned cDNA is an exact match for the genomic sequence present after nt 1318 of the cDNA. The sequence at the beginning of the 51 bp insertion is a match for the consensus sequence at the 5' end of an intron (Moore *et al.*, 1993), indicating that the 51 bp sequence is intronic. The *NUP98* genomic sequence present after the 51 bp insertion is a match for a consensus splice acceptor site (Moore *et al.*, 1993). This indicates that the 51 bp insertion in the *NUP98* RT-PCR product arose as a result of use of an alternate splice acceptor site.

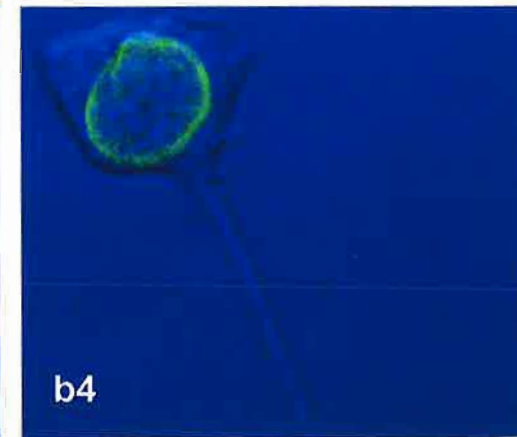
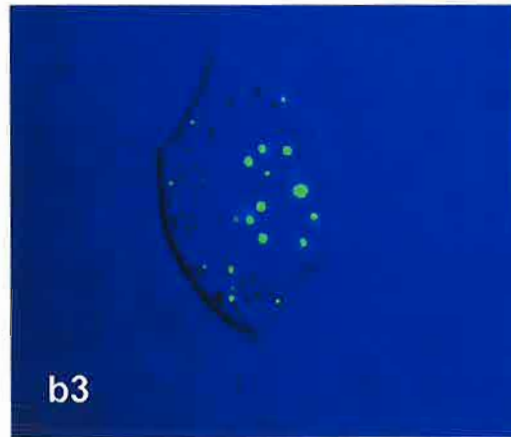
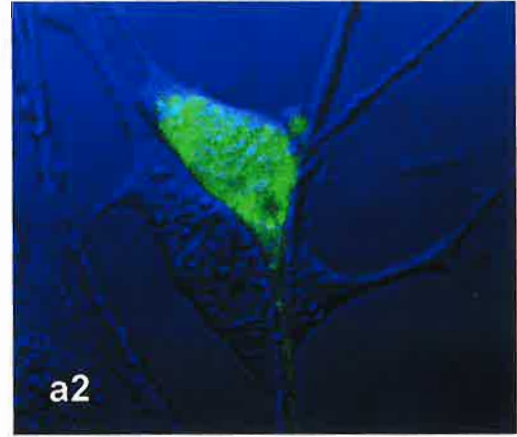
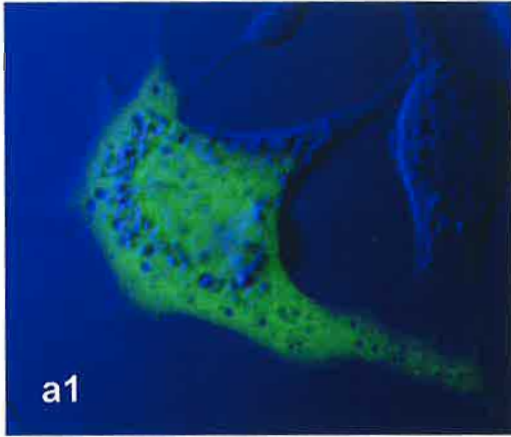
RT-PCR using primers flanking the 51 bp sequence (N981265F + N1384R) showed that the *NUP98* isoform containing the 51 bp insertion was expressed at a lower level than the isoform without the insertion (data not shown). A review of recent publications on *NUP98* translocations revealed that the isoform of *NUP98* containing the 51 bp insertion has been noted by another group (Hatano *et al.*, 1999). In agreement with the findings described in this thesis, the isoform containing the insertion was present at a lower level than the isoform corresponding to the *NUP98* sequence in Genbank (Hatano *et al.*, 1999).

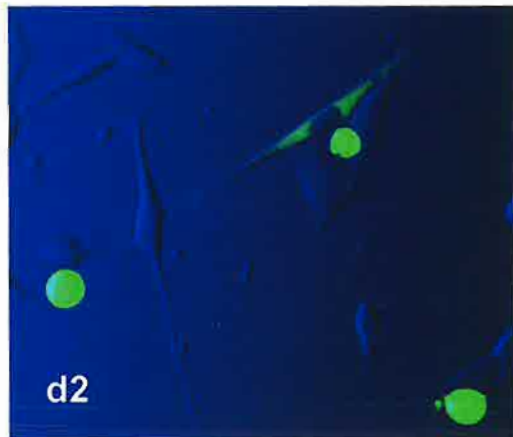
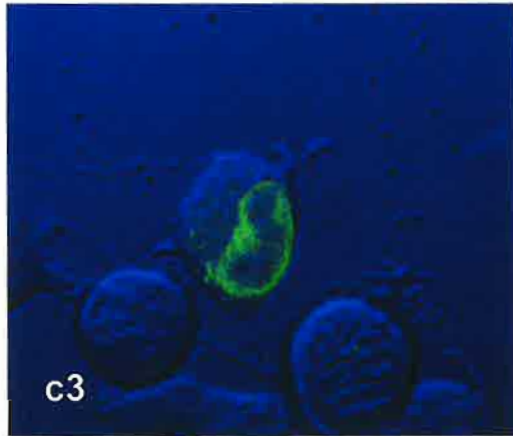
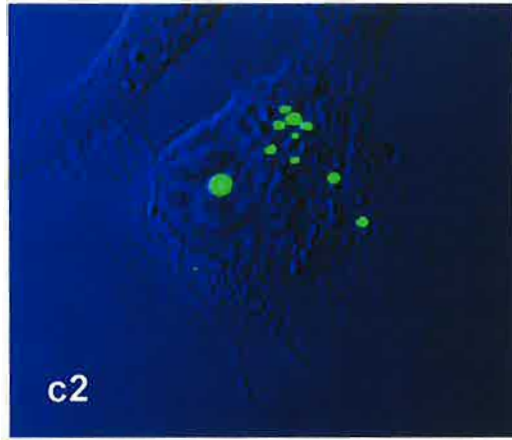
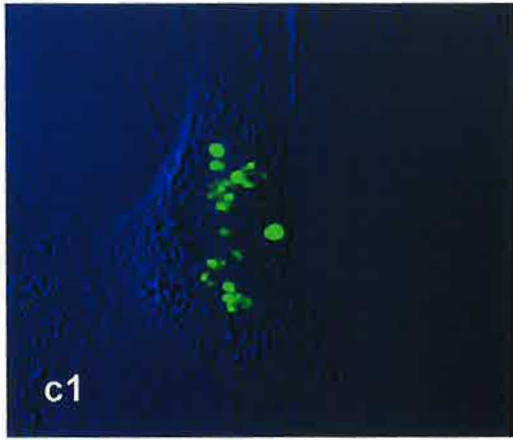
A BLASTN search of the cloned *NUP98* cDNA sequence showed that it had increased homology to the rat *NUP98* cDNA sequence relative to the human *NUP98* cDNA sequence in Genbank. In fact, the 51bp insertion was 100% homologous to the rat sequence and removed a gap in the alignment between the human and rat *NUP98* cDNA sequences in Genbank. The insertion maintains the open reading frame of *NUP98* and introduces an extra FG repeat into the protein, creating a total of 39 FG repeats.

### **Figure 5.5 Subcellular localisation of gfp tagged proteins in NIH3T3 cells**

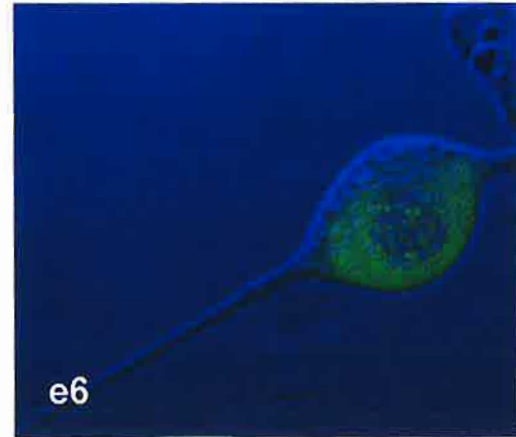
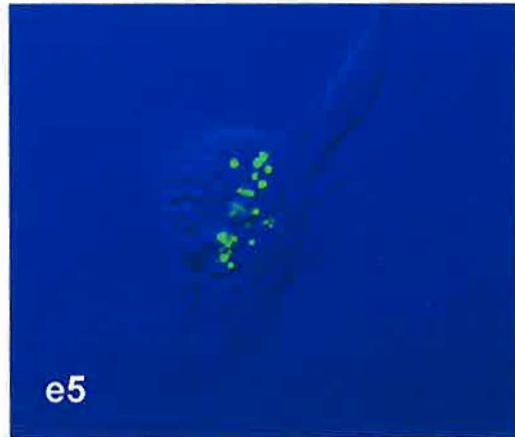
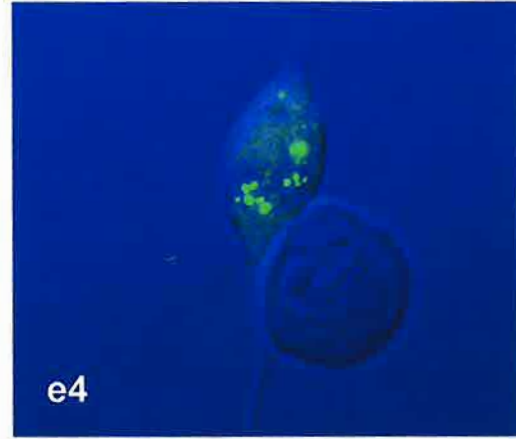
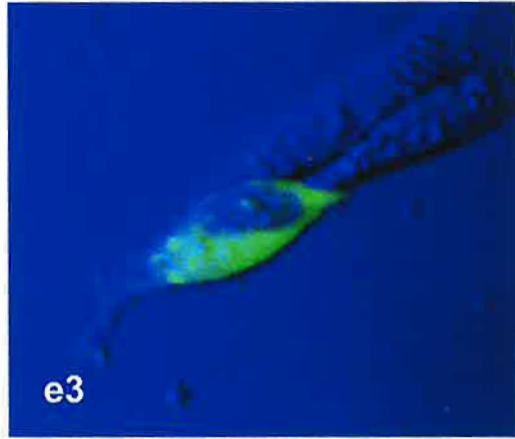
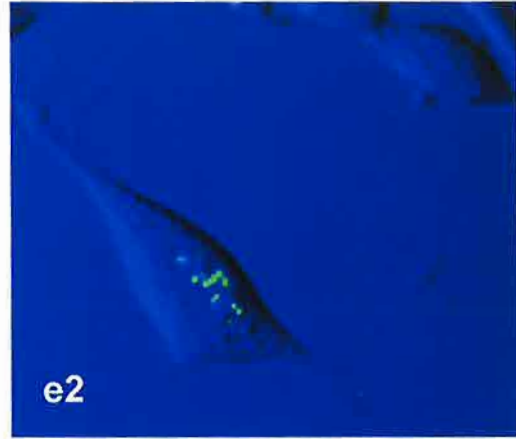
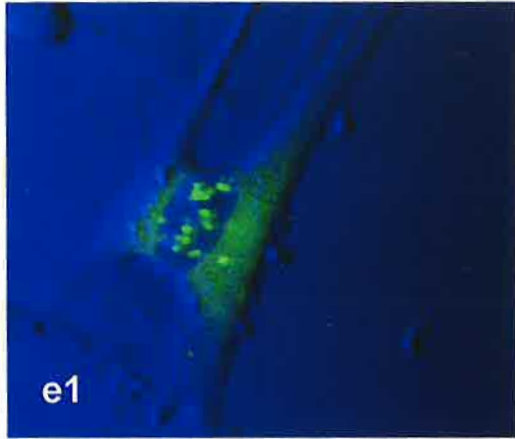
The images in this Figure were made by merging the confocal (green) and light transmission (blue) pictures of the same cell using Confocal Assistant© (freeware computer program created by Todd Clark Brelje, University of Minnesota, Minneapolis, MN). When more than one pattern was seen for a particular protein the representative images of each pattern are shown. All photographs were taken at  $\times 40$  magnification with a digital zoom of  $\times 5$ , except d2 which was taken at low magnification.

- (a) **gfp alone** (1) transient expression, (2) stable expression
- (b) **gfp-nup98** (1-3) transient expression, (4) stable expression
- (c) **gfp-nup98t** (1 and 2) transient expression, (3) stable expression
- (d) **gfp-smgGDS** (1 and 2) transient expression, (3) stable expression
- (e) **gfp-nrg** (1-3) transient expression, (4-6) stable expression









## 5.2 Cellular localisation of gfp tagged nup98, nup98t, smgGDS and nrg

Once the inserts were verified as correct they were excised from pTarget with *Xho* I and *Cfr* 9I and subcloned into the same sites in pEGFP-C2 (Fig. 5.1). This was done to create an in frame fusion of *GFP* to the 5'end of the cDNA of interest. Transfection of the plasmid constructs into mouse NIH 3T3 cells (Sections 2.2.2.2 and 2.2.2.3) resulted in constitutive expression, driven by the CMV promoter in pEGFP-C2, of the desired proteins tagged at their amino-terminus with gfp.

The proteins were tagged at the amino-terminus rather than the carboxy-terminus for two reasons. Firstly, others have shown that tagging nup98 at its amino-terminus with haemagglutinin (ha) does not prevent its correct localisation (Fontoura *et al.*, 1999; Kasper *et al.*, 1999; Pritchard *et al.*, 1999; Zolotukhin and Felber, 1999). Secondly, a yeast protein called Yeb3p, which is structurally similar to smgGDS had been shown to mislocalise when tagged at its carboxy-terminus, but it localised correctly when tagged at its amino-terminus (Pan and Goldfarb, 1998).

The subcellular location of the amino-terminal gfp tagged proteins was assessed by confocal microscopy (Fig. 5.5). Localisation data was obtained from cells transiently expressing the proteins 48 h after transfection. Additionally, cells that were selected for stable expression of the proteins (Section 2.2.2.3) were examined four weeks after transfection. Stably transfected cells generally provide a more accurate representation of protein localisation because they express the protein at a lower level than transiently transfected cells, and are therefore less prone to localisation artefacts caused by overexpression.

Transiently transfected cells were found to express gfp tagged proteins at very high levels 48 h after transfection. The proteins were easily visible at 30% of the laser capacity of the confocal microscope. Therefore, the localisation of transiently expressed gfp tagged proteins was observed at 30% laser capacity to avoid photobleaching the gfp. In contrast, all stably transfected cells observed 4 weeks after transfection expressed the gfp tagged proteins at levels 3-10 times lower than that of transiently transfected cells, regardless of the particular construct they were transfected with. Consequently, it was necessary to use full laser capacity to observe localisation of gfp tagged proteins in stably transfected cells.

The diffuse, non-specific cellular distribution of the gfp alone control, observed with both transient (Fig. 5.5a1) and stable expression (Fig. 5.5a2), is expected because the amino acid sequence of gfp does not contain any subcellular localisation motifs. The non-specific distribution of gfp shows that the localisation patterns described below for gfp tagged nup98, nup98t, smgGDS and nrg are due solely to each of these proteins and are not influenced by the gfp tag.

Transiently expressed gfp-nup98 localised in a punctate discontinuous ring around the nuclear envelope (NE) in about 90% of the cells (Fig. 5.5b1-2). In these cells gfp-nup98 also occurred diffusely throughout the nucleus. Gfp-nup98 was also found in the cytoplasm, but at a significantly lower level than in the nucleus. Occasionally cytoplasmic speckles were also observed. In about 10% of cells transiently expressing gfp-nup98, the protein did not localise to the NE but occurred as large protein aggregates mostly within the nucleus and occasionally within the cytoplasm (Fig. 5.5b3).

In cells which stably expressed gfp-nup98 (Fig. 5.5b4), the protein consistently localised in a punctate discontinuous ring around the NE, and was present diffusely within the nucleus and

cytoplasm. There were no stably expressing cells in which gfp-nup98 was absent from the NE.

In cells which transiently expressed gfp-nup98t, the protein was not present in a discontinuous ring around the nucleus in any of the cells. Instead, gfp-nup98t occurred in nuclear aggregates in all cells examined (Fig. 5.5c1). The occasional cell (~10%) also had cytoplasmic aggregates containing gfp-nup98t (Fig. 5.5c2).

In cells which stably expressed gfp-nup98t (Fig. 5.5c3), the protein localised around the NE in a pattern identical to gfp-nup98. No cells were observed in which stably expressed gfp-nup98t was absent from the NE.

Gfp-smgGDS accumulated in a continuous diffuse pattern throughout the cytoplasm (Fig. 5.5d1-3). The most striking feature of gfp-smgGDS was its exclusion from the nucleus in all cells that displayed typical fibroblast morphology. The absence of a nucleus in bright green round cells (Fig. 5.5d2), and the observation that these cells were floating rather than attached to the slide, are evidence that these cells were in mitosis. The localisation pattern of transiently (Fig. 5.5d1-2) and stably (Fig. 5.5d3) expressed gfp-smgGDS was identical.

Finally, gfp-nrg was located in dots throughout the nucleus and also in a diffuse cytoplasmic pattern (Fig. 5.5e1 and Fig. 5.5e4). The nuclear dots varied in intensity and size, as assessed by 3D modelling of Z sections through the cell (data not shown). About 20% of cells expressing gfp-nrg were biased in varying degrees either towards a nuclear dot pattern (Fig. 5.5e2 and Fig. 5.5e5), or towards a diffuse cytoplasmic pattern (Fig. 5.5e3 and Fig. 5.5e6). In contrast to gfp-nup98 and gfp-nup98t, no cells were seen in which gfp-nrg localised around the NE. Cells which stably expressed gfp-nrg (Fig. 5.5e4-e6) displayed identical localisation

patterns to cells transiently expressing gfp-nrg (Fig. 5.5e1-e3), thereby indicating that the nuclear dot pattern was not simply an artefact of overexpression.

## 5.3 Discussion

This chapter describes experiments involving an *in vitro* study of the intracellular location of nrg. The purpose of these experiments was to gain an understanding of the mechanism(s) involved in the transforming ability of nrg. By determining the subcellular location of proteins, one can then more fruitfully speculate on their roles and what other proteins they may interact with.

In order to generate plasmid constructs encoding gfp tagged versions of nrg and its components, cDNAs were amplified by RT-PCR and then cloned and sequenced. It was essential to sequence the constructs because of the possibility of polymerase induced point mutations. Clones that contained point mutations may have encoded prematurely truncated proteins, or proteins with amino acid substitutions, that were at risk of producing aberrant localisation results. Sequencing of the cloned RT-PCR products revealed a number of naturally occurring variations in the *RAP1GDS1* cDNA sequence. The most interesting of all noted sequence variations was the inclusion of a novel exon in the predominant isoform of *RAP1GDS1* in PBMNC.

### 5.3.1 The predominant isoform of smgGDS in PBMNC contains 12 armadillo repeats

The DNA sequence of the novel 147 bp exon in the predominant isoform of *RAP1GDS1* in PBMNC was analysed using the BLAST X algorithm. This analysis revealed that it encoded a total of 49 amino acids that were 79% identical (95% similar) to the 3rd armadillo repeat of *Xenopus* smgGDS (Fig. 5.6a). Therefore the presence of the additional 147 bp exon results in

a variant of smgGDS containing 12 tandem armadillo repeats instead of the 11 reported in Genbank entry X63465. BLAST X analysis of the entire *RAP1GDS1* coding sequence including the additional 147 bp revealed that the isoform of smgGDS containing 12 armadillo repeats is actually a better match for the putative invertebrate orthologues of smgGDS, in *Xenopus laevis* (XsmgGDS) (Iouzalén *et al.*, 1998), *Dictyostelium discoideum* (darlin) (Vithalani *et al.*, 1998), and *Drosophila melanogaster* (vimar) (Lo and Frasch, 1998) than the isoform containing 11 armadillo repeats (Fig. 5.6b). The improved alignment of the 12 armadillo repeat isoform of human smgGDS with XsmgGDS, darlin and vimar supports the idea that these invertebrate proteins are indeed orthologues of human smgGDS.

It is unclear why the sequence encoding the extra armadillo repeat has not been reported in the previous cloning of vertebrate *RAP1GDS1* from bovine and humans. In both of these cases the cloning was performed from brain tissue and it may be that the 147 bp exon is preferentially spliced out in brain whereas in PBMNC it is preferentially spliced in.

### **5.3.2 Localisation of gfp-nup98**

It has been well established, using non-gfp based tagging, that nup98 normally accumulates as a discontinuous ring around the NE and is also present diffusely within the nucleus and cytoplasm (Fontoura *et al.*, 1999; Kasper *et al.*, 1999; Powers *et al.*, 1995; Pritchard *et al.*, 1999; Radu *et al.*, 1995b; Zolotukhin and Felber, 1999). Therefore, determining the localisation pattern of gfp-nup98 was initially not one of the aims of this study. However, because transiently expressed gfp-nup98t did not localise around the NE, in contrast to published results (Kasper *et al.*, 1999), it was thought that the gfp tag may have caused mislocalisation of gfp-nup98t. In order to assess this possibility, the transient localisation pattern of gfp-nup98 was examined. Cells expressing gfp-nup98 transiently showed the

**Figure 5.6 The additional 147 bp exon in *RAP1GDS1* encodes an armadillo repeat**

(a) Selected BLASTX output using the alternatively spliced 147bp exon from the major isoform of *RAP1GDS1* as the query sequence. The amino acids of the subject (sbjct) sequence represent the 3rd armadillo repeat of XSmgGDS (Iouzalén *et al.*, 1998).

(b) The isoform of *RAP1GDS1* encoding 12 armadillo repeats (1824bp, Genbank AF237413) is more similar to the putative invertebrate homologues of smgGDS than the isoform encoding 11 armadillo repeats (1677 bp, Genbank X63465). The outputs of BLAST X searches are shown in tabular format so that the similarity score and e value can be easily compared. The e value describes the number of hits one can "expect" to see just by chance when searching a database of a particular size, and essentially indicates the random background noise that exists for matches between sequences. An e value of less than 0.05 indicates that the sequences are significantly related.

a)

BLASTP 2.0.12

Reference:

Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25:3389-3402.

Query= smgGDS 147 bp alternatively spliced exon  
(49 letters)

Database: nr  
496,499 sequences; 155,672,339 total letters

|  |                 |            |
|--|-----------------|------------|
| Sequences producing significant alignments:  | Score<br>(bits) | E<br>Value |
| emb CAA06746.1  (AJ005070) Rab-binding protein [ <i>Xenopus laevis</i> ]<br>Length = 607 | 90              | 2e-18      |

Score = 90.5 bits (221), Expect = 2e-18  
Identities = 39/49 (79%), Positives = 47/49 (95%)

Query: 1 EGRSAVDQAGGAQIVIDHLRSLCSITDPANEKLLTVFCGMLMNYSENEND 49  
EGR AVDQ GGAQIV+DHLRS+C++TDP++EKL+TVFCGMLMNYSENEND  
Sbjct: 122 EGRRAVDQEGGAQIVVDHLRSMCTLTDPSSSEKLMTVFCGMLMNYSENEND 170

b)

|                                     | 1824 bp<br><i>RAP1GDS1</i> | 1677 bp<br><i>RAP1GDS1</i> | 1824 bp<br><i>RAP1GDS1</i> | 1677 bp<br><i>RAP1GDS1</i> |
|-------------------------------------|----------------------------|----------------------------|----------------------------|----------------------------|
|                                     | score                      |                            | e value                    |                            |
| Human smgGDS                        | 1012                       | 1042                       | 0.0                        | 0.0                        |
| Bovine smgGDS                       | 971                        | 1001                       | 0.0                        | 0.0                        |
| Cattle smgGDS                       | 701                        | 1001                       | 0.0                        | 0.0                        |
| XsmgGDS<br>( <i>X. laevis</i> )     | 998                        | 888                        | 0.0                        | 0.0                        |
| Vimar<br>( <i>D. melanogaster</i> ) | 222                        | 188                        | 4e-57                      | 8e-47                      |
| Darlin<br>( <i>D. discoideum</i> )  | 61                         | 52                         | 3e-08                      | 2e-05                      |



predicted localisation pattern, thereby excluding the possibility that the gfp tag was causing gfp-nup98t to localise incorrectly.

Gfp-nup98 accumulated predominantly as a discontinuous ring around the NE and was also present diffusely throughout the nucleus and the cytoplasm consistent with previous reports. It has been shown that nup98, but not other GLFG nucleoporins such as nup214 and nup153, can completely translocate from the nucleus to the cytoplasm under certain conditions (Zolotukhin and Felber, 1999). This demonstrates that nup98 is a dynamic rather than simply a stationary component of the NPC.

In some cells transiently expressing gfp-nup98, the protein failed to localise to the NE and occurred as protein aggregates mostly within the nucleus and occasionally within the cytoplasm. In contrast, all cells stably expressing gfp-nup98 showed the expected NE localisation and lacked nuclear aggregates, indicating that failure to localise at the NE was an artefact of overexpression of nup98 in transiently transfected cells.

### **5.3.3 Localisation of gfp-nup98t**

Published results show that an amino-terminus ha tagged truncated form of nup98 localises to the NE in a pattern similar to full length nup98 (Kasper *et al.*, 1999). This is not in agreement with the finding that transiently expressed gfp-nup98t failed to localise to the NE, and typically accumulated as aggregates inside the nucleus. The localisation of gfp-nup98t was not due to the gfp tagging because gfp-nup98 localised correctly. However in cells stably expressing gfp-nup98t the protein consistently localised to the NE, in agreement with the results of Kasper *et al.* (1999). The aberrant localisation of transiently expressed gfp-nup98t was therefore an artefact of overexpression of nup98t. It is possible that cells in which transiently expressed gfp-nup98t was located at the NE were overlooked because the

microscope was not set at the same high power setting used for observation of the stably expressing cells.

Cells which transiently overexpressed gfp-nup98 or gfp-nup98t 48 h after transfection shared a common feature, ie the absence of NE localisation and an associated nuclear (and sometimes cytoplasmic) aggregation of gfp-nup98 or gfp-nup98t. In stably transfected cells observed four weeks after transfection, gfp-nup98 and gfp-nup98t were always located at the NE. This difference between the transient and stable localisation patterns of gfp-nup98 and gfp-nup98t suggests that the failure of these proteins to localise at the NE is associated with a lethal phenotype. It is possible that a lethal effect of overexpressed gfp-nup98 and gfp-nup98t is caused by a disruption of nucleocytoplasmic transport. In support of this hypothesis, overexpression of the isolated FG repeat region of nup98, or the nup98 related nucleoporin, nup214, is known to disrupt nucleocytoplasmic transport by interfering with a number of soluble transport factors (Boer *et al.*, 1998; Pritchard *et al.*, 1999; Zolotukhin and Felber, 1999).

#### **5.3.3.1 Possible disruption of crm1 mediated nucleocytoplasmic transport**

Nup214, like nup98, is a member of the FG repeat family of nucleoporins (Kraemer *et al.*, 1994). Boer *et al.* (1998) induced transient overexpression of nup214 in the human monoblast cell line U937. The localisation pattern of overexpressed nup214 was observed over 3 days and was found to alter dramatically over this period. On the first day of overexpression, nup214 localised mainly to the NE and cytoplasmic speckles, which was consistent with the localisation pattern observed for endogenous nup214. During the second day, a number of cells were observed in which overexpressed nup214 was absent from the NE and was located within the nucleus. On the third day, overexpressed nup214 was absent from the NE, and localised diffusely within the nucleus and in bright nuclear aggregates in 90% of the cells.

The timing of the change in localisation of overexpressed nup214 coincided with the a decrease in export of polyA RNA from the nucleus and the induction of growth arrest and apoptosis (Boer *et al.*, 1998). Transient overexpression of the isolated FG repeat region of nup214 was found to have exactly the same effect on polyA RNA export and cell viability (Boer *et al.*, 1998).

The FG repeat region of nup214 is known to interact with the nuclear-cytoplasmic shuttling RNA and protein transport factor, crm1 (Fomerod *et al.*, 1997). Accordingly, Boer *et al.* (1998) assessed whether the lethal effect of overexpressed nup214 resulted from interference with the transport function of crm1. After 3 days of nup214 overexpression it was found that crm1 colocalised with nup214 inside the nucleus and was depleted from the NE. The timing of the disruption of crm1 localisation coincided with nuclear accumulation of polyA RNA and induction of growth arrest and apoptosis. It was also shown that overexpressed nup214 bound to and interfered with the function of another transport factor, p97/importin  $\beta$ , by depleting it from the NE and sequestering it inside the nucleus (Boer *et al.*, 1998).

Nup98 directly interacts via its FG repeat region with crm1 (Zolotukhin and Felber, 1999) and importin  $\beta$  (Radu *et al.*, 1995a). Given that overexpressed nup214 interferes with the normal function of crm1 and importin  $\beta$ , by depeleting these transport factors from the NE and sequestering them in the nucleus, it seems likely that overexpression of nup98 and nup98t would disrupt nucleocytoplasmic transport in a similar manner. In fact, it has been shown that overexpression of the isolated FG repeat region of nup98 interferes with crm1 mediated nuclear export of HIV Rev by sequestering crm1 inside the nucleus and preventing its interaction with the NE (Zolotukhin and Felber, 1999). Disruption of the normal function of crm1, and perhaps importin  $\beta$ , caused by overexpressed nup98 would presumably coincide with absence of nup98 from the NE and its accumulation as aggregates inside the nucleus, as

is seen with overexpressed nup214. This is consistent with the observation that overexpressed gfp-nup98t, and to a lesser extent gfp-nup98, failed to localise to the NE and accumulated as aggregates inside the nucleus of cells observed 48 h after transfection.

Overexpression of nup98 and nup98t may also result in the induction of growth arrest and apoptosis, as is caused by overexpression of nup214 or its isolated FG repeat region. Perhaps cells, in which overexpressed gfp-nup98 or gfp-nup98t was absent from the NE, died days after transfection because of defects in nucleocytoplasmic transport. This would explain why gfp-nup98 and gfp-nup98t were found to localise at the NE in all cells which stably expressed these proteins four weeks after transfection.

#### **5.3.3.2 Possible disruption of rae1 mediated nucleocytoplasmic transport**

Embedded within the FG repeat region of nup98 is a region of highly charged amino acids referred to as the GLEBS-like domain, or the rae1 binding domain (Pritchard *et al.*, 1999). This domain directly interacts with rae1, which is an mRNA export factor that shuttles between the nucleus and the cytoplasm (Pritchard *et al.*, 1999). Over-expression of the GLEBS motif region of nup98 diminishes the transient association of rae1 with the NE (Pritchard *et al.*, 1999). This inhibits rae1 mediated RNA export and results in the accumulation of large aggregates of polyA RNA in the nucleus (Pritchard *et al.*, 1999). It is therefore likely that transient over-expression of gfp-nup98, or gfp-nup98t, which both contain the rae1 binding domain, acted to disrupt rae1 mediated nuclear export of polyA RNA by sequestering rae1 in the nucleus

Overexpressed gfp-nup98t failed to localise to the NE and accumulated in nuclear (and to a lesser extent cytoplasmic) aggregates in all cells observed 48 h after transfection. In contrast, overexpressed gfp-nup98 was absent from the NE in only 10% of cells observed 48 h after

transfection. This may suggest that nup98t is able to block the association of rae1 with the NE more effectively than full length nup98. In support of this hypothesis, Pritchard *et al.* (1999) found that overexpression of the GLEBS motif of nup98 blocked the association of rae1 with the NE in all cells observed 16-18 h after transfection. In contrast, rae1 was present at the NE in all cells that overexpressed full length nup98 observed 16-18 h after transfection (Pritchard *et al.*, 1999).

The FG repeat region of nup98 also interacts with another nuclear-cytoplasmic shuttling mRNA transport factor called tap (Bachi *et al.*, 2000). Although the effect of over-expression of truncated nup98 on the RNA export function of tap has not been assessed, it is possible that this would also block nuclear export of RNA by inhibiting association of tap with the NE.

In summary, overexpression of nup98 or nup98t has the potential to disrupt nucleocytoplasmic transport by interfering with a number of soluble transport factors. Perhaps the nuclear aggregates observed in all cells that overexpressed gfp-nup98t, and some cells that overexpressed gfp-nup98, are complexes that contain gfp-nup98t or gfp-nup98 bound to rae1 or any other nucleocytoplasmic transport factors that interact with the FG repeats of nup98.

### **5.3.4 Localisation of gfp-smgGDS**

The cellular localisation pattern of smgGDS has not previously been reported. Gfp-smgGDS localised diffusely throughout the cytoplasm and was excluded from the nucleus. This pattern is in accordance with its reported role as a guanine nucleotide disassociation stimulator for cytoplasmic GTPases (Mizuno *et al.*, 1991). Both isoforms of smgGDS identified in this study were analysed using the PSORTII World Wide Web Server (<http://psort.nibb.ac.jp/form2.html>) which predicts protein sorting and localisation signals in amino acid sequences. The inclusion of the extra armadillo repeat was not predicted to alter

the localisation of smgGDS, since proteins with and without the extra armadillo repeat were predicted to be most likely cytoplasmic. No nuclear localisation sequences were found by PSORTII and this is in accordance with the absence of gfp-smgGDS from the nucleus.

### **5.3.5 Localisation of gfp-nrg**

Nrg consists of the amino-terminal end of nup98 fused to the entire smgGDS protein. Gfp-nrg localised to distinct subnuclear domains in the form of nuclear dots, and was also present diffusely throughout the cytoplasm. Unlike stably expressed gfp-nup98 and gfp-nup98t, stably expressed gfp-nrg was located in nuclear dots in 90% of cells observed four weeks after transfection. This result indicates that nrg does not act to disrupt nucleocytoplasmic transport, otherwise cells in which gfp-nrg was located inside the nucleus would not have been viable.

The most striking feature of the nuclear localisation of nrg is that it brings the armadillo repeats of smgGDS into the nucleus. This is in complete contrast to the normal situation in which the armadillo repeats of smgGDS are excluded from the nucleus.

The nuclear dot pattern of gfp-nrg is characteristic of a number of proteins which localise to specific subnuclear domains and have known roles in transcription eg amll1 (McNeil *et al.*, 1999) and mll (Yano *et al.*, 1997). It has been shown by two independent groups that the FG repeats of nup98 are very potent transactivators of gene transcription (Kasper *et al.*, 1999; Nakamura *et al.*, 1999). The FG repeat region of nup98 directly interacts with the transcriptional co-activators cbp and p300 (Kasper *et al.*, 1999). The appearance of gfp-nrg as nuclear dots is therefore in accordance with the known transcriptional transactivation property of the nup98 portion of nrg.

In some cells expressing *gfp-nrg*, the protein showed a bias toward nuclear localisation, while other cells were biased toward a cytoplasmic localisation similar to that of *gfp-smgGDS*. This observation suggests that *nrg* may shuttle between the nucleus and the cytoplasm, perhaps in cell cycle dependent fashion. *Nrg* may exert different transforming properties in the nucleus and the cytoplasm, each contributing to leukemogenesis.

#### 5.3.5.1 *Nrg* in the nucleus

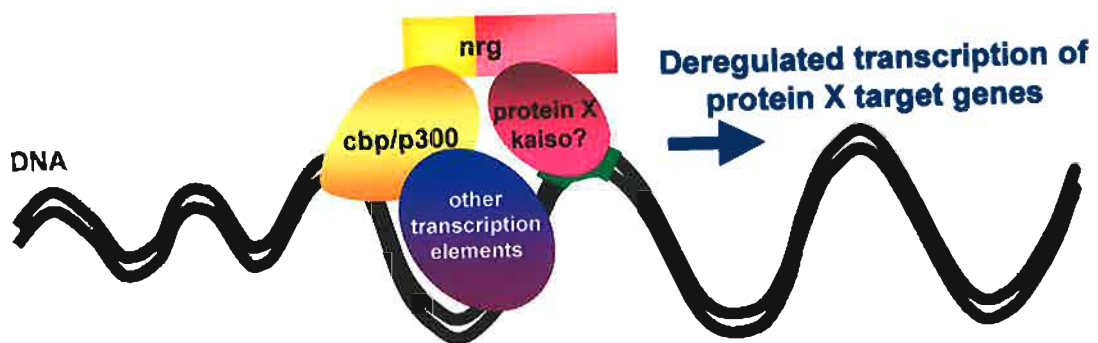
For *gfp-nrg* to accumulate as bright nuclear dots this would mean that each dot contains multiple copies of the protein. Since the *nup98* portion of *nrg* has known transcriptional transactivation potential (Kasper *et al.*, 1999; Nakamura *et al.*, 1999), it is feasible to propose a model in which *nrg* functions as a novel bipartite transcriptional regulator (Fig. 5.7). In this model, the FG repeats of *nrg* would provide general promoter transactivation by interacting with co-activators *cbp* and *p300*. *Cbp/p300* possesses histone acetyltransferase activity, which acts to unwind the DNA and allows access to other elements of the transcriptional machinery (reviewed in Blobel, 2000). The *smgGDS* portion of *nrg* contains twelve tandem armadillo repeats which are generally thought to act as a protein/protein interaction motif (Peifer *et al.*, 1994). Therefore it is conceivable that the armadillo repeats in *nrg* (not present in the nucleus of a normal cell) may interact with a sequence specific DNA binding protein, arbitrarily named “protein X”, and recruit this into the complex. In this way, the pattern of transcription of genes containing a “protein X” DNA interaction motif in their promoters would be altered to bring about cellular transformation. For example “protein X” may activate a transcription pathway(s) that would normally be absent in precursor T-lymphocytes.

The classic example of an armadillo protein interacting with a DNA binding protein in cancer is the interaction between  $\beta$ -catenin and lymphoid enhancing factor 1/T-cell transcription factor (*lef-1/tcf*) (reviewed in Waltzer and Bienz, 1999).  $\beta$ -catenin is an armadillo repeat

**Figure 5.7 A model for the transforming mechanism of nrg in the nucleus**

Diagrammatic representation of the model proposed in Chapter 5, for a detailed explanation see text. The DNA segment containing the “protein X” interaction motif is shown in green.





protein which is normally destabilised in the cytoplasm by binding a multiprotein complex which includes the adenomatous polyposis coli (APC) tumour suppressor protein. In the wingless signalling pathway, which is involved in cell fate determination, the activity of the destabilising complex is inhibited and  $\beta$ -catenin accumulates in the cytoplasm and translocates to the nucleus. Once inside the nucleus,  $\beta$ -catenin forms a transcriptional activator complex with lef-1/tcf which activates specific lef-1/tcf targets. Mutations of the *APC* or  *$\beta$ -catenin* genes, which occur in heritable colon cancer, prevent the interaction of  $\beta$ -catenin with APC and allow  $\beta$ -catenin to translocate to the nucleus in the absence of wingless signalling. This in turn results in constitutive activation of lef-1/tcf target genes and results in tumorigenesis (Aoki *et al.*, 1999).

Another recently characterised example of armadillo-DNA binding protein interaction is that between the Catenin p120<sup>cm</sup> and a protein called kaiso (Daniel and Reynolds, 1999). Kaiso is a member of the BTB/POZ (Broad complex, Tramtrak, Bric a brac/Pox virus and zinc finger) family of zinc finger transcription factors. All BTB/POZ proteins identified to date are involved in development and/or cancer, therefore it is likely that kaiso has an as yet undetermined role in oncogenesis. For example the bcl-6 and plzf proteins, which are BTB/POZ members, are causally involved in non-Hodgkins lymphoma (Dallery *et al.*, 1995; Kerckaert *et al.*, 1993) and acute promyelocytic leukaemia (Chen *et al.*, 1993) respectively. Bcl-6 and plzf both form macromolecular complexes with nuclear transcriptional machinery regulators (David *et al.*, 1998; Dhordain *et al.*, 1997). In accordance with a role in transcriptional regulation, kaiso localises to the nucleus and is associated with nuclear dots (Daniel and Reynolds, 1999). The intracellular location of p120<sup>cm</sup>-kaiso interaction has not been determined but is probably nuclear given the nuclear dot distribution of kaiso. In support of this, it was recently found that p120<sup>cm</sup> can translocate to the nucleus under certain conditions, and that nuclear translocation of p120<sup>cm</sup> is associated with increased lef-1/tcf

signalling (Eger *et al.*, 2000). Thus the p120<sup>cm</sup>-kaiso interaction is an example of an armadillo repeat protein interacting with a sequence specific DNA binding transcription factor with a potential role in leukemogenesis.

Armadillo repeats 1 to 7 of p120<sup>cm</sup> are necessary and sufficient for p120<sup>cm</sup>-kaiso interaction, however for technical reasons it has not been possible to determine which of these repeats are essential (Daniel and Reynolds, 1999). Using BLAST P analysis of all 10 armadillo repeats from p120<sup>cm</sup>, Peifer *et al.* (1994) found that p120<sup>cm</sup> is significantly related to smgGDS (e value = 0.039, for an explanation of e value see the legend for Fig. 5.6). The observation that p120<sup>cm</sup> and smgGDS are significantly related raises the intriguing possibility that the armadillo repeat region of nrg may interact with kaiso, or a related BTB/POZ protein, in the nucleus. To strengthen this possibility, it was important to determine if the region of similarity between p120<sup>cm</sup> and smgGDS corresponded to the p120<sup>cm</sup>-kaiso interaction domain, ie armadillo repeats 1 to 7 of p120<sup>cm</sup>. Armadillo repeats 1 to 7 from p120<sup>cm</sup> (Genbank Z17804) were analysed using BLAST P. The results confirmed those of Peifer *et al.* (1994), and showed that p120<sup>cm</sup> is significantly similar to the previously reported isoform of smgGDS. The region of similarity includes the second half of p120<sup>cm</sup> armadillo repeat 1, all of p120<sup>cm</sup> armadillo repeats 2 and 3, and the first half of p120<sup>cm</sup> armadillo repeat 4 (e value = 0.037). More importantly, p120<sup>cm</sup> is more similar to the isoform of smgGDS with 12 armadillo repeats than it is to the previously reported isoform of smgGDS. The inclusion of the additional armadillo repeat in smgGDS raises the e value for the match between p120<sup>cm</sup> and smgGDS from 0.036 to 0.016. Accordingly, the region of similarity between p120<sup>cm</sup> and the 12 armadillo repeat isoform of smgGDS extends from the second half of p120<sup>cm</sup> armadillo repeat 1 to the end of p120<sup>cm</sup> armadillo repeat 5.

The significant similarity between p120<sup>cm</sup> and smgGDS, over almost the entire region of p120<sup>cm</sup> that is necessary and sufficient for p120<sup>cm</sup>-kaiso interaction, supports the hypothesis that the armadillo repeats of nrg could interact with kaiso in the nucleus. Furthermore, the increased similarity between the kaiso interaction domain of p120<sup>cm</sup> and the smgGDS isoform with 12 armadillo repeats suggests that the additional armadillo repeat in nrg may strengthen the interaction between nrg and kaiso. Perhaps kaiso, or a related BTB/POZ protein, is the proposed “protein X” that interacts with nrg inside the nucleus to promote leukemogenesis.

### 5.3.5.2 The role of SMAP

SMAP is a protein with 9 armadillo repeats which was identified in a screen for proteins that interact with smgGDS (Shimizu *et al.*, 1996). SMAP is ubiquitously expressed in all tissues examined and is present in the nucleus and highly concentrated at the endoplasmic reticulum area. The interaction of SMAP and smgGDS does not interfere with the normal regulatory effects that smgGDS has on rhoA (Shimizu *et al.*, 1996). SMAP is tyrosine phosphorylated by v-Src and this reduces its affinity for smgGDS (Shimizu *et al.*, 1996). SMAP is the human orthologue of mouse KAP3, which is part of the KIF3A/3B protein motor complex involved in ATP dependent transport of organelles along microtubules (Yamazaki *et al.*, 1996). Although the role of the smgGDS/SMAP interaction has not been determined, it has been speculated that SMAP acts as an adaptor for both smgGDS and the KIF3A/3B ATPase motor protein and links them with the smgGDS regulated small G protein and Src tyrosine kinase signalling pathways.

In the nucleus SMAP is known to associate with HCAP (Human chromosome-associated polypeptide), which is a nuclear protein with a role in regulating the assembly and maintaining the structure of mitotic chromosomes (Shimizu *et al.*, 1998). The SMAP/HCAP complex is associated via SMAP with the KIF3A/3B ATPase dependent organelle transport

motor (Shimizu *et al.*, 1998). These proteins, and possibly others proteins, form a large complex which is thought to be involved in chromosome movement during mitosis. However the exact role of this nuclear protein complex has not yet been determined.

There are no reports in which the potential association between smgGDS and the SMAP/HCAP/KIF3A/3B complex has been tested. The experiments described in this thesis show that smgGDS is absent from the nucleus, nevertheless interaction between smgGDS and the SMAP/HCAP/motor protein complex could occur during mitosis when there is no nuclear membrane. However, in cells expressing nrg, the smgGDS portion of nrg is present inside the nucleus and may therefore interact with the SMAP/HCAP/ KIF3A/3B complex for the greater part of the cell cycle. It is difficult to speculate how this interaction could result in leukemogenesis because the role of the complex has not been elucidated.

### **5.3.5.3 Nrg in the cytoplasm**

Gfp-nrg is present in the cytoplasm in a pattern very similar to that of gfp-smgGDS, suggesting that nrg may perform the same role in the cytoplasm as smgGDS. It is possible that the (4;11) translocation effectively increases the amount of cytoplasmic smgGDS because *RAP1GDS1* is expressed under the control of the *NUP98* promoter. This may exert an oncogenic effect by disrupting the balance of GDP/GTP exchange of ras, rho, and rap1 family proteins and favouring their active GTP bound form. This proposed scenario for the oncogenic mechanism of nrg is analogous with the oncogenic property of constitutively active mutant ras (reviewed in de Vries *et al.*, 1996). However it is not yet possible to confirm this scenario because the relative levels of expression of nup98 and smgGDS have not been formally assessed. Even if the amount of smgGDS in the cytoplasm is not effectively increased by nrg, it is possible that smgGDS may have a more stable protein configuration in

the context of *nrg*. In a similar fashion to over-expression of *smgGDS*, increased stability of *smgGDS* would disrupt the balance of GDP/GTP exchange and promote oncogenesis

Recent experiments on *smgGDS* knockout mice suggest that *smgGDS* is involved in anti-apoptotic cell survival signalling through *Ki-ras* (Takakura *et al.*, 2000). *SmgGDS* knockout mice have smaller thymi than normal and their thymocytes show an increased rate of apoptosis. When *smgGDS* is exogenously expressed in *smgGDS* *-/-* primary thymocytes, by transfection of a *smgGDS* encoding plasmid, the cells are rescued from apoptosis (Takakura *et al.*, 2000). If *nrg* effectively results in increased cytoplasmic activity of *smgGDS* this may provide precursor T-cells with increased resistance to apoptosis and allow them to accumulate disproportionately over other haematopoietic cell types. Increased resistance to apoptosis would also allow the development of extra genetic changes that could lead to leukaemia. This possible scenario is analogous to the anti-apoptotic phenotype of B cells caused by increased levels of *bcl-2* in *t(14;18)* follicular lymphoma (Hua *et al.*, 1988; reviewed in Sanchez-Garcia, 1997).

## 5.4 Conclusion

The localisation data obtained from this study, in combination with the known properties of the *nup98* and *smgGDS* portions of *nrg*, are sufficient to speculate on models for the transforming properties of *nrg*. The favoured model predicts that *nrg* has dual transforming properties in the nucleus and the cytoplasm. In the nucleus *nrg* may act to deregulate transcription pathways critical to normal T-cell development, while in the cytoplasm *nrg* may act to increase cell survival by promoting an anti-apoptotic phenotype.

## **Chapter 6**

### **Conclusions and future studies**

## Chapter 6

### Conclusions and future studies

The aim of this study was to determine the molecular basis of a (4;11)(q21;p15) translocation in a 21 year old male with T-cell acute lymphocytic leukaemia. Using a combination of Southern analysis and PCR on material from the patient it was shown that the (4;11)(q21;p15) translocation fuses the 5' portion of the chromosome 11 *NUP98* gene to the open reading frame of the chromosome 4 *RAP1GDS1* gene. Fusion of the *NUP98* and *RAP1GDS1* genes produces a hybrid message named *NRG*.

During the course of the study, another 2 ALL patients were identified with a cytogenetically similar translocation to that of the initial patient. *NRG* was detectable by RT-PCR in the leukaemic cells of all three ALL patients with the (4;11)(q21;p15) translocation. The reciprocal transcript, *RGN* was only detected in one of the three patients, indicating that *NRG* is the more important fusion transcript. Further support for this hypothesis came from Northern analysis of the initial patient which showed that *NRG* was expressed at presentation and relapse of ALL but not during remission.

Cellular localisation experiments using gfp tagged proteins showed that gfp-nrg is present in a nuclear dot pattern similar to that of a number of transcription factors. Gfp-nrg also showed a diffuse cytoplasmic distribution similar to that of gfp-smgGDS. In contrast to gfp-nrg, gfp-smgGDS was excluded from the nucleus. Therefore, nrg results in aberrant nuclear localisation of the smgGDS protein by fusing it to the FG repeat portion of nup98. A model for the leukemogenic action of nrg has been proposed whereby nrg acts to alter normal transcriptional pathways in a precursor T-cell by acting as a bipartite transcriptional regulator. In this model transcriptional transactivation activity is supplied by the nup98 portion and



indirect DNA binding activity is supplied by the smgGDS portion through recruiting of an as yet undetermined DNA binding protein called “protein X”.

## 6.1 A common theme for nup98 fusions

The obvious common theme for all nup98 fusion proteins is the retention of amino-terminal FG repeat region of nup98. This common feature suggests that truncation and fusion of nup98 is somehow leukemogenic in itself. This is true for the mll protein, which is similar to nup98 in that its amino-terminal region is fused to a variety of different proteins by translocations associated with acute leukaemias (reviewed in Rowley, 1999). Experiments involving embryonic stem (ES) cell knock-in to disrupt the *MLL* gene have shown that truncation of mll does not cause leukaemia (Corral *et al.*, 1996; Dobson *et al.*, 2000). However, ES cell knock-in to create a fusion of the amino-terminus of mll to the lacZ protein, produces mice which develop embryonic stem cell derived acute leukaemias (Dobson *et al.*, 2000). As the lacZ protein has no demonstrated role in tumour formation, this result indicates that truncation and fusion of mll is all that is required for leukemogenesis.

The truncated FG repeat region of nup98, as well as the nup98-hoxa9 and nup98-pmx1 fusion proteins, have a transcriptional transactivation property (Kasper *et al.*, 1999; Nakamura *et al.*, 1999), which is due to direct interaction of the nup98 FG repeat region with the transcriptional coactivators cbp and p300 (Kasper *et al.*, 1999). In contrast, the full length nup98 protein does not possess a transcriptional transactivation property and does not detectably interact with cbp/p300 (Kasper *et al.*, 1999).

The truncated FG repeat region of nup98 alone does not possess transforming properties (Kasper *et al.*, 1999). However the nup98-hoxa9 fusion protein transforms NIH3T3 cells to an anchorage independent phenotype, and interaction with cbp/p300 is essential for this

transforming property (Kasper *et al.*, 1999). In accordance with its lack of interaction with cbp/p300, full length nup98 does not share the transforming property of nup98-hoxa9 (Kasper *et al.*, 1999). These observations indicate that inappropriate interaction of the amino-terminus of nup98 with cbp and p300 is the unifying theme that underlies the leukemogenic ability of nup98 fusion proteins. This leads to the question; how is disruption of cbp/p300 function able to cause leukaemia?

### **6.1.1 Cbp and p300 are key regulators of haematopoietic differentiation**

*CBP* and *p300* encode highly related, and in some cases interchangeable, proteins that function as transcriptional coactivators by interacting with a number of key haematopoietic transcription factors (reviewed in Blobel, 2000). The nuclear pool of cbp/p300 is limited (Horvai *et al.*, 1997; Kamei *et al.*, 1996), which means that unscheduled interaction with cbp/p300 could deplete its availability in the nucleus. This could have a dominant negative effect on the transcriptional pathways required for normal haematopoiesis.

#### **6.1.1.1 Cbp/p300 interacts with aml1**

Cbp and p300 directly interact with aml1 (Kitabayashi *et al.*, 1998), which is a sequence specific DNA binding transcription factor for a number of genes required for normal haematopoietic development (reviewed in Lutterbach and Hiebert, 2000). The amino-terminus of aml1 is fused to a variety of different proteins by translocations associated with acute leukaemia, one of the most common being t(8;21) which fuses aml1 to eto (reviewed in Rowley, 1999). Overexpression of the aml1-eto fusion protein in the myeloid cell line L-G interferes with G-CSF-induced differentiation. However, forced expression of wildtype aml1 is able to overcome this block and induce differentiation (Kitabayashi *et al.*, 1998), indicating that the aml1-eto protein has a dominant negative effect on normal aml-1 mediated transcription. The ability of wildtype aml1 to induce differentiation is significantly enhanced

by overexpression of p300 and interaction of p300 with *aml1* (Kitabayashi *et al.*, 1998). This suggests that the *aml1-eto* protein may exert its dominant negative effect by interfering with the limited nuclear pool of p300.

#### **6.1.1.2 Haploinsufficiency of *cbp* causes haematopoietic defects**

Direct evidence for a dominant negative effect on haematopoietic differentiation caused by depletion of *cbp/p300* comes from observations of mice with monoallelic inactivation of *CBP*. Mice derived from ES cells with an inactivated *CBP* allele display multiple haematopoietic defects, including abnormalities in B and T cell differentiation (Kung *et al.*, 2000). Later in life these mice also develop myeloid and lymphoid malignancies (Kung *et al.*, 2000). These results indicate that a full complement of *cbp* is required for normal haematopoiesis.

The above two examples support the idea that aberrant interaction of the *nup98* FG repeat region with *cbp/p300* could deplete the limited nuclear pool of *cbp/p300*, thereby interfering with transcriptional pathways required for normal haematopoietic differentiation. They also support the idea that truncation and fusion of the *nup98* FG repeat region is leukemogenic, regardless of the fusion partner.

## **6.2 What is the role of the fusion partners of *nup98*?**

If truncation and fusion of the *nup98* FG repeat region is sufficient for deregulation of haematopoiesis, this leads to the question of how the fusion partners of *nup98* contribute to leukemogenesis. It is possible that the fusion partners of *nup98* may determine the specificity of the lineage in which the leukaemia arises, and they may also possess features which contribute to leukemogenesis.

The normal functions of all known nup98 fusion partners have been discussed in chapter 4. Of these 6 partners only the homeobox genes, *HOXA9*, *HOXD13* and *PMX1*, encode proteins with a known ability to bind DNA in a sequence specific manner. Several studies have demonstrated that homeobox genes are targets for disruption in leukaemia. For example, *HoxA9* knockout mice have multiple haematopoietic defects, including reduced numbers of peripheral blood granulocytes and lymphocytes, as well as myeloid and pre-B-cell progenitors (Lawrence *et al.*, 1997). Further to this, the homeobox genes, *HoxA9*, *HoxA7* and *Meis1* have all been identified as frequent targets of proviral insertion in murine myeloid leukaemia (Li *et al.*, 1999; Nakamura *et al.*, 1996b). The proviral insertion leads to inappropriate activation of these genes in haematopoietic precursors (Nakamura *et al.*, 1996b). It has been suggested that cooperative activation of *Hoxa* family and *Meis1* genes may block myeloid differentiation through altered transcription of downstream target genes (Nakamura *et al.*, 1996b). Perhaps fusion of homeobox genes to *NUP98* leads to their constitutive expression and acts to block myeloid differentiation in a similar fashion.

*Ddx10* and *top1* have not been reported to be specifically associated with myeloid cell differentiation and it is therefore difficult to speculate on how their fusion with *nup98* may block myelopoiesis. By virtue of their respective functions, both proteins may be expected to bring the *nup98* moiety of the fusion protein into close proximity with DNA. Kasper *et al.* (1999) pointed out that *in vitro* interaction of *cbp* with RNA polymerase II requires RNA helicase A, which contains a domain highly related to the DEAD box of *ddx10* (Nakajima *et al.*, 1997). They suggested that the DEAD motif in *ddx10* may also interact with *cbp* and therefore that the *nup98-ddx10* protein may act by deregulating *cbp/p300* function. In support of this hypothesis, it was recently shown that another DEAD box RNA helicase, called *p68*, acts as a transcriptional co-activator through its interaction with *cbp* (Endoh *et al.*, 1999). It is possible that *nup98-top1* may act to deregulate *cbp/p300* function in a similar fashion.

### 6.2.1 Why is nrg specific to T-cell ALL?

As shown by the cellular localisation experiments described in chapter 5 of this thesis, smgGDS is normally absent from the nucleus. Fusion of smgGDS to nup98 brings the armadillo repeats of smgGDS into the nucleus, and nrg localises to distinct subnuclear domains that are typical of proteins involved in transcriptional regulation. The potential for the armadillo repeats of nrg to interact with other proteins involved in transcription, such as lef-1/tcf and kaiso, has been discussed in chapter 5.

Since nrg is the only nup98 fusion protein associated with lymphoid leukaemia, specifically T-cell ALL, this suggests that the smgGDS portion of nrg is involved in selection of the T-cell ALL phenotype. It seems unlikely that the (4;11) translocation arose by chance in the same precursor T cell type in the three patients reported here, and in the 3 cases reported in the literature (Bloomfield *et al.*, 1986; Inoue *et al.*, 1985; Pui *et al.*, 1991). Perhaps nrg is only able to exert its leukemogenic effects in precursor T cells because they possess a property that supports the leukemogenic ability of nrg. This peculiar property may be the expression of “protein X”, the hypothesised sequence specific DNA binding protein recruited by the armadillo repeats in nrg (discussed in chapter 5). This hypothesis could be tested experimentally, as described in the next section.

The involvement of smgGDS in thymocyte resistance to apoptosis (Takakura *et al.*, 2000) may suggest that nrg exerts a specific anti-apoptotic activity in precursor T-cells (see Section 5.3.5.3). It is not clear from the experiments described by Takakura *et al.* (2000) if the anti-apoptotic activity of smgGDS is restricted to thymocytes, or if it has this activity in other precursor haematopoietic cells. If the anti-apoptotic activity of smgGDS is restricted to thymocytes this would explain the specific association between nrg and T-cell ALL.

## 6.3 Future studies

The knowledge gained and the material derived as a result of the study described in this thesis can be used to further the understanding of the leukemogenic mechanism(s) of *nrg*.

### 6.3.1 Identifying proteins that interact with *nrg* and pathways affected by *nrg*

*NRG* can be subcloned from the *gfp* expression vector and used as bait in a yeast two-hybrid system to capture proteins that interact with *nrg*. This system is widely used for identifying specific protein-protein interactions. In fact, the *smgGDS*-*SMAP* (Shimizu *et al.*, 1996) and *p120*-*kaiso* interactions (Daniel and Reynolds, 1999), described in chapter 5, were discovered using the yeast two-hybrid system. A yeast two-hybrid assay could reveal the identity of the putative DNA binding “protein X”, and may show that “protein X” is *kaiso* or a related BTB/POZ protein. It may be necessary to perform this assay on a cDNA library derived from *t(4;11)* blast cells in case “protein X” is expressed principally in a restricted population of cells. Using *nup98t* and *smgGDS* as bait would show whether *nrg* has affinity for proteins that its counterparts do not interact with.

In addition to determining the proteins that *nrg* interacts with, it would also be instructive to assess the changes in gene expression that are caused by *nrg*. This can be achieved using cDNA arrays that contain genes involved in diverse signalling pathways. By comparing the pattern of hybridisation of cDNA probes derived from *t(4;11)* leukaemic and remission RNA to the same array, this would detect some of the genes and/ or pathways that are deregulated by *nrg*. This approach has been successfully used to identify genes, including transcription factors and growth factors, whose expression is deregulated by the *PAX3-FKHR* fusion oncogene associated with alveolar rhabdomyosarcoma (Khan *et al.*, 1999).

### 6.3.1.1 A common pathway for T-cell ALL

The *c-myc*, *SCL/TAL1*, *TAL2*, *LYL1*, and *BHLHB1* genes all encode members of the bHLH (basic helix-loop-helix motif) family of DNA binding transcription factors. All five of these genes become inappropriately expressed as a result of T-cell ALL associated translocations (Table 1.2). Similarly, the *RBTN1/LMO1* and *RBTN2/LMO2* genes, which encode members of the LIM family of DNA binding transcription factors, are also inappropriately expressed as a result of T-cell ALL associated translocations. In all cases, the inappropriate expression is caused by juxtaposition with the enhancer elements of TCR genes (reviewed in Rabbitts, 1994).

A common pathway for T-cell ALL has been proposed based on the fact that the bHLH and LIM domain transcription factors are able to inhibit the normal growth suppressing function of the *e2a* protein (reviewed in Hwang and Baer, 1995). This model is well supported by the following; (i) it has been shown that *e2a* deficient mice rapidly develop T-cell ALL (Bain *et al.*, 1997), and that this T-cell ALL is very similar to that which develops rapidly in *SCL LMO1* double transgenic mice (Chervinsky *et al.*, 1999); (ii) when *e2a* activity is restored in T-cell ALL cells that overexpress the E2A inhibiting *tall* protein, the cells undergo growth inhibition and apoptosis (Park *et al.*, 1999); and (iii) it was recently shown that *bhlhb1* can inhibit *e2a* function in T-lymphoblasts (Wang *et al.*, 2000).

It is possible that *nrg* may act through this common pathway and increase transcription of bHLH and LIM family transcription factors, thereby inhibiting the function of E2A. To test whether a common set of bHLH and/or LIM genes are activated by t(4;11) and the translocations described above, the expression profiles of bHLH and LIM genes should be compared between t(4;11) patients and patients with translocations that directly disrupt these genes. Expression of E2A should also be compared between these patients.

### 6.3.2 Does *nrg* effectively cause increased expression/activity of smgGDS?

The (4;11) translocation places *RAP1GDS1* under the control of the *NUP98* promoter and may effectively result in increased *RAP1GDS1* expression. This may in turn lead to increased cytoplasmic levels of smgGDS, and this could disrupt the balance of GDP/GTP exchange of ras, rho, and rap1 family proteins to favour their active GTP bound form. As both *RAP1GDS1* and *NRG* are expressed as multiple mRNA isoforms, some that are indistinguishable from each other because they are so similar in size, it would be difficult to accurately determine their overall relative expression levels using Northern analysis. In order to accurately compare the expression level of *RAP1GDS1* in normal T-cells to that of *NRG* in t(4;11) cells, it would be necessary to perform quantitative RT-PCR.

Rather than increasing the overall level of cytoplasmic smgGDS, *nrg* may be a more active GDP/GTP exchange factor than smgGDS. The relative activation potentials of smgGDS and *nrg* for ras could be measured and compared using the ras-binding domain assay described by de Rooij and Bos (1997). Briefly, this assay involves the use of a fusion protein consisting of glutathione-S-transferase (GST) fused to the ras-binding domain of raf1. This fusion protein is called GST-rbd. GST-rbd acts as an activation specific probe for ras because it only interacts with rasGTP and not rasGDP. Therefore the amount of rasGTP present can be measured by determining the amount of ras bound to GST-rbd (de Rooij and Bos, 1997). Recent studies using this assay have shown that expression of the bcr-abl and tel-abl fusion proteins results in higher levels of GTP-bound ras than that resulting from expression of wildtype abl (Voss *et al.*, 2000).



### **6.3.3 Further studies on nrg localisation**

The localisation studies described in this thesis, although sufficient to frame hypotheses on the leukemogenic mechanisms of *nrg*, may not provide an accurate representation of the cellular distribution of *nrg* in t(4;11) leukaemic cells. In the future it would be desirable to use antibody staining of blast cells from t(4;11) patients to check this. Although antibodies for *nup98* and *smgGDS* are available, it is uncertain if these would cross-react with and therefore detect *nrg*, since *nrg* may have different epitopes to its counterparts. In addition, even if the wildtype antibodies did detect *nrg* it would be very difficult to interpret the localisation data because the signal from *nrg*, *nup98* and *smgGDS* would not be easily distinguishable. These problems could be overcome by using an antibody that reacts only with *nrg* and not with *nup98* or *smgGDS*. Such an antibody is currently being developed collaboratively in the laboratory of Dr John Schrader (Biomedical Research Institute, University of British Columbia, Vancouver).

### **6.3.4 Assessing the transforming properties of nrg**

All three t(4;11) patients described in this thesis were shown to express *NRG* mRNA, indicating that expression of *NRG* is recurrently associated with this translocation. There are a total of six known reported cases of t(4;11) in T-cell leukaemia patients and in all of these cases this is the primary translocation at presentation. Further to this, t(4;11) is the only detected cytogenetic abnormality detected at presentation for two of the patients. These observations suggest that the formation of *nrg* may be the sole factor required for initiating leukemogenesis. Assays which test the transforming ability of *nrg* will indicate if *nrg* is the sole leukemogenic factor or whether additional genetic changes are required.

#### 6.3.4.1 *in vitro* transformation studies

There are two principal model systems that are used to study the *in vitro* transforming abilities of fusion proteins arising from chromosome translocations; (i) transformation of mouse NIH-3T3 fibroblast cells and (ii) abolition of growth factor dependence in haematopoietic cell lines.

Transformation of mouse NIH-3T3 cells can be assessed in a number of ways including focus formation (loss of contact inhibition), and ability to grow in low serum. NIH-3T3 cells that overexpress smgGDS exhibit mild focus formation and increased ability to grow in low serum compared to wildtype NIH-3T3 cells (Fujioka *et al.*, 1992). It is important to test nrg in these assays to see if it is more strongly transforming than smgGDS.

Other signs of transformation include morphological changes, growth in soft agar (anchorage independent growth) and tumour formation in nude mice. The aml1-eto fusion protein, which arises from the common (8;21) translocation, induces all three of these transformed phenotypes when expressed in NIH-3T3 cells (Frank *et al.*, 1999). It is expected that nrg will possess at least one of these transforming properties since expression of nup98-hoxa9 results in anchorage independence (Kasper *et al.*, 1999).

The ability to convert factor dependent haematopoietic cell lines to a growth factor independent phenotype is another standard transformation assay. The bcr-abl (Hariharan *et al.*, 1988) and tel-jak2 (Schwaller *et al.*, 1998) fusion proteins can abolish IL-3 dependence in established haematopoietic cell lines. *RAS* oncogenes abolish growth factor dependence in some instances (Boswell *et al.*, 1990), so if nrg results in increased signalling through the ras pathway then it may impart a factor independent phenotype.

#### 6.3.4.2 *in vivo* transformation studies

Reconstitution of the bone marrow of lethally irradiated mice with stem cells that have been retrovirally induced to express fusion proteins is a powerful approach used for assessing the transformation of haematopoietic cells *in vivo*. Daley *et al.* (1990) infected murine bone marrow with a retrovirus encoding p210<sup>bcr-abl</sup> and then transplanted this into an irradiated syngeneic strain of mice. This resulted in mice which developed a range of haematopoietic malignancies including a myeloproliferative disorder resembling the chronic phase of human CML (Daley *et al.*, 1990). The tumour tissue contained the retroviral construct, thus implicating the bcr-abl protein as the primary cause of malignancy in CML. These results were confirmed by a more recent study which used the murine stem cell virus retroviral promoter and showed that mice transplanted with p210<sup>bcr-abl</sup> all developed CML between 3 and 5 weeks after transplantation (Pear *et al.*, 1998).

Kroon *et al.* (1999) found that mice reconstituted with bone marrow expressing the nup98-hoxa9 fusion protein, which is associated with AML, developed a myeloproliferative disorder within three months post-transplantation. By eight months post-transplantation, 5 of the 10 mice transplanted with *NUP98-HOXA9*-transduced bone marrow cells had died due to the myeloproliferative disorder and another two had acquired AML (Kroon *et al.*, 1999).

Reconstitution of mouse bone marrow with stem cells expressing nrg will be informative on a number of levels. First and foremost it will show whether nrg is capable of causing leukaemia in mice. Secondly it will tell us whether nrg can result in haematological malignancies other than T-cell ALL. If stem cells expressing nrg only give rise to T-cell ALL then this will be direct evidence of the specificity of nrg for T-cell ALL. Finally, the latency of leukaemia development will indicate if extra genetic changes are required.

Like the fusion proteins previously discovered, elucidating the biological mechanisms by which nrg causes leukaemia promises to yield new insights into the mechanisms of leukemogenesis and thereby indicate pathways for which new therapies can be developed.

## References

## References

- Abraham, K.M., Levin, S.D., Marth, J.D., Forbush, K.A. and Perlmutter, R.M. (1991). Thymic tumorigenesis induced by overexpression of p56lck. *Proc Natl Acad Sci USA*, **88**, 3977-81.
- Ahuja, H.G., Felix, C.A. and Aplan, P.D. (1999). The t(11;20)(p15;q11) chromosomal translocation associated with therapy-related myelodysplastic syndrome results in an NUP98-TOP1 fusion. *Blood*, **94**, 3258-61.
- Amati, B., Alevizopoulos, K. and Vlach, J. (1998). Myc and the cell cycle. *Front Biosci*, **3**, D250-68.
- Amson, R.B., Marcelle, C. and Telerman, A. (1989). Identification of a 130 KDa bcr related gene product. *Oncogene*, **4**, 243-7.
- Aoki, M., Hecht, A., Kruse, U., Kemler, R. and Vogt, P.K. (1999). Nuclear endpoint of Wnt signaling: neoplastic transformation induced by transactivating lymphoid-enhancing factor 1. *Proc Natl Acad Sci USA*, **96**, 139-44.
- Aplan, P.D., Lombardi, D.P., Ginsberg, A.M., Cossman, J., Bertness, V.L. and Kirsch, I.R. (1990). Disruption of the human SCL locus by "illegitimate" V-(D)-J recombinase activity. *Science*, **250**, 1426-9.
- Arai, Y., Hosoda, F., Kobayashi, H., Arai, K., Hayashi, Y., Kamada, N., Kaneko, Y. and Ohki, M. (1997). The inv(11)(p15q22) chromosome translocation of de novo and therapy-related myeloid malignancies results in fusion of the nucleoporin gene, NUP98, with the putative RNA helicase gene, DDX10. *Blood*, **89**, 3936-44.
- Bachi, A., Braun, I.C., Rodrigues, J.P., Pante, N., Ribbeck, K., von Kobbe, C., Kutay, U., Wilm, M., Gorlich, D., Carmo-Fonseca, M. and Izaurralde, E. (2000). The C-terminal domain of TAP interacts with the nuclear pore complex and promotes export of specific CTE-bearing RNA substrates. *RNA*, **6**, 136-58.

- Baer, R. (1993). TAL1, TAL2 and LYL1: a family of basic helix-loop-helix proteins implicated in T cell acute leukaemia. *Semin Cancer Biol*, **4**, 341-7.
- Bain, B.J. (1999). *Leukaemia Diagnosis*. Blackwell Science Ltd: Oxford.
- Bain, G., Engel, I., Robanus Maandag, E.C., te Riele, H.P., Volland, J.R., Sharp, L.L., Chun, J., Huey, B., Pinkel, D. and Murre, C. (1997). E2A deficiency leads to abnormalities in alphabeta T-cell development and to rapid development of T-cell lymphomas. *Mol Cell Biol*, **17**, 4782-91.
- Begley, C.G., Aplan, P.D., Davey, M.P., Nakahara, K., Tchorz, K., Kurtzberg, J., Hershfield, M.S., Haynes, B.F., Cohen, D.I., Waldmann, T.A. and *et al.* (1989). Chromosomal translocation in a human leukemic stem-cell line disrupts the T-cell antigen receptor delta-chain diversity region and results in a previously unreported fusion transcript. *Proc Natl Acad Sci U S A*, **86**, 2031-5.
- Bellefroid, E.J., Lecocq, P.J., Benhida, A., Poncelet, D.A., Belayew, A. and Martial, J.A. (1989). The human genome contains hundreds of genes coding for finger proteins of the Kruppel type. *DNA*, **8**, 377-87.
- Bellefroid, E.J., Marine, J.C., Ried, T., Lecocq, P.J., Riviere, M., Amemiya, C., Poncelet, D.A., Coulie, P.G., de Jong, P., Szpirer, C. and *et al.* (1993). Clustered organization of homologous KRAB zinc-finger genes with enhanced expression in human T lymphoid cells. *Embo J*, **12**, 1363-74.
- Bellefroid, E.J., Poncelet, D.A., Lecocq, P.J., Revelant, O. and Martial, J.A. (1991). The evolutionarily conserved Kruppel-associated box domain defines a subfamily of eukaryotic multifingered proteins. *Proc Natl Acad Sci U S A*, **88**, 3608-12.
- Bene, M.C., Bernier, M., Castoldi, G., Faure, G.C., Knapp, W., Ludwig, W.D., Matutes, E., Orfao, A. and van't Veer, M. (1999). Impact of immunophenotyping on management of acute leukemias. *Haematologica*, **84**, 1024-34.

- Bennett, J.M., Catovsky, D., Daniel, M.T., Flandrin, G., Galton, D.A., Gralnick, H.R. and Sultan, C. (1985). Proposed revised criteria for the classification of acute myeloid leukemia. A report of the French-American-British Cooperative Group. *Ann Intern Med*, **103**, 620-5.
- Besnard-Guerin, C., Newsham, I., Winqvist, R. and Cavenee, W.K. (1996). A common region of loss of heterozygosity in Wilms' tumor and embryonal rhabdomyosarcoma distal to the D11S988 locus on chromosome 11p15.5. *Hum Genet*, **97**, 163-70.
- Blobel, G.A. (2000). CREB-binding protein and p300: molecular integrators of hematopoietic transcription. *Blood*, **95**, 745-55.
- Bloomfield, C.D., Goldman, A.I., Alimena, G., Berger, R., Borgstrom, G.H., Brandt, L., Catovsky, D., de la Chapelle, A., Dewald, G.W., Garson, O.M. and *et al.* (1986). Chromosomal abnormalities identify high-risk and low-risk patients with acute lymphoblastic leukemia. *Blood*, **67**, 415-20.
- Boehm, T., Foroni, L., Kaneko, Y., Perutz, M.F. and Rabbitts, T.H. (1991). The rhombotin family of cysteine-rich LIM-domain oncogenes: distinct members are involved in T-cell translocations to human chromosomes 11p15 and 11p13. *Proc Natl Acad Sci U S A*, **88**, 4367-71.
- Boer, J., Bonten-Surtel, J. and Grosveld, G. (1998). Overexpression of the nucleoporin CAN/NUP214 induces growth arrest, nucleocytoplasmic transport defects, and apoptosis. *Mol Cell Biol*, **18**, 1236-47.
- Borrow, J., Shearman, A.M., Stanton, V.P., Jr., Becher, R., Collins, T., Williams, A.J., Dube, I., Katz, F., Kwong, Y.L., Morris, C., Ohyashiki, K., Toyama, K., Rowley, J. and Housman, D.E. (1996). The t(7;11)(p15;p15) translocation in acute myeloid leukaemia fuses the genes for nucleoporin NUP98 and class I homeoprotein HOXA9. *Nat Genet*, **12**, 159-67.



- Boswell, H.S., Nahreini, T.S., Burgess, G.S., Srivastava, A., Gabig, T.G., Inhorn, L., Srour, E.F. and Harrington, M.A. (1990). A RAS oncogene imparts growth factor independence to myeloid cells that abnormally regulate protein kinase C: a nonautocrine transformation pathway. *Exp Hematol*, **18**, 452-60.
- Brownell, E., Mittereder, N. and Rice, N.R. (1989). A human rel proto-oncogene cDNA containing an Alu fragment as a potential coding exon. *Oncogene*, **4**, 935-42.
- Burnett, R.C., David, J.C., Harden, A.M., Le Beau, M.M., Rowley, J.D. and Diaz, M.O. (1991). The LCK gene is involved in the t(1;7)(p34;q34) in the T-cell acute lymphoblastic leukemia derived cell line, HSB-2. *Genes Chromosomes Cancer*, **3**, 461-7.
- Call, K.M., Glaser, T., Ito, C.Y., Buckler, A.J., Pelletier, J., Haber, D.A., Rose, E.A., Kral, A., Yeager, H., Lewis, W.H. and *et al.* (1990). Isolation and characterization of a zinc finger polypeptide gene at the human chromosome 11 Wilms' tumor locus. *Cell*, **60**, 509-20.
- Campana, D. and Pui, C.H. (1995). Detection of minimal residual disease in acute leukemia: methodologic advances and clinical significance. *Blood*, **85**, 1416-34.
- Catovsky, D.F.R. (1990). *The lymphoid leukaemias*. Pub. by Butterworth London.
- Chen, Z., Brand, N.J., Chen, A., Chen, S.J., Tong, J.H., Wang, Z.Y., Waxman, S. and Zelent, A. (1993). Fusion between a novel Kruppel-like zinc finger gene and the retinoic acid receptor-alpha locus due to a variant t(11;17) translocation associated with acute promyelocytic leukaemia. *Embo J*, **12**, 1161-7.
- Chervinsky, D.S., Zhao, X.F., Lam, D.H., Ellsworth, M., Gross, K.W. and Aplan, P.D. (1999). Disordered T-cell development and T-cell malignancies in SCL LMO1 double-transgenic mice: parallels with E2A-deficient mice. *Mol Cell Biol*, **19**, 5025-35.

- Childs, C.C., Hirsch-Ginsberg, C., Culbert, S.J., Ahearn, M., Reuben, J., Trujillo, J.M., Cork, A., Walters, R.R., Freireich, E.J. and Stass, S.A. (1988). Lineage heterogeneity in acute leukemia with the t(4;11) abnormality: implications for acute mixed lineage leukemia. *Hematol Pathol*, **2**, 145-57.
- Chuang, T.H., Xu, X., Quilliam, L.A. and Bokoch, G.M. (1994). SmgGDS stabilizes nucleotide-bound and -free forms of the Rac1 GTP- binding protein and stimulates GTP/GDP exchange through a substituted enzyme mechanism. *Biochem J*, **303**, 761-7.
- Chung, W.Y., Yuan, L., Feng, L., Hensle, T. and Tycko, B. (1996). Chromosome 11p15.5 regional imprinting: comparative analysis of KIP2 and H19 in human tissues and Wilms' tumors. *Hum Mol Genet*, **5**, 1101-8.
- Cobaleda, C., Gutierrez-Cianca, N., Perez-Losada, J., Flores, T., Garcia-Sanz, R., Gonzalez, M. and Sanchez-Garcia, I. (2000). A primitive hematopoietic cell is the target for the leukemic transformation in human philadelphia-positive acute lymphoblastic leukemia. *Blood*, **95**, 1007-13.
- Copelan, E.A. and McGuire, E.A. (1995). The biology and treatment of acute lymphoblastic leukemia in adults. *Blood*, **85**, 1151-68.
- Cormack, B.P., Valdivia, R.H. and Falkow, S. (1996). FACS-optimized mutants of the green fluorescent protein (GFP). *Gene*, **173**, 33-8.
- Corral, J., Lavenir, I., Impey, H., Warren, A.J., Forster, A., Larson, T.A., Bell, S., McKenzie, A.N., King, G. and Rabbitts, T.H. (1996). An Mll-AF9 fusion gene made by homologous recombination causes acute leukemia in chimeric mice: a method to create fusion oncogenes. *Cell*, **85**, 853-61.
- Crosier, P.S., Ricciardi, S.T., Hall, L.R., Vitas, M.R., Clark, S.C. and Crosier, K.E. (1993). Expression of isoforms of the human receptor tyrosine kinase c-kit in leukemic cell lines and acute myeloid leukemia. *Blood*, **82**, 1151-8.

- Daley, G.Q., Van Etten, R.A. and Baltimore, D. (1990). Induction of chronic myelogenous leukemia in mice by the P210bcr/abl gene of the Philadelphia chromosome. *Science*, **247**, 824-30.
- Dallery, E., Galiegue-Zouitina, S., Collyn-d'Hooghe, M., Quief, S., Denis, C., Hildebrand, M.P., Lantoine, D., Deweindt, C., Tilly, H., Bastard, C. and *et al.* (1995). TTF, a gene encoding a novel small G protein, fuses to the lymphoma-associated LAZ3 gene by t(3;4) chromosomal translocation. *Oncogene*, **10**, 2171-8.
- Daniel, J.M. and Reynolds, A.B. (1999). The catenin p120(ctn) interacts with Kaiso, a novel BTB/POZ domain zinc finger transcription factor. *Mol Cell Biol*, **19**, 3614-23.
- Dao, D., Walsh, C.P., Yuan, L., Gorelov, D., Feng, L., Hensle, T., Nisen, P., Yamashiro, D.J., Bestor, T.H. and Tycko, B. (1999). Multipoint analysis of human chromosome 11p15/mouse distal chromosome 7: inclusion of H19/IGF2 in the minimal WT2 region, gene specificity of H19 silencing in Wilms' tumorigenesis and methylation hyper-dependence of H19 imprinting. *Hum Mol Genet*, **8**, 1337-52.
- David, G., Alland, L., Hong, S.H., Wong, C.W., DePinho, R.A. and Dejean, A. (1998). Histone deacetylase associated with mSin3A mediates repression by the acute promyelocytic leukemia-associated PLZF protein. *Oncogene*, **16**, 2549-56.
- de Klein, A., van Kessel, A.G., Grosveld, G., Bartram, C.R., Hagemeijer, A., Bootsma, D., Spurr, N.K., Heisterkamp, N., Groffen, J. and Stephenson, J.R. (1982). A cellular oncogene is translocated to the Philadelphia chromosome in chronic myelocytic leukaemia. *Nature*, **300**, 765-7.
- de Rooij, J. and Bos, J.L. (1997). Minimal Ras-binding domain of Raf1 can be used as an activation-specific probe for Ras. *Oncogene*, **14**, 623-5.
- de Vries, J.E., ten Kate, J. and Bosman, F.T. (1996). p21ras in carcinogenesis. *Pathol Res Pract*, **192**, 658-68.

- Dhordain, P., Albagli, O., Lin, R.J., Ansieau, S., Quief, S., Leutz, A., Kerckaert, J.P., Evans, R.M. and Leprince, D. (1997). Corepressor SMRT binds the BTB/POZ repressing domain of the LAZ3/BCL6 oncoprotein. *Proc Natl Acad Sci U S A*, **94**, 10762-7.
- Dierick, H. and Bejsovec, A. (1999). Cellular mechanisms of wingless/Wnt signal transduction. *Curr Top Dev Biol*, **43**, 153-90.
- Dobrovic, A., Finch, J., Peters, G.B., Hardingham, J.E., Kalatzis, V., Fitzgerald, D. and Sage, R.E. (1994). Localisation of chromosome 11 breakpoint in a translocation t(4:11)(q21:p15) in T cell acute lymphoblastic leukemia. *Amer J Hum Genet* **55: suppl**, **298**.
- Dobrovic, A., Peters, G.B. and Ford, J.H. (1991). Molecular analysis of the Philadelphia chromosome. *Chromosoma*, **100**, 479-86.
- Dobson, C.L., Warren, A.J., Pannell, R., Forster, A. and Rabbitts, T.H. (2000). Tumorigenesis in mice with a fusion of the leukaemia oncogene Mll and the bacterial lacZ gene. *Embo J*, **19**, 843-51.
- Domer, P.H., Fakharzadeh, S.S., Chen, C.S., Jockel, J., Johansen, L., Silverman, G.A., Kersey, J.H. and Korsmeyer, S.J. (1993). Acute mixed-lineage leukemia t(4;11)(q21;q23) generates an MLL-AF4 fusion product. *Proc Natl Acad Sci U S A*, **90**, 7884-8.
- Dube, I.D., Kamel-Reid, S., Yuan, C.C., Lu, M., Wu, X., Corpus, G., Raimondi, S.C., Crist, W.M., Carroll, A.J., Minowada, J. and *et al.* (1991). A novel human homeobox gene lies at the chromosome 10 breakpoint in lymphoid neoplasias with chromosomal translocation t(10;14). *Blood*, **78**, 2996-3003.
- Eger, A., Stockinger, A., Schaffhauser, B., Beug, H. and Foisner, R. (2000). Epithelial mesenchymal transition by c-Fos estrogen receptor activation involves nuclear translocation of beta-catenin and upregulation of beta-catenin/lymphoid enhancer binding factor-1 transcriptional activity. *J Cell Biol*, **148**, 173-88.

- Ellisen, L.W., Bird, J., West, D.C., Soreng, A.L., Reynolds, T.C., Smith, S.D. and Sklar, J. (1991). TAN-1, the human homolog of the *Drosophila* notch gene, is broken by chromosomal translocations in T lymphoblastic neoplasms. *Cell*, **66**, 649-61.
- Endoh, H., Maruyama, K., Masuhiro, Y., Kobayashi, Y., Goto, M., Tai, H., Yanagisawa, J., Metzger, D., Hashimoto, S. and Kato, S. (1999). Purification and identification of p68 RNA helicase acting as a transcriptional coactivator specific for the activation function 1 of human estrogen receptor alpha. *Mol Cell Biol*, **19**, 5363-72.
- Erickson, P., Gao, J., Chang, K.S., Look, T., Whisenant, E., Raimondi, S., Lasher, R., Trujillo, J., Rowley, J. and Drabkin, H. (1992). Identification of breakpoints in t(8;21) acute myelogenous leukemia and isolation of a fusion transcript, AML1/ETO, with similarity to *Drosophila* segmentation gene, runt. *Blood*, **80**, 1825-31.
- Faderl, S., Kantarjian, H.M., Talpaz, M. and Estrov, Z. (1998). Clinical significance of cytogenetic abnormalities in adult acute lymphoblastic leukemia. *Blood*, **91**, 3995-4019.
- Fontoura, B.M., Blobel, G. and Matunis, M.J. (1999). A conserved biogenesis pathway for nucleoporins: proteolytic processing of a 186-kilodalton precursor generates Nup98 and the novel nucleoporin, Nup96. *J Cell Biol*, **144**, 1097-112.
- Fornerod, M., van Deursen, J., van Baal, S., Reynolds, A., Davis, D., Murti, K.G., Fransen, J. and Grosveld, G. (1997). The human homologue of yeast CRM1 is in a dynamic subcomplex with CAN/Nup214 and a novel nuclear pore component Nup88. *Embo J*, **16**, 807-16.
- Frank, R.C., Sun, X., Berguido, F.J., Jakubowiak, A. and Nimer, S.D. (1999). The t(8;21) fusion protein, AML1/ETO, transforms NIH3T3 cells and activates AP-1. *Oncogene*, **18**, 1701-10.
- Freter, R.R., Irrminger, J.C., Porter, J.A., Jones, S.D. and Stiles, C.D. (1992). A novel 7-nucleotide motif located in 3' untranslated sequences of the immediate-early gene set

- mediates platelet-derived growth factor induction of the JE gene. *Mol Cell Biol*, **12**, 5288-300.
- Frohman, M.A., Dush, M.K. and Martin, G.R. (1988). Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc Natl Acad Sci U S A*, **85**, 8998-9002.
- Fujioka, H., Kaibuchi, K., Kishi, K., Yamamoto, T., Kawamura, M., Sakoda, T., Mizuno, T. and Takai, Y. (1992). Transforming and c-fos promoter/enhancer-stimulating activities of a stimulatory GDP/GTP exchange protein for small GTP-binding proteins. *J Biol Chem*, **267**, 926-30.
- Galaktionov, K., Chen, X. and Beach, D. (1996). Cdc25 cell-cycle phosphatase as a target of c-myc. *Nature*, **382**, 511-7.
- Gao, J., Erickson, P., Gardiner, K., Le Beau, M.M., Diaz, M.O., Patterson, D., Rowley, J.D. and Drabkin, H.A. (1991). Isolation of a yeast artificial chromosome spanning the 8;21 translocation breakpoint t(8;21)(q22;q22.3) in acute myelogenous leukemia. *Proc Natl Acad Sci U S A*, **88**, 4882-6.
- Gebuhr, T.C., Bultman, S.J. and Magnuson, T. (2000). Pc-G/trx-G and the SWI/SNF connection: developmental gene regulation through chromatin remodeling. *Genesis*, **26**, 189-97.
- Gilliland, D.G. (1998). Molecular genetics of human leukemia. *Leukemia*, **12 Suppl 1**, S7-12.
- Golub, T.R., Barker, G.F., Lovett, M. and Gilliland, D.G. (1994). Fusion of PDGF receptor beta to a novel ets-like gene, tel, in chronic myelomonocytic leukemia with t(5;12) chromosomal translocation. *Cell*, **77**, 307-16.
- Gorlich, D., Henklein, P., Laskey, R.A. and Hartmann, E. (1996). A 41 amino acid motif in importin-alpha confers binding to importin- beta and hence transit into the nucleus. *Embo J*, **15**, 1810-7.

- Grignani, F., Fagioli, M., Alcalay, M., Longo, L., Pandolfi, P.P., Donti, E., Biondi, A., Lo Coco, F. and Pelicci, P.G. (1994). Acute promyelocytic leukemia: from genetics to treatment. *Blood*, **83**, 10-25.
- Groffen, J., Stephenson, J.R., Heisterkamp, N., de Klein, A., Bartram, C.R. and Grosveld, G. (1984). Philadelphia chromosomal breakpoints are clustered within a limited region, bcr, on chromosome 22. *Cell*, **36**, 93-9.
- Gu, Y., Nakamura, T., Alder, H., Prasad, R., Canaani, O., Cimino, G., Croce, C.M. and Canaani, E. (1992). The t(4;11) chromosome translocation of human acute leukemias fuses the ALL-1 gene, related to *Drosophila trithorax*, to the AF-4 gene. *Cell*, **71**, 701-8.
- Hardingham, J.E., Peters, G.B., Dobrovic, A., Dale, B.M., Kotasek, D., Ford, H.E., Story, C.J. and Sage, R.E. (1991). A rare translocation (4;11)(q21;p14-15) in an acute lymphoblastic leukemia expressing T-cell and myeloid markers. *Cancer Genet Cytogenet*, **56**, 255-62.
- Hariharan, I.K., Adams, J.M. and Cory, S. (1988). bcr-abl oncogene renders myeloid cell line factor independent: potential autocrine mechanism in chronic myeloid leukemia. *Oncogene Res*, **3**, 387-99.
- Harris, N.L., Jaffe, E.S., Diebold, J., Flandrin, G., Muller-Hermelink, H.K., Vardiman, J., Lister, T.A. and Bloomfield, C.D. (2000). The World Health Organization classification of hematological malignancies report of the Clinical Advisory Committee Meeting, Airlie House, Virginia, November 1997. *Mod Pathol*, **13**, 193-207.
- Hatano, M., Roberts, C.W., Minden, M., Crist, W.M. and Korsmeyer, S.J. (1991). Deregulation of a homeobox gene, HOX11, by the t(10;14) in T cell leukemia. *Science*, **253**, 79-82.

- Hatano, Y., Miura, I., Nakamura, T., Yamazaki, Y., Takahashi, N. and Miura, A.B. (1999). Molecular heterogeneity of the NUP98/HOXA9 fusion transcript in myelodysplastic syndromes associated with t(7;11)(p15;p15). *Br J Haematol*, **107**, 600-604.
- Heisterkamp, N., Jenster, G., ten Hoeve, J., Zovich, D., Pattengale, P.K. and Groffen, J. (1990). Acute leukaemia in bcr/abl transgenic mice. *Nature*, **344**, 251-3.
- Honda, H., Oda, H., Suzuki, T., Takahashi, T., Witte, O.N., Ozawa, K., Ishikawa, T., Yazaki, Y. and Hirai, H. (1998). Development of acute lymphoblastic leukemia and myeloproliferative disorder in transgenic mice expressing p210bcr/abl: a novel transgenic model for human Ph1-positive leukemias. *Blood*, **91**, 2067-75.
- Hoovers, J.M., Mannens, M., John, R., Bliet, J., van Heyningen, V., Porteous, D.J., Leschot, N.J., Westerveld, A. and Little, P.F. (1992). High-resolution localization of 69 potential human zinc finger protein genes: a number are clustered. *Genomics*, **12**, 254-63.
- Horvai, A.E., Xu, L., Korzus, E., Brard, G., Kalafus, D., Mullen, T.M., Rose, D.W., Rosenfeld, M.G. and Glass, C.K. (1997). Nuclear integration of JAK/STAT and Ras/AP-1 signaling by CBP and p300. *Proc Natl Acad Sci U S A*, **94**, 1074-9.
- Hu, R.J., Lee, M.P., Connors, T.D., Johnson, L.A., Bum, T.C., Su, K., Landes, G.M. and Feinberg, A.P. (1997). A 2.5-Mb transcript map of a tumor-suppressing subchromosomal transferable fragment from 11p15.5, and isolation and sequence analysis of three novel genes. *Genomics*, **46**, 9-17.
- Hua, C., Zorn, S., Jensen, J.P., Coupland, R.W., Ko, H.S., Wright, J.J. and Bakhshi, A. (1988). Consequences of the t(14;18) chromosomal translocation in follicular lymphoma: deregulated expression of a chimeric and mutated BCL-2 gene. *Oncogene Res*, **2**, 263-75.



- Hunger, S.P. (1996). Chromosomal translocations involving the E2A gene in acute lymphoblastic leukemia: clinical features and molecular pathogenesis. *Blood*, **87**, 1211-24.
- Hussey, D.J., Nicola, M., Moore, S., Peters, G.B. and Dobrovic, A. (1999). The (4;11)(q21;p15) translocation fuses the NUP98 and RAP1GDS1 genes and is recurrent in T-cell acute lymphocytic leukemia. *Blood*, **94**, 2072-9.
- Hussey, D.J., Parker, N.J., Hussey, N.D., Little, P.F.R. and Dobrovic, A. (1997). Characterization of a KRAB family zinc finger gene, ZNF195, mapping to chromosome band 11p15.5. *Genomics*, **45**, 451-455.
- Hwang, L.Y. and Baer, R.J. (1995). The role of chromosome translocations in T cell acute leukemia. *Curr Opin Immunol*, **7**, 659-64.
- Inoue, S., Tyrkus, M., Ravindranath, Y. and Gohle, N. (1985). A variant translocation between chromosomes 4 and 11, t(4q;11p) in a child with acute leukemia. *Am J Pediatr Hematol Oncol*, **7**, 211-4.
- Iouzalet, N., Camonis, J. and Moreau, J. (1998). Identification and characterization in *Xenopus* of XsmgGDS, a RalB binding protein. *Biochem Biophys Res Commun*, **250**, 359-63.
- James, M.R., Richard, C.W., 3rd, Schott, J.J., Yousry, C., Clark, K., Bell, J., Terwilliger, J.D., Hazan, J., Dubay, C., Vignal, A. and *et al.* (1994). A radiation hybrid map of 506 STS markers spanning human chromosome 11. *Nat Genet*, **8**, 70-6.
- Kalatzis, V., Peters, G.B. and Dobrovic, A. (1993). Mapping of the chromosome 11 breakpoint of the t(4;11)(q21;p14-15) translocation. *Cancer Genet Cytogenet*, **69**, 122-5.
- Kamei, Y., Xu, L., Heinzl, T., Torchia, J., Kurokawa, R., Gloss, B., Lin, S.C., Heyman, R.A., Rose, D.W., Glass, C.K. and Rosenfeld, M.G. (1996). A CBP integrator

- complex mediates transcriptional activation and AP-1 inhibition by nuclear receptors. *Cell*, **85**, 403-14.
- Kashige, N., Carpino, N. and Kobayashi, R. (2000). Tyrosine phosphorylation of p62dok by p210bcr-abl inhibits RasGAP activity. *Proc Natl Acad Sci U S A*, **97**, 2093-2098.
- Kasper, L.H., Brindle, P.K., Schnabel, C.A., Pritchard, C.E., Cleary, M.L. and van Deursen, J.M. (1999). CREB binding protein interacts with nucleoporin-specific FG repeats that activate transcription and mediate NUP98-HOXA9 oncogenicity. *Mol Cell Biol*, **19**, 764-76.
- Kennedy, M.A., Gonzalez-Sarmiento, R., Kees, U.R., Lampert, F., Dear, N., Boehm, T. and Rabbitts, T.H. (1991). HOX11, a homeobox-containing T-cell oncogene on human chromosome 10q24. *Proc Natl Acad Sci U S A*, **88**, 8900-4.
- Kerckaert, J.P., Deweindt, C., Tilly, H., Quief, S., Lecocq, G. and Bastard, C. (1993). LAZ3, a novel zinc-finger encoding gene, is disrupted by recurring chromosome 3q27 translocations in human lymphomas. *Nat Genet*, **5**, 66-70.
- Khan, J., Bittner, M.L., Saal, L.H., Teichmann, U., Azorsa, D.O., Gooden, G.C., Pavan, W.J., Trent, J.M. and Meltzer, P.S. (1999). cDNA microarrays detect activation of a myogenic transcription program by the PAX3-FKHR fusion oncogene. *Proc Natl Acad Sci U S A*, **96**, 13264-9.
- Kitabayashi, I., Yokoyama, A., Shimizu, K. and Ohki, M. (1998). Interaction and functional cooperation of the leukemia-associated factors AML1 and p300 in myeloid cell differentiation. *Embo J*, **17**, 2994-3004.
- Kitayama, H., Sugimoto, Y., Matsuzaki, T., Ikawa, Y. and Noda, M. (1989). A ras-related gene with transformation suppressor activity. *Cell*, **56**, 77-84.
- Kjellman, C., Sjogren, H.O., Salford, L.G. and Widegren, B. (1999a). HERV-F (XA34) is a full-length human endogenous retrovirus expressed in placental and fetal tissues. *Gene*, **239**, 99-107.

- Kjellman, C., Sjogren, H.O. and Widegren, B. (1999b). HERV-F, a new group of human endogenous retrovirus sequences. *J Gen Virol*, **80**, 2383-92.
- Kjoller, L. and Hall, A. (1999). Signaling to Rho GTPases. *Exp Cell Res*, **253**, 166-79.
- Kobzev, Y.N. and Rowley, J.D. (1999). *NUP98* gene rearrangements in leukemia detected by fluorescence in situ hybridization (FISH). *Blood*, **94**, 496a.
- Konopka, J.B., Watanabe, S.M. and Witte, O.N. (1984). An alteration of the human c-abl protein in K562 leukemia cells unmasks associated tyrosine kinase activity. *Cell*, **37**, 1035-42.
- Kourlas, P.J., Strout, M.P., Becknell, B., Veronese, M.L., Croce, C.M., Theil, K.S., Krahe, R., Ruutu, T., Knuutila, S., Bloomfield, C.D. and Caligiuri, M.A. (2000). Identification of a gene at 11q23 encoding a guanine nucleotide exchange factor: Evidence for its fusion with MLL in acute myeloid leukemia. *Proc Natl Acad Sci U S A*, **97**, 2145-2150.
- Kozak, M. (1991). Structural features in eukaryotic mRNAs that modulate the initiation of translation. *J Biol Chem*, **266**, 19867-70.
- Kraemer, D., Wozniak, R.W., Blobel, G. and Radu, A. (1994). The human CAN protein, a putative oncogene product associated with myeloid leukemogenesis, is a nuclear pore complex protein that faces the cytoplasm. *Proc Natl Acad Sci U S A*, **91**, 1519-23.
- Kroon, E., Thorsteinsdottir, U., Nakamura, T. and Sauvageau, G. (1999). Over-expression of the human AML-associated fusion gene *NUP98-HOXA9* in murine bone marrow cells leads to myelo-proliferation and AML. *Blood*, **94**, 694a.
- Kung, A.L., Rebel, V.I., Bronson, R.T., Ch'ng, L.E., Sieff, C.A., Livingston, D.M. and Yao, T.P. (2000). Gene dose-dependent control of hematopoiesis and hematologic tumor suppression by CBP. *Genes Dev*, **14**, 272-7.

- Kwong, Y.L. and Pang, A. (1999). Low frequency of rearrangements of the homeobox gene HOXA9/t(7;11) in adult acute myeloid leukemia. *Genes Chromosomes Cancer*, **25**, 70-4.
- Lankiewicz, S., Gisselmann, G. and Hatt, H. (1997). Enhanced RACE method using specific enrichment by biotinylated oligonucleotides bound to streptavidin coated magnetic particles. *Nucleic Acids Res*, **25**, 2037-8.
- Lawrence, H.J., Helgason, C.D., Sauvageau, G., Fong, S., Izon, D.J., Humphries, R.K. and Largman, C. (1997). Mice bearing a targeted interruption of the homeobox gene HOXA9 have defects in myeloid, erythroid, and lymphoid hematopoiesis. *Blood*, **89**, 1922-30.
- Le Beau, M.M., McKeithan, T.W., Shima, E.A., Goldman-Leikin, R.E., Chan, S.J., Bell, G.I., Rowley, J.D. and Diaz, M.O. (1986). T-cell receptor alpha-chain gene is split in a human T-cell leukemia cell line with a t(11;14)(p15;q11). *Proc Natl Acad Sci U S A*, **83**, 9744-8.
- Lee, T.C., Li, L., Philipson, L. and Ziff, E.B. (1997). Myc represses transcription of the growth arrest gene gas1. *Proc Natl Acad Sci U S A*, **94**, 12886-91.
- Li, J., Shen, H., Himmel, K.L., Dupuy, A.J., Largaespada, D.A., Nakamura, T., Shaughnessy, J.D., Jr., Jenkins, N.A. and Copeland, N.G. (1999). Leukaemia disease genes: large-scale cloning and pathway predictions. *Nat Genet*, **23**, 348-53.
- Lister, T.A., Norton, A.J. and Rohatiner, A. (1995). The proposed revised European-American classification of lymphoma. *Blood*, **85**, 1975.
- Lo, P.C. and Frasch, M. (1998). bagpipe-Dependent expression of vimar, a novel Armadillo-repeats gene, in Drosophila visceral mesoderm. *Mech Dev*, **72**, 65-75.
- Look, A.T. (1997). Oncogenic transcription factors in the human acute leukemias. *Science*, **278**, 1059-1064.

- Lovering, R. and Trowsdale, J. (1991). A gene encoding 22 highly related zinc fingers is expressed in lymphoid cell lines. *Nucleic Acids Res*, **19**, 2921-8.
- Lu, M., Gong, Z.Y., Shen, W.F. and Ho, A.D. (1991). The tcl-3 proto-oncogene altered by chromosomal translocation in T-cell leukemia codes for a homeobox protein. *Embo J*, **10**, 2905-10.
- Lutterbach, B. and Hiebert, S.W. (2000). Role of the transcription factor AML-1 in acute leukemia and hematopoietic differentiation. *Gene*, **245**, 223-35.
- Maguire, R.T., Robins, T.S., Thorgeirsson, S.S. and Heilman, C.A. (1983). Expression of cellular myc and mos genes in undifferentiated B cell lymphomas of Burkitt and non-Burkitt types. *Proc Natl Acad Sci U S A*, **80**, 1947-50.
- Margolin, J.F., Friedman, J.R., Meyer, W.K., Vissing, H., Thiesen, H.J. and Rauscher, F.J., 3rd. (1994). Kruppel-associated boxes are potent transcriptional repression domains. *Proc Natl Acad Sci U S A*, **91**, 4509-13.
- Martin, J.F., Bradley, A. and Olson, E.N. (1995). The paired-like homeo box gene MHox is required for early events of skeletogenesis in multiple lineages. *Genes Dev*, **9**, 1237-49.
- Maru, Y. and Witte, O.N. (1991). The BCR gene encodes a novel serine/threonine kinase activity within a single exon. *Cell*, **67**, 459-68.
- McCrea, P.D., Turck, C.W. and Gumbiner, B. (1991). A homolog of the armadillo protein in *Drosophila* (plakoglobin) associated with E-cadherin. *Science*, **254**, 1359-61.
- McNeil, S., Zeng, C., Harrington, K.S., Hiebert, S., Lian, J.B., Stein, J.L., van Wijnen, A.J. and Stein, G.S. (1999). The t(8;21) chromosomal translocation in acute myelogenous leukemia modifies intranuclear targeting of the AML1/CBFalpha2 transcription factor. *Proc Natl Acad Sci U S A*, **96**, 14882-7.
- McNoe, L.A., Eccles, M.R. and Reeve, A.E. (1992). Dinucleotide repeat polymorphism at the D11S860 locus. *Nucleic Acids Res*, **20**, 1161.

- Mellentin, J.D., Smith, S.D. and Cleary, M.L. (1989). lyl-1, a novel gene altered by chromosomal translocation in T cell leukemia, codes for a protein with a helix-loop-helix DNA binding motif. *Cell*, **58**, 77-83.
- Melo, J.V., Gordon, D.E., Cross, N.C. and Goldman, J.M. (1993a). The ABL-BCR fusion gene is expressed in chronic myeloid leukemia. *Blood*, **81**, 158-65.
- Melo, J.V., Gordon, D.E., Tuszynski, A., Dhut, S., Young, B.D. and Goldman, J.M. (1993b). Expression of the ABL-BCR fusion gene in Philadelphia-positive acute lymphoblastic leukemia. *Blood*, **81**, 2488-91.
- Miller, J., McLachlan, A.D. and Klug, A. (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *Embo J*, **4**, 1609-14.
- Miller, M. and Zeller, K. (1997). Alternative splicing in lecithin:cholesterol acyltransferase mRNA: an evolutionary paradigm in humans and great apes. *Gene*, **190**, 309-13.
- Miwa, T., Sudo, K., Nakamura, Y. and Imai, T. (1993). Fifty sequenced-tagged sites on human chromosome 11. *Genomics*, **17**, 211-4.
- Mizuno, T., Kaibuchi, K., Yamamoto, T., Kawamura, M., Sakoda, T., Fujioka, H., Matsuura, Y. and Takai, Y. (1991). A stimulatory GDP/GTP exchange protein for smg p21 is active on the post-translationally processed form of c-Ki-ras p21 and rhoA p21. *Proc Natl Acad Sci U S A*, **88**, 6442-6.
- Monaco, A.P. (1994). Isolation of genes from cloned DNA. *Current Opinion in Genetics and Development*, **4**, 360-365.
- Moore, M.J., Query, C.C. and Sharp, P.A. (1993). Splicing of precursors to messenger RNAs by the spliceosome. In *The RNA World*, Atkins, R.G.J. (ed) pp. 303. Cold Spring Harbor Laboratory: New York.
- Moroianu, J., Hijikata, M., Blobel, G. and Radu, A. (1995). Mammalian karyopherin alpha 1 beta and alpha 2 beta heterodimers: alpha 1 or alpha 2 subunit binds nuclear

- localization signal and beta subunit interacts with peptide repeat-containing nucleoporins. *Proc Natl Acad Sci U S A*, **92**, 6532-6.
- Muragaki, Y., Mundlos, S., Upton, J. and Olsen, B.R. (1996). Altered growth and branching patterns in synpolydactyly caused by mutations in HOXD13. *Science*, **272**, 548-51.
- Nakajima, T., Uchida, C., Anderson, S.F., Lee, C.G., Hurwitz, J., Parvin, J.D. and Montminy, M. (1997). RNA helicase A mediates association of CBP with RNA polymerase II. *Cell*, **90**, 1107-12.
- Nakamura, T., Largaespada, D.A., Lee, M.P., Johnson, L.A., Ohyashiki, K., Toyama, K., Chen, S.J., Willman, C.L., Chen, I.M., Feinberg, A.P., Jenkins, N.A., Copeland, N.G. and Shaughnessy, J.D., Jr. (1996a). Fusion of the nucleoporin gene NUP98 to HOXA9 by the chromosome translocation t(7;11)(p15;p15) in human myeloid leukaemia. *Nat Genet*, **12**, 154-8.
- Nakamura, T., Largaespada, D.A., Shaughnessy, J.D., Jr., Jenkins, N.A. and Copeland, N.G. (1996b). Cooperative activation of Hoxa and Pbx1-related genes in murine myeloid leukaemias. *Nat Genet*, **12**, 149-53.
- Nakamura, T., Yamazaki, Y., Hatano, Y. and Miura, I. (1999). NUP98 is fused to PMX1 homeobox gene in human acute myelogenous leukemia with chromosome translocation t(1;11)(q23;p15). *Blood*, **94**, 741-7.
- Narumiya, S. (1996). The small GTPase Rho: cellular functions and signal transduction. *J Biochem (Tokyo)*, **120**, 215-28.
- Nishiyama, M., Arai, Y., Tsunematsu, Y., Kobayashi, H., Asami, K., Yabe, M., Kato, S., Oda, M., Eguchi, H., Ohki, M. and Kaneko, Y. (1999). 11p15 translocations involving the NUP98 gene in childhood therapy-related acute myeloid leukemia/myelodysplastic syndrome. *Genes Chromosomes Cancer*, **26**, 215-220.
- Nowell, P.C. and Hungerford, D.A. (1960). A minute chromosome in human chronic granulocytic leukemia. *Science*, **132**, 1497.

- Okuda, T., van Deursen, J., Hiebert, S.W., Grosveld, G. and Downing, J.R. (1996). AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. *Cell*, **84**, 321-30.
- Pak, B.J., Park, H., Chang, E.R., Pang, S.C. and Graham, C.H. (1998). Differential display analysis of oxygen-mediated changes in gene expression in first trimester human trophoblast cells. *Placenta*, **19**, 483-8.
- Pan, X. and Goldfarb, D.S. (1998). YEB3/VAC8 encodes a myristylated armadillo protein of the *Saccharomyces cerevisiae* vacuolar membrane that functions in vacuole fusion and inheritance. *J Cell Sci*, **111**, 2137-47.
- Papadopoulos, P., Ridge, S.A., Boucher, C.A., Stocking, C. and Wiedemann, L.M. (1995). The novel activation of ABL by fusion to an ets-related gene, TEL. *Cancer Res*, **55**, 34-8.
- Park, S.T., Nolan, G.P. and Sun, X.H. (1999). Growth inhibition and apoptosis due to restoration of E2A activity in T cell acute lymphoblastic leukemia cells. *J Exp Med*, **189**, 501-8.
- Pear, W.S., Miller, J.P., Xu, L., Pui, J.C., Soffer, B., Quackenbush, R.C., Pendergast, A.M., Bronson, R., Aster, J.C., Scott, M.L. and Baltimore, D. (1998). Efficient and rapid induction of a chronic myelogenous leukemia-like myeloproliferative disease in mice receiving P210 bcr/abl-transduced bone marrow. *Blood*, **92**, 3780-92.
- Peifer, M., Berg, S. and Reynolds, A.B. (1994). A repeating amino acid motif shared by proteins with diverse cellular roles. *Cell*, **76**, 789-91.
- Peifer, M. and Wieschaus, E. (1990). The segment polarity gene armadillo encodes a functionally modular protein that is the *Drosophila* homolog of human plakoglobin. *Cell*, **63**, 1167-76.



- Powers, M.A., Forbes, D.J., Dahlberg, J.E. and Lund, E. (1997). The vertebrate GLFG nucleoporin, Nup98, is an essential component of multiple RNA export pathways. *J Cell Biol*, **136**, 241-50.
- Powers, M.A., Macaulay, C., Masiarz, F.R. and Forbes, D.J. (1995). Reconstituted nuclei depleted of a vertebrate GLFG nuclear pore protein, p97, import but are defective in nuclear growth and replication. *J Cell Biol*, **128**, 721-36.
- Pritchard, C.E., Fornerod, M., Kasper, L.H. and van Deursen, J.M. (1999). RAE1 is a shuttling mRNA export factor that binds to a GLEBS-like NUP98 motif at the nuclear pore complex through multiple domains. *J Cell Biol*, **145**, 237-54.
- Pui, C.H. (1998). Recent advances in the biology and treatment of childhood acute lymphoblastic leukemia. *Curr Opin Hematol*, **5**, 292-301.
- Pui, C.H., Behm, F.G. and Crist, W.M. (1993). Clinical and biologic relevance of immunologic marker studies in childhood acute lymphoblastic leukemia. *Blood*, **82**, 343-62.
- Pui, C.H. and Evans, W.E. (1998). Acute lymphoblastic leukemia. *N Engl J Med*, **339**, 605-15.
- Pui, C.H., Frankel, L.S., Carroll, A.J., Raimondi, S.C., Shuster, J.J., Head, D.R., Crist, W.M., Land, V.J., Pullen, D.J., Steuber, C.P. and *et al.* (1991). Clinical characteristics and treatment outcome of childhood acute lymphoblastic leukemia with the t(4;11)(q21;q23): a collaborative study of 40 cases. *Blood*, **77**, 440-7.
- Rabbitts, T.H. (1994). Chromosomal translocations in human cancer. *Nature*, **372**, 143-9.
- Radu, A., Blobel, G. and Moore, M.S. (1995a). Identification of a protein complex that is required for nuclear protein import and mediates docking of import substrate to distinct nucleoporins. *Proc Natl Acad Sci U S A*, **92**, 1769-73.

- Radu, A., Moore, M.S. and Blobel, G. (1995b). The peptide repeat domain of nucleoporin Nup98 functions as a docking site in transport across the nuclear pore complex. *Cell*, **81**, 215-22.
- Raza-Egilmez, S.Z., Jani-Sait, S.N., Grossi, M., Higgins, M.J., Shows, T.B. and Aplan, P.D. (1998). NUP98-HOXD13 gene fusion in therapy-related acute myelogenous leukemia. *Cancer Res*, **58**, 4269-73.
- Redeker, E., Hoovers, J.M., Alders, M., van Moorsel, C.J., Ivens, A.C., Gregory, S., Kalikin, L., Bliet, J., de Galan, L., van den Bogaard, R. and *et al.* (1994). An integrated physical map of 210 markers assigned to the short arm of human chromosome 11. *Genomics*, **21**, 538-50.
- Riess, O., Epplen, C., Siedlaczek, I. and Epplen, J.T. (1993). Chromosomal assignment of the human smg GDP dissociation stimulator gene to human chromosome 4q21-q25. *Hum Genet*, **92**, 629-30.
- Rosenberg, U.B., Schroeder, C., Preiss, A., Kienlin, A., Cote, S., Riede, I. and Jaekle, H. (1996). Structural homology of the product of the *Drosophila* Kruppel gene with *Xenopus* transcription factor IIIA. *Nature*, **319**, 336-339.
- Rowley, J.D. (1973). Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature*, **243**, 290-3.
- Rowley, J.D. (1999). The role of chromosome translocations in leukemogenesis. *Semin Hematol*, **36**, 59-72.
- Rowley, J.D., Diaz, M.O., Espinosa, R.d., Patel, Y.D., van Melle, E., Ziemin, S., Taillon-Miller, P., Lichter, P., Evans, G.A., Kersey, J.H. and *et al.* (1990). Mapping chromosome band 11q23 in human acute leukemia with biotinylated probes: identification of 11q23 translocation breakpoints with a yeast artificial chromosome. *Proc Natl Acad Sci U S A*, **87**, 9358-62.

- Russell, N.H. (1997). Biology of acute leukaemia. *Lancet*, **349**, 118-22.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory: New York.
- Sanchez-Garcia, I. (1997). Consequences of chromosomal abnormalities in tumor development. *Annu Rev Genet*, **31**, 429-53.
- Savitsky, K., Ziv, Y., Bar-Shira, A., Gilad, S., Tagle, D.A., Smith, S., Uziel, T., Sfez, S., Nahmias, J., Sartiel, A., Eddy, R.L., Shows, T.B., Collins, F.S., Shiloh, Y. and Rotman, G. (1996). A human gene (DDX10) encoding a putative DEAD-box RNA helicase at 11q22- q23. *Genomics*, **33**, 199-206.
- Sawyers, C.L. (1997). Molecular genetics of acute leukaemia. *Lancet*, **349**, 196-200.
- Sawyers, C.L., Denny, C.T. and Witte, O.N. (1991). Leukemia and the disruption of normal hematopoiesis. *Cell*, **64**, 337-50.
- Schroeder, W.T., Chao, L.Y., Dao, D.D., Strong, L.C., Pathak, S., Riccardi, V., Lewis, W.H. and Saunders, G.F. (1987). Nonrandom loss of maternal chromosome 11 alleles in Wilms tumors. *Am J Hum Genet*, **40**, 413-420.
- Schwaller, J., Frantsve, J., Aster, J., Williams, I.R., Tomasson, M.H., Ross, T.S., Peeters, P., Van Rompaey, L., Van Etten, R.A., Ilaria, R., Jr., Marynen, P. and Gilliland, D.G. (1998). Transformation of hematopoietic cell lines to growth-factor independence and induction of a fatal myelo- and lymphoproliferative disease in mice by retrovirally transduced TEL/JAK2 fusion genes. *Embo J*, **17**, 5321-33.
- Shaul, Y. (2000). c-abl: activation and nuclear targets. *Cell Death Differ*, **7**, 10-6.
- Shima, E.A., Le Beau, M.M., McKeithan, T.W., Minowada, J., Showe, L.C., Mak, T.W., Minden, M.D., Rowley, J.D. and Diaz, M.O. (1986). Gene encoding the alpha chain of the T-cell receptor is moved immediately downstream of c-myc in a chromosomal 8;14 translocation in a cell line from a human T-cell leukemia. *Proc Natl Acad Sci U S A*, **83**, 3439-43.

- Shimizu, K., Kawabe, H., Minami, S., Honda, T., Takaishi, K., Shirataki, H. and Takai, Y. (1996). SMAP, an Smg GDS-associating protein having arm repeats and phosphorylated by Src tyrosine kinase. *J Biol Chem*, **271**, 27013-7.
- Shimizu, K., Shirataki, H., Honda, T., Minami, S. and Takai, Y. (1998). Complex formation of SMAP/KAP3, a KIF3A/B ATPase motor-associated protein, with a human chromosome-associated polypeptide. *J Biol Chem*, **273**, 6591-4.
- Shows, T.B., Alders, M., Bennett, S., Burbee, D., Cartwright, P., Chandrasekharappa, S., Cooper, P., Courseaux, A., Davies, C., Devignes, M.D., Devilee, P., Elliott, R., Evans, G., Fantes, J., Garner, H., Gaudray, P., Gerhard, D.S., Gessler, M., Higgins, M., Hummerich, H., James, M., Lagercrantz, J., Litt, M., Little, P., Zabel, B. and *et al.* (1996). Report of the Fifth International Workshop on Human Chromosome 11 Mapping 1996. *Cytogenet Cell Genet*, **74**, 1-56.
- Slany, R.K., Lavau, C. and Cleary, M.L. (1998). The oncogenic capacity of HRX-ENL requires the transcriptional transactivation activity of ENL and the DNA binding motifs of HRX. *Mol Cell Biol*, **18**, 122-9.
- Song, W.J., Sullivan, M.G., Legare, R.D., Hutchings, S., Tan, X., Kufrin, D., Ratajczak, J., Resende, I.C., Haworth, C., Hock, R., Loh, M., Felix, C., Roy, D.C., Busque, L., Kurnit, D., Willman, C., Gewirtz, A.M., Speck, N.A., Bushweller, J.H., Li, F.P., Gardiner, K., Poncz, M., Maris, J.M. and Gilliland, D.G. (1999). Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. *Nat Genet*, **23**, 166-75.
- Strassheim, D., Porter, R.A., Phelps, S.H. and Williams, C.L. (2000). Unique in vivo associations with SmgGDS and RhoGDI and different guanine nucleotide exchange activities exhibited by RhoA, dominant negative RhoA(Asn-19), and activated RhoA(Val-14). *J Biol Chem*, **275**, 6699-702.
- Sverdlov, E.D. (2000). Retroviruses and primate evolution. *Bioessays*, **22**, 161-171.

- Taagepera, S., McDonald, D., Loeb, J.E., Whitaker, L.L., McElroy, A.K., Wang, J.Y. and Hope, T.J. (1998). Nuclear-cytoplasmic shuttling of C-ABL tyrosine kinase. *Proc Natl Acad Sci U S A*, **95**, 7457-62.
- Takakura, A., Miyoshi, J., Ishizaki, H., Tanaka, M., Togawa, A., Nishizawa, Y., Yoshida, H., Nishikawa, S. and Takai, Y. (2000). Involvement of a small GTP-binding protein (G protein) regulator, small G protein GDP dissociation stimulator, in antiapoptotic cell survival signaling. *Mol Biol Cell*, **11**, 1875-86.
- Taub, R., Kirsch, I., Morton, C., Lenoir, G., Swan, D., Tronick, S., Aaronson, S. and Leder, P. (1982). Translocation of the c-myc gene into the immunoglobulin heavy chain locus in human Burkitt lymphoma and murine plasmacytoma cells. *Proc Natl Acad Sci U S A*, **79**, 7837-41.
- Tycko, B., Smith, S.D. and Sklar, J. (1991). Chromosomal translocations joining LCK and TCRB loci in human T cell leukemia. *J Exp Med*, **174**, 867-73.
- Uckun, F.M., Sensel, M.G., Sun, L., Steinherz, P.G., Trigg, M.E., Heerema, N.A., Sather, H.N., Reaman, G.H. and Gaynon, P.S. (1998). Biology and treatment of childhood T-lineage acute lymphoblastic leukemia. *Blood*, **91**, 735-46.
- Van Etten, R.A., Jackson, P. and Baltimore, D. (1989). The mouse type IV c-abl gene product is a nuclear protein, and activation of transforming ability is associated with cytoplasmic localization. *Cell*, **58**, 669-78.
- Vithalani, K.K., Parent, C.A., Thorn, E.M., Penn, M., Larochelle, D.A., Devreotes, P.N. and De Lozanne, A. (1998). Identification of darlin, a Dictyostelium protein with Armadillo-like repeats that binds to small GTPases and is important for the proper aggregation of developing cells. *Mol Biol Cell*, **9**, 3095-106.
- von Lindern, M., Fornerod, M., van Baal, S., Jaegle, M., de Wit, T., Buijs, A. and Grosveld, G. (1992a). The translocation (6;9), associated with a specific subtype of acute

- myeloid leukemia, results in the fusion of two genes, *dek* and *can*, and the expression of a chimeric, leukemia-specific *dek-can* mRNA. *Mol Cell Biol*, **12**, 1687-97.
- von Lindern, M., van Baal, S., Wiegant, J., Raap, A., Hagemeijer, A. and Grosveld, G. (1992b). *Can*, a putative oncogene associated with myeloid leukemogenesis, may be activated by fusion of its 3' half to different genes: characterization of the *set* gene. *Mol Cell Biol*, **12**, 3346-55.
- Voss, J., Posem, G., Hannemann, J.R., Wiedemann, L.M., Turhan, A.G., Poirel, H., Bernard, O.A., Adermann, K., Kardinal, C. and Feller, S.M. (2000). The leukaemic oncoproteins Bcr-Abl and Tel-Abl (ETV6/Abl) have altered substrate preferences and activate similar intracellular signalling pathways. *Oncogene*, **19**, 1684-90.
- Waltzer, L. and Bienz, M. (1999). The control of beta-catenin and TCF during embryonic development and cancer. *Cancer Metastasis Rev*, **18**, 231-46.
- Wang, J., Jani-Sait, S.N., Escalon, E.A., Carroll, A.J., de Jong, P.J., Kirsch, I.R. and Aplan, P.D. (2000). The t(14;21)(q11.2;q22) chromosomal translocation associated with T-cell acute lymphoblastic leukemia activates the BHLHB1 gene. *Proc Natl Acad Sci U S A*, **97**, 3497-502.
- Wang, J.C. (1996). DNA topoisomerases. *Annu Rev Biochem*, **65**, 635-92.
- Ward, W.W., Cody, C.W., Hart, R.C. and Cormier, M.J. (1980). Spectrophotometric identity of the energy transfer chromophores in *Renilla* and *Aequorea* green-fluorescent proteins. *Photochemistry and Photobiology*, **31**, 611-615.
- Wente, S.R., Rout, M.P. and Blobel, G. (1992). A new family of yeast nuclear pore complex proteins. *J Cell Biol*, **119**, 705-23.
- Wildin, R.S., Garvin, A.M., Pawar, S., Lewis, D.B., Abraham, K.M., Forbush, K.A., Ziegler, S.F., Allen, J.M. and Perlmutter, R.M. (1991). Developmental regulation of *lck* gene expression in T lymphocytes. *J Exp Med*, **173**, 383-93.

- Wimmer, C., Doye, V., Grandi, P., Nehrbass, U. and Hurt, E.C. (1992). A new subclass of nucleoporins that functionally interact with nuclear pore protein NSP1. *Embo J*, **11**, 5051-61.
- Witzgall, R., O'Leary, E., Leaf, A., Onaldi, D. and Bonventre, J.V. (1994). The Kruppel-associated box-A (KRAB-A) domain of zinc finger proteins mediates transcriptional repression. *Proc Natl Acad Sci U S A*, **91**, 4514-8.
- Wong, K.F., So, C.C. and Kwong, Y.L. (1999). Chronic myelomonocytic leukemia with t(7;11)(p15;p15) and NUP98/HOXA9 fusion. *Cancer Genet Cytogenet*, **115**, 70-2.
- Yamamoto, T., Kaibuchi, K., Mizuno, T., Hiroyoshi, M., Shirataki, H. and Takai, Y. (1990). Purification and characterization from bovine brain cytosol of proteins that regulate the GDP/GTP exchange reaction of smg p21s, ras p21-like GTP-binding proteins. *J Biol Chem*, **265**, 16626-34.
- Yamazaki, H., Nakata, T., Okada, Y. and Hirokawa, N. (1996). Cloning and characterization of KAP3: a novel kinesin superfamily-associated protein of KIF3A/3B. *Proc Natl Acad Sci U S A*, **93**, 8443-8.
- Yang, T.T., Cheng, L. and Kain, S.R. (1996). Optimized codon usage and chromophore mutations provide enhanced sensitivity with the green fluorescent protein. *Nucleic Acids Res*, **24**, 4592-3.
- Yano, T., Nakamura, T., Blechman, J., Sorio, C., Dang, C.V., Geiger, B. and Canaani, E. (1997). Nuclear punctate distribution of ALL-1 is conferred by distinct elements at the N terminus of the protein. *Proc Natl Acad Sci U S A*, **94**, 7286-91.
- Ye, B.H., Lista, F., Lo Coco, F., Knowles, D.M., Offit, K., Chaganti, R.S. and Dalla-Favera, R. (1993). Alterations of a zinc finger-encoding gene, BCL-6, in diffuse large-cell lymphoma. *Science*, **262**, 747-50.
- Zolotukhin, A.S. and Felber, B.K. (1999). Nucleoporins nup98 and nup214 participate in nuclear export of human immunodeficiency virus type 1 Rev. *J Virol*, **73**, 120-7.

**Appendix**  
**Journal Papers**



D.J. Hussey, N.J. Parker, N.D. Hussey, P.F.R. Little and A. Dobrovic (1997)  
Characterization of a KRAB Family Zinc Finger Gene, ZNF195, Mapping to  
Chromosome Band 11p15.5.  
*Genomics*, v. 45 (2), pp. 451–455, October 1997

NOTE: This publication is included in the print copy of the thesis  
held in the University of Adelaide Library.

It is also available online to authorised users at:

<http://dx.doi.org/10.1006/geno.1997.4958>

D.J. Hussey, M. Nicola, S. Moore, G.B. Peters and A. Dobrovic (1999) The (4;11) (q21;p15) Translocation Fuses the NUP98 andRAP1GDS1 Genes and Is Recurrent in T-Cell Acute Lymphocytic Leukemia.  
*Blood*, v. 94 (6), pp. 2072-2079, September 1999

NOTE: This publication is included in the print copy of the thesis held in the University of Adelaide Library.

## ERRATA

Table 4.2: For patient 5, “T” should read “T-cell type, individual surface markers not reported”.

p103 section 4.5: “...referred to patients 2 and 3...” should read “referred to as patients 2 and 3...”.

Fig 5.3 Lane 5: This lane contains DNA target as a PCR control. No band is observed and this is a confirmation that the PCR is only amplifying reverse transcribed cDNA.