MOLECULAR CHARACTERISATION OF TRANSLOCATIONS INVOLVING CHROMOSOME BAND 1p36 IN ACUTE MYELOID LEUKAEMIA

Christopher Slape, B.Sc. (Hons)

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

to

The University of Adelaide

Department of Medicine

(The Queen Elizabeth Hospital)

October 2002

Table of Contents

Chapter 3

List of Tables

List of Figures

ABSTRACT

This thesis describes the mapping of the breakpoints of three different chromosome rearrangements, all involving 1p36, in acute myeloid leukaemia (AML) patients, and an investigation into the molecular outcomes of these rearrangements.

The t(1;3)(p36;q21) is the most frequently occurring translocation involving 1p36 in AML. The breakpoints in a patient with this translocation were mapped and the translocation shown to result in the upregulation of the *MEL1* gene, consistent with other reports. A panel of non t(1;3)(p36;q21) AML patients were also tested for *MEL1* expression, and a subset of these were found to be positive for expression. *MEL1* had previously been reported to be expressed specifically in patients with the t(1;3)(p36;q21). Contrary to previous reports, *MEL1* was found to be expressed in normal bone marrow. These results suggest that the expression of *MEL1* in non-t(1;3)(p36;q21) patients may be due to the clonal expansion of a normally-expressing haematopoietic progenitor. The 3q21 gene *GATA2* was also found to be aberrantly upregulated by the t(1;3)(p36;q21), as well as the t(3;3)(q21;q26) and the inv(3)(q21q26). These three translocations define a subset of AML patients with distinct features, namely elevated platelet counts and dysmegakaryopoiesis. *GATA2* expression is proposed as a likely cause of this phenotype.

The ins(12;1)(p13;p36p21) is a novel rearrangement. Investigation of this rearrangement in Patient 2 revealed the 1p36 breakpoint was similar to those of the t(1;3)(p36;q21) patients. The 12p13 breakpoint was within the *TEL* gene, which is frequently involved in fusion genes in leukaemia. No fusion transcript was identified, and the primary outcome of the rearrangement was determined to be upregulation of *MEL1*, most likely caused by position effects of the juxtaposed 12p13 sequences.

The t(1;3)(p36;p21) is a recurrent translocation in several types of leukaemia. Investigation of this translocation in Patient 3 mapped the 3p21 breakpoint to an intron of *GOLGA4*, which has not previously been implicated in leukaemia. Northern analysis strongly suggested the formation of a fusion transcript. Possible leukemogenic mechanisms by which this fusion may function are proposed.

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, being available for loan and photocopying.

Signed:

Date: 14/01/03

ACKNOWLEDGEMENTS

I am very grateful to my principal supervisor, Dr Alex Dobrovic, who provided me with an unexpected opportunity to undertake a Ph.D., and who went out of his way to find a project that suited my liking and abilities. I thank him for his guidance and especially for his approachable and easy-going approach to supervision, which allowed me to develop the project over time.

I am also in debt to Dr Damian Hussey for his supervision, his excellent knowledge of all laboratory techniques, and for his expert guidance in the analysis and understanding of chromosome translocations. His example inspired me on many occasions and is one I still aspire to live up to.

I am very grateful to Dr Sally Stephenson for her supervision and support. She was always willing and able to help with any problem, and she encouraged me many times when I had reason to despair and always got me back on track.

I would also like to thank my external supervisor, Dr Simon Koblar, for providing moral support, especially early in the project, and for always being available on the rare occasions I felt the need to go outside the department.

I gratefully acknowledge the assistance of Dr Ian Lewis and Dr David Westerman, who read and corrected drafts of this thesis.

Thanks also go to the staff of the Department of Haematology-Oncology at The Queen Elizabeth Hospital. I would especially like to thank Professor Ed Sage for fostering research within the department throughout his long tenure. I wish him well in his retirement.

I would like to acknowledge the financial support provided by The Queen Elizabeth Hospital Research Foundation and The Cancer and Bowel Research Trust. Without this support, I would not have been able to complete this project, and I am extremely grateful.

Sarah Moore from the Department of Haematology at the IMVS was of great assistance in performing and analysing karyotypes, and also in providing patient material and background information upon request. Her contribution to this work is greatly appreciated.

Members of other labs within the Queen Elizabeth Hospital were generous with their resources and expertise, and I would like to thank them also: Jenny Hardingham, Prue Cowled, Henry Betts, Ravi Krishnan, Nicole Hussey, Anke Warner, Belinda Farmer, Lefta Leonardos, Sue Millard, Terry Gooley and Jenny Kennedy. Thanks also to Jing Xian Mi, Adrian Brenton and Logi Bauer for their excellent technical assistance.

I appreciate the support and friendship of the other research students at the Queen Elizabeth Hospital. I have made many friends over my years here, and this is surely an important component of any studentship. They are too many to list but I hope to remain in touch with them all.

I am extremely grateful to Tanya Sanders, who has helped me in more different ways through my project than anyone. She provided great support during some difficult times and was always excellent company during the more tedious experiments. She helped me through the difficult experiment-finalising and thesis-writing period, and provided the same emotional and additional culinary support during this time. I wish her well in her own Ph.D. and for all her endeavours in life, and hope to remain her friend throughout.

(Dr) Tina Bianco, who suffered through the thesis write-up period almost concurrently with me, is owed an enormous debt of gratitude for allegedly suggesting to Alex that I was a worthy Ph.D. student and so landing me in the position I find myself. Our friendship extends back to our undergraduate days, and she has remained an excellent friend, lab companion and fellow student. She guided me through my first inept days in the laboratory, and for this and much more she has my eternal gratitude. She still owes me a lasagne, though.

I am indebted to Nathaniel Albanese, who shared an office with me for two years (and a radio show for one), and who provided great friendship and support over the first part of my project. His friendly and laid-back manner made work more fun and with this attitude to life I have no doubt that he will be successful in everything he does. I wish him nothing less.

I thank Michael Raynor for sharing an office with me for almost the duration of his own Ph.D. Among my many bad habits that he has never commented on is that of monopolising the computer, especially during the final six months. This tolerance was noted and appreciated, and I hope the benefit of having the office to himself now will allow him to complete his Ph.D. in a relatively pain-free manner.

Thanks also to Evelyn Douglas for sharing an office with me for as long as she could bear (about six months), for being a good and supportive friend and such a deft hand with glue and scissors. I wish her well in completing her Ph.D. also, and in all things to follow.

Cassandra Woithe and Rebecca Gully have been similarly good friends and fun to have around the lab. I wish them both well in their endeavours and marriages.

Thanks to my parents, who have provided me not only with love, support and guidance, but also my education. This thesis was only possible because of them. I also appreciate the love and support of the rest of my family. They have all assisted me at every opportunity throughout this project, and before (and hopefully beyond).

My friends, who have provided me with amusement, distraction and fun, as well as support and friendship, know who they are and know how much they mean to me. I apologise to them all for being relatively unavailable for about three years, and promise to be less so in the future. Really.

Finally, the biggest thanks goes to Ai, not only for her love and support during the work presented here, but also for being an unexpectedly good motivator. She taught me that hard work can pay off, and without that, much of the work presented here would not have been performed. I owe this thesis to her as much as anyone. I cannot thank her enough, for that and everything else.

ABBREVIATIONS

AML	-	acute myeloid leukemia
ALL	-	acute lymphoid leukemia
BAC	-	bacterial artificial chromosome
BLAST	-	basic local alignment search tool
bp	-	base pair(s)
BSA	-	bovine serum albumin
CELU	-	cohesive end ligation unit
CLL	-	chronic lymphoid leukemia
CML	-	chronic myeloid leukemia
dATP	-	deoxyadenosine triphosphate
dCTP	-	deoxycytosine triphosphate
dGTP	-	deoxyguanine triphosphate
DIG	-	Digoxigenin
DNA	-	deoxyribonucleic acid
dNTP	-	deoxynucleoside triphosphate
dTTP	-	deoxythymine triphosphate
EDTA	-	ethylene diaminetetraacetic acid
EST	-	expressed sequence tag
EtBr	-	ethidium bromide
FAB	-	French - American - British
FCS	-	fetal calf serum
FISH	-	fluorescent in situ hybridisation
IMVS	-	Institue of Medical and Vetinary Science (Adelaide)
ISH	-	*in situ* hybridisation
kb	-	kilobases (10^3 base pairs)
Mb	-	megabases (10^6 base pairs)
MDS	-	myelodysplastic syndrome
MMLV	-	Moloney Murine Leukaemia Virus
NCBI	-	National Centre for Biotechnology Information
nt	-	nucleotide(s)
PBMNC	-	peripheral blood mononuclear cells
PBS	-	phosphate buffered saline
PCR	-	polymerase chain reaction
RACE	-	rapid amplification of cDNA ends
RNA	-	ribonucleic acid
RT-PCR	-	reverse transcription PCR
UPW	-	ultrapure water
UV	-	ultraviolet
WHO	-	World Health Organisation

SI (Systeme International) units and the international code for DNA bases are used throughout this thesis.

CHAPTER 1

Introduction

and

Review of Literature

CHAPTER 1

Introduction and Review of Literature

1.0 Overview

The aim of the work described in this thesis was to characterise the chromosomal translocation breakpoints in three acute myeloid leukaemia (AML) patients with three different rearrangements involving chromosome band 1p36: t(1;3)(p36;q21), ins(12;1)(p13;p36p21) and t(1;3)(p36;p21).

Past studies characterising translocation breakpoints have led to the discovery of many genes important in leukaemia, as well as genes with important functions in normal processes such as haematopoiesis and development (Wang *et al* 1998, Okuda *et al* 2001, Crans and Sakamoto 2001). These studies have proven that translocations are important mechanisms in the development of leukaemia, as many events caused by chromosome translocation are leukemogenic when reconstructed in cell lines or transgenic mice (for example Heisterkamp *et al* 1990, Pandolfi 2001, van Etten 2001, Yuan *et al* 2001). Moreover, many translocations are highly correlated with particular subtypes of leukaemia (reviewed in Barr 1998, reviewed in Rowley 1999).

The importance of studying translocations is two-fold. Firstly, mapping individual translocation breakpoints may identify new genes important in the promotion of leukemogenesis (or indicate novel leukemogenic functions for known genes). As translocation breakpoints act as markers for the location of these genes, translocation-affected genes are among the easiest disease genes to identify. Secondly, greater understanding of the role of translocations in leukaemia leads to a better understanding of genetic leukemogenic mechanisms in general. Better understanding of the causes of the disease may allow more effective treatments to be developed and implemented.

This chapter provides an introduction to leukaemia (Section 1.1) and chromosome translocations (Section 1.2). It demonstrates the value of mapping translocation breakpoints and studying the genes affected by the breakpoints to determine their contribution to leukemogenesis.

1.1 Leukaemia

Leukaemia is a disease of the haematopoietic system, arising from the disruption of the regulation of haematopoiesis. It is characterised by the uncontrolled proliferation or expansion of haematopoietic cells that do not retain the capacity to differentiate normally to mature blood cells (Sawyers *et al* 1991).

1.1.1 Prevalence of Leukaemia

Leukaemia is the seventh most common cancer in South Australia, after prostate cancer, breast cancer, colorectal cancer, lung cancer, melanoma and lymphoma. There were 4510 cases of leukaemia diagnosed in South Australia between January 1977 and December 2000, representing 3.4 % of all cancer diagnoses. Most diagnoses occur in the over 50 years age group, with the majority of the paediatric cases being acute lymphoid leukaemias (ALL). Acute myeloid leukaemias (AML), the primary focus of this thesis, represent 1.0 % of all cancers and 30.8 % of all leukaemias. Rates of occurrence of leukaemia are slightly higher in males than in females, but this is reflective of an overall trend for most types of cancer (South Australian Cancer Registry, 2001).

1.1.2 Haematopoiesis

Haematopoiesis is the means by which new blood cells are produced. The process originates with pluripotent stem cells which have the capacity to give rise to all of the different types of specialised blood cells. Haematopoietic stem cells possess the crucial capacity to divide either symmetrically (stem cell self-renewal) or asymmetrically, such that one daughter cell becomes a stem cell which retains the pluripotency of the parental cell, while the other daughter cell differentiates to one of several different fates. The differentiation pathways leading to these fates are illustrated in Figure 1.1.

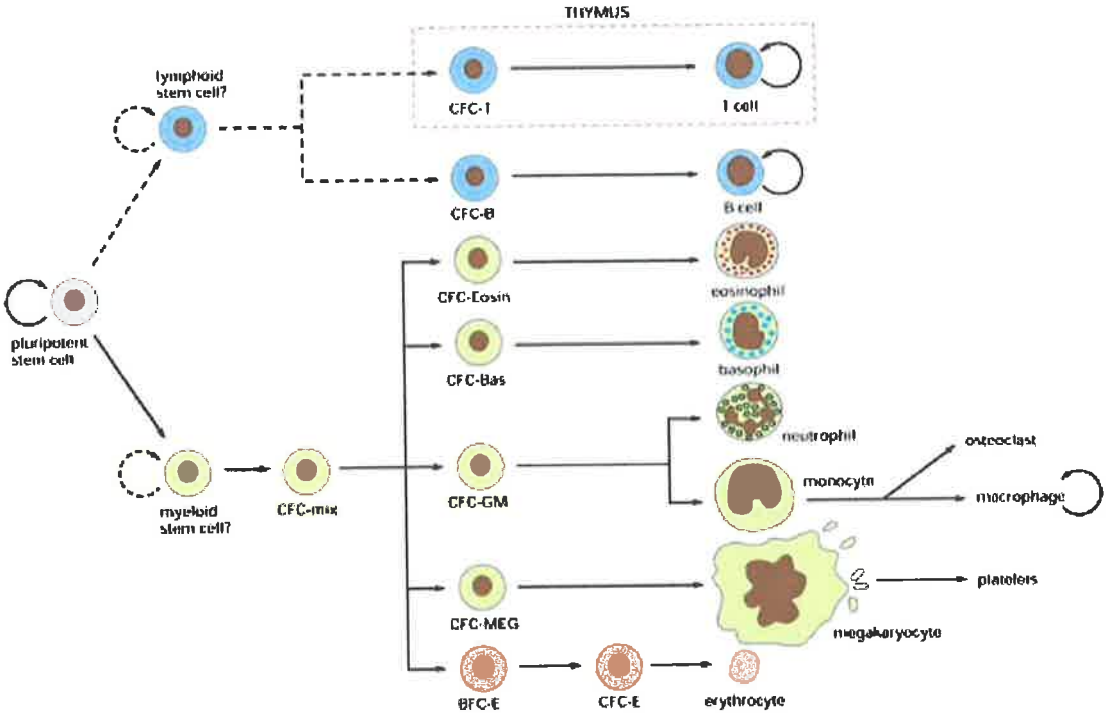


Figure 1.1. A Scheme of Haematopoiesis. The pluripotent stem cell divides to generate either more pluripotent stem cells or committed progenitor cells (CFCs - colony-forming cells). Progenitor cells proliferate and differentiate to eventually develop into terminally differentiated blood cells. T cells are formed in the thymus; all other cell types are produced in the bone marrow. Dashed lines reflect uncertainty over the nature of the pathway of differentiation. Figure reproduced from Alberts *et al* (1994).

1.1.3 Classification of the Leukaemias

1.1.3.1 Types of Leukaemia

Leukaemia is a neoplasm derived from haematopoietic cells. The leukaemias are classified broadly according to the apparent cell of origin (that is, lymphoid or myeloid) and also according to the rapidity of the clinical course (that is, acute or

chronic). The diagnosis of the type of leukaemia is now based on a combination of morphologic, immunophenotypic, cytochemical and cytogenetic characterisation using the World Health Organisation (WHO) Classification. This is a relatively new proposal for classification of haematological malignancies, formulated on behalf of the WHO by the European Association of Pathologists and the Society for Hematopathology, and was formally adopted in 2000 (Bennett 2000). This classification system places greater emphasis on immunological and cytogenetic characteristics of the disease than did the French - American - British (FAB) system which preceded it (discussed in Section 1.1.3.2).

The translocations studied in this thesis all arose in patients with AML. Leukemic cells in AML are referred to as blasts, as they resemble undifferentiated cells. The definition of acute leukaemia requires that there be both a loss of differentiation capacity and a loss of proliferative control (Sawyers *et al* 1991). In addition to the AML patients investigated, some myelodysplastic syndrome (MDS) patients were also studied. MDS is a clonal disorder of the myeloid lineage in which affected cells have a block preventing normal differentiation, leading to morphologic abnormalities and an accumulation of haematopoietic progenitor cells. Under the FAB system, if blasts must be present at less than 30 % in the marrow for diagnosis as an MDS, as a higher blast count would lead to a diagnosis of leukaemia. Under the new WHO system, this threshold has been lowered to 20 % (Bennett 2000). In approximately 20 % of cases, MDS can transform into AML as the affected cells undergo additional changes.

1.1.3.2 Subtypes of AML

While the WHO classification system has been widely adopted for the classification of leukaemias, the FAB system is discussed here for two reasons. Firstly, the WHO classification categories used to describe subtypes of AML are based on the FAB categories. Secondly, the patients studied in this thesis presented prior to the adoption of the WHO classification system and were diagnosed under the FAB system.

The FAB classification system for AML was first devised by a consortium of haematologists in 1976 (Bennett *et al* 1976) and comprehensively revised in 1985 (Bennett *et al* 1985a). The system at first defined six AML subtypes (M1 - M6), later adding M7 (Bennett *et al* 1985b) and M0 (Bennett *et al* 1991) (see Table 1.1). In a broad sense, the subtypes are based on morphologic appearance of the blasts and reflect the degree of maturation and lineage commitment of the leukemic blasts, with M0 being derived from undifferentiated cells, M1 being AML derived from haematopoietic cells without maturation and M2 from cells showing partial maturation. AML M3 through M7 are classified according to the lineage of the haematopoietic cells from which the leukaemia derived.

Classification by the FAB system requires morphological examination and cytochemical staining of the leukemic blast cells. The system takes into account cell size and shape, nucleus size and shape, number of nucleoli present and the staining pattern of cytoplasmic granules (reviewed in Bennett *et al* 1985a). A number of cytochemical stains are useful in the diagnosis and classification of AML subtypes.

AML FAB Subtype	Diagnostic Criteria
AML M0 Minimally differentiated acute myeloid leukemia	Blast cells ≥ 30 % of non-erythroid cells (NEC). ≤ 3 % of blasts positive for myeloperoxidase. Confirmation of myeloid lineage by immunophenotyping and electron microscopy.
AML M1 Acute myeloblastic leukemia without maturation	Blast cells ≥ 90 % of NEC. ≥ 3 % of blasts positive for myeloperoxidase or sudan black staining. Maturing granulocytic component ≤ 10 %. Maturing monocytic component ≤ 10 %.
AML M2 Acute myeloblastic leukemia with maturation	Blast cells between 30 % and 89 % of NEC. Monocytic cells ≤ 20 %. Maturing granulocytic cells (promyelocytes) ≥ 10 %.
AML M3 Acute promyelocytic leukemia	Blast cells ≥ 30 % of NEC. Majority of cells are abnormal promyelocytes. Heavy granulation of promyelocytes is prominent.
AML M4 Acute myelomonocytic leukemia	Blast cells ≥ 30 % of NEC in bone marrow. 30 % - 80 % of NEC in bone marrow are of the myelomonocytic lineage. 20 % - 80 % of NEC in bone marrow are of the monocytic lineage. M4 diagnosis also requires substantial presence of monocytic cells in the peripheral blood, either by cell count or other method.
AML M5 Acute monocytic leukemia	Blast cells ≥ 30 % of NEC. ≥ 80 % of NEC in bone marrow are of the monocytic lineage. M5a: ≥ 80 % of monocytic cells are monoblasts. M5b: ≤ 80 % of monocytic cells are monoblasts.
AML M6 Acute erythroleukemia	Blast cells ≥ 30 % of NEC. Erythroblasts ≥ 50 % of nucleated cells.
AML M7 Acute megakaryocytic leukemia	Blast cells ≥ 30 % of all cells. Megakaryocytic lineage demonstrated by the platelet peroxidase reaction. ≤ 3 % blasts positive for myeloperoxidase or Sudan Black. Negative for lymphoid markers.

Table 1.1. The FAB Subtypes of Acute Myeloid Leukemia. Diagnostic criteria compiled from Bennet *et al* (1976), Bennet *et al* (1985a), Bennett *et al* (1985b) and Bennett *et al* (1991).

For example, myeloperoxidase and Sudan Black positivity is primarily indicative of granulocytic lineage cells. Nonspecific esterase (α-naphthyl butyrate esterase) stains specifically for cells of the monocytic lineage (reviewed in Elghetany *et al* 1990).

Leukaemia is a disease of acquired genetic dysregulation, so it would appear desirable to classify the different types of leukaemia on a genetic basis rather than a phenotypic basis. This would allow diagnosis to be more closely related to the genetic cause of the leukaemia than to the morphological outcome of the disease. The development of the WHO classification system reflects a desire among clinicians for this to occur. Certain cytogenetically detectable chromosomal rearrangements are correlated with a particular clinical AML subtype, and in many cases this is a useful means of confirming a subtype diagnosis. For example, the t(15;17)(q22;q12) is present in virtually all cases of AML M3, and is never present in other AML subtypes (Pandolfi 2001). However, the genetic classification of leukaemia will progress to become much more sophisticated than the employment of cytogenetic data alone. Recently, microarray technology has been used to correctly categorise leukaemias as either ALL or AML purely on the basis of the gene expression profile of the leukaemia (Golub *et al* 1999). It has also been used to correctly categorise AML subtypes which correlate with a cytogenetic marker (Schoch *et al* 2002), and even to identify mixed lineage leukaemia (MLL) involving an 11q23 translocation (affecting the *MLL* gene) as a disease genetically distinct from both ALL and AML (Armstrong *et al* 2002). The power of microarray technology may eventually lead to more robust classification and more individualised therapy, but the clinical application of this technology is some years away (reviewed in Radich 2002).

1.2 Chromosome Translocations

1.2.1 Translocations in Leukaemia and Other Malignancies

Cancer is a disease of acquired genetic dysregulation, meaning that it arises out of disruption of normal genetic function. This disruption can come in the form of an epigenetic mechanism, such as methylation of a gene promoter region which affects expression of a gene, or a genetic mechanism, which involves changes to the actual DNA sequence of the genome. The most gross form of mutation is chromosomal rearrangement, where the normal structure of the chromosomes is altered. These rearrangements come in the form of deletions, amplifications, inversions and translocations, and are often discernible cytogenetically. Figure 1.2 is a schematic representation of each of these types of rearrangements. Translocations are the reciprocal exchange of material between two chromosomes. Inversions occur when chromosomal material is, in effect, removed from the chromosome, rotated and reinserted in the original position. Inversions and translocations have similar molecular outcomes; either existing genes are expressed aberrantly due to the new genomic context in which they are placed, or two existing genes are combined to form a fusion gene which retains some properties of both original genes which, in combination, are usually oncogenic (see Section 1.2.2). A deletion is the loss of chromosomal material, and generally indicates the loss of a tumour suppressor gene, although there is at least one example of a deletion producing a fusion gene (Kourlas *et al* 2000).

Figure 1.2. Schematic Representation of the Various Types of Chromosomal Rearrangements.

A) Reciprocal Translocation. The parental chromosomes are represented in yellow and purple respectively. Each parental chromosome breaks and there is reciprocal exchange of material between the two.

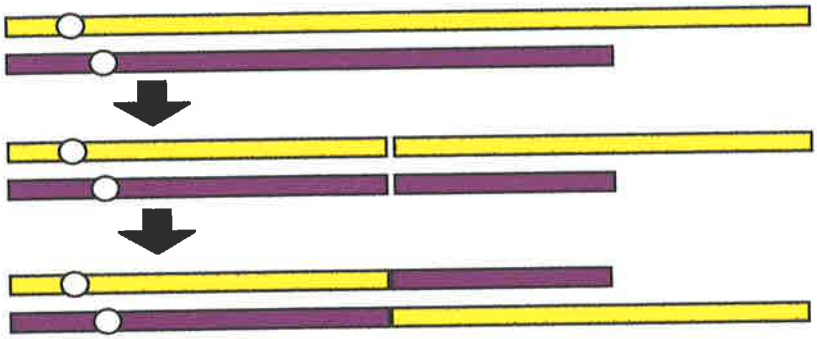
B) Inversion. Genetic markers A, B, C and D from the parental chromosome are shown to illustrate their movement as a result of the rearrangement. Two breaks occur within the parental chromosome and the centre piece of the chromosome is inverted before reinsertion. Some inversions involve the centromere (pericentric inversions); this one does not (paracentric inversion).

C) Deletion. Genetic markers A, B, C, D and E from the parental chromosome are shown to illustrate their loss as a result of the deletion. In this case, the deletion spans the markers C and D. Deletions can have an oncogenic effect through the loss of expression of genes which are deleted. However, only one allele of each gene is lost in the deletion.

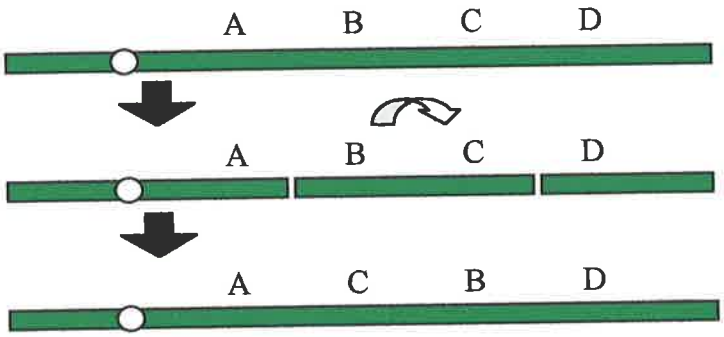
D) Amplification. Genetic markers A, B, C, D and E from the parental chromosome are shown to illustrate the effect of the amplification on them. In this case, the portion of chromosome contining marker C has been duplicated, resulting in the presence of two copies of the marker on the chromosome. This can lead to overexpression of amplified genes.

Figure is adapted from Griffiths *et al* (2000).

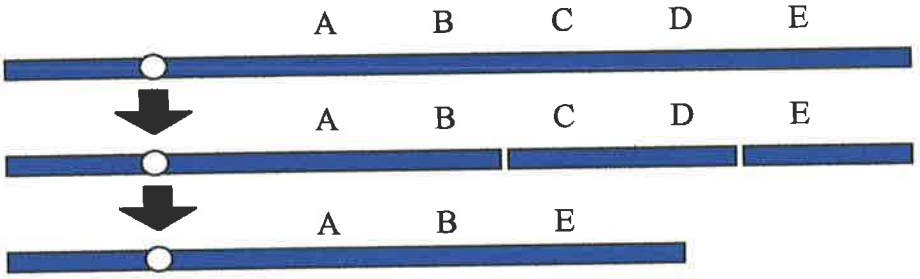
A) Translocation



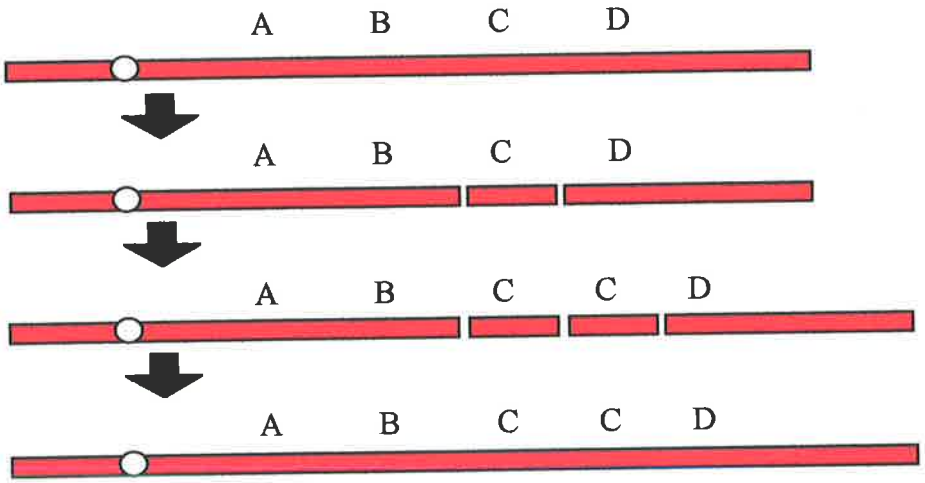
B) Inversion



C) Deletion



D) Amplification



Translocations are more common in haematological malignancies than in other cancers, and are often specific to a certain type of leukaemia. For example, the Philadelphia chromosome, which is created by the translocation t(9;22)(q34;q11), which was the first reported recurrent translocation, is a hallmark of chronic myeloid leukaemia (CML), as it is present in 90 - 95% of CML patients (reviewed in Dobrovic *et al* 1991). Similarly, the t(15;17)(q22;q12) translocation is highly specific to acute promyelocytic leukaemia, a subtype of AML (FAB subtype M3) (Pandolfi 2001). This specificity is thought to reflect both the different cellular environments in which the translocation initially takes place (that is, the different cell types from which the leukaemia derives) as well as the different molecular outcomes of the different translocations and the impact this has on the cell phenotype (reviewed in Barr 1998). The high degree of association between specific translocations and types of leukaemia makes translocation characterisation an important field of leukaemia research, both in order to understand mechanisms of leukemogenesis and to provide more specific, less subjective bases for diagnosis.

By convention, chromosomes that are derived from a translocation or other rearrangement are referred to as derivative chromosomes. To differentiate between derivative chromosomes, chromosomes are denoted according to which parental centromere they retain. For example, in the t(9;22)(q34;q11), the translocated chromosome which retains the centromere of the parental chromosome 9 is referred to as the der 9 chromosome. The other translocated chromosome is the der 22 chromosome.

1.2.2 Effects of Translocations on Gene Expression

As gross as the changes at the chromosomal level may appear under the microscope, it is the changes that translocations cause at the sequence level of the genome that are significant in discerning their importance. The genes located at or near the translocation breakpoint can be affected in two major ways, either of which may promote oncogenesis.

1.2.2.1 Dysregulation of Expression of Oncogenes by Translocation

An oncogene which does not have its coding region interrupted by a translocation but which is located in proximity to the breakpoint may be affected by juxtaposition to new regulatory sequences such as a promoter or transcriptional enhancer. In this case, the cell may undergo aberrant over-expression of the oncogene. This inappropriate expression can contribute to oncogenesis. Examples of this include the *cyclin D1* (*BCL1*) gene, a cell cycle regulatory gene normally expressed only during the G1 phase of the cell cycle, which is translocated adjacent to the immunoglobulin heavy chain (*IgH*) locus enhancers by the translocation t(11;14)(q13;q32) in mantle cell lymphomas (reviewed in de Boer *et al* 1997). While the *IgH* enhancers have long been believed to be the mechanism for upregulation of *cyclin D1* in translocations, this long-standing hypothesis has only recently been demonstrated experimentally (Marculescu *et al* 2002). *LMO2*, *TAL1* and *TAL2* are all upregulated by translocation to an *Ig* locus in T-cell acute lymphoblastic leukaemia, as are *BCL1* and *BCL2* in B cell non-Hodgkin's lymphoma (Marculescu *et al* 2002). Another example is the *c-myc*

gene at 8q24, which is translocated to the *Ig* loci at either 14q32, 22q11 or 2p11 in Burkitt's lymphoma (reviewed in Boxer and Dang 2001). Overexpression of *c-myc* promotes cell growth and oncogenesis through numerous target genes and pathways, including cell cycle regulation (reviewed in Dang 1999).

1.2.2.2 Formation of Fusion Genes by Translocation

If the coding region of a gene is disrupted by the translocation breakpoints on each of the two affected chromosomes, then the two fragments of one gene may be recombined with two fragments of the other gene to form two new genes, termed fusion genes. The best characterised example of a fusion gene is the *BCR-ABL1* fusion created by the Philadelphia chromosome translocation in CML patients (see Figure 1.3). Generally, only one of the two fusion genes formed by a translocation will be consistently expressed and functional, as most of the functional domains of the two original genes are recombined into the one fusion gene. An exception to this is the fusion genes formed by the t(15;17)(q22;q12) translocation in PML, where both fusion genes, *RARα-PML* and *PML-RARα*, are expressed and have leukemogenic effects in transgenic mice (reviewed in Rego and Pandolfi 2002).

As the number of identified fusion genes increases, it is becoming apparent that only a limited number of genes are involved. Many genes form fusion genes with more than one partner in a variety of different translocations. Examples of these include the *MLL* gene at 11q23, the *AML1* gene at 21q22 and the *ETV6* (*TEL*) gene at 12p12 (reviewed in Bohlander 2000) Whether this is a result of these genes existing at preferred sites of

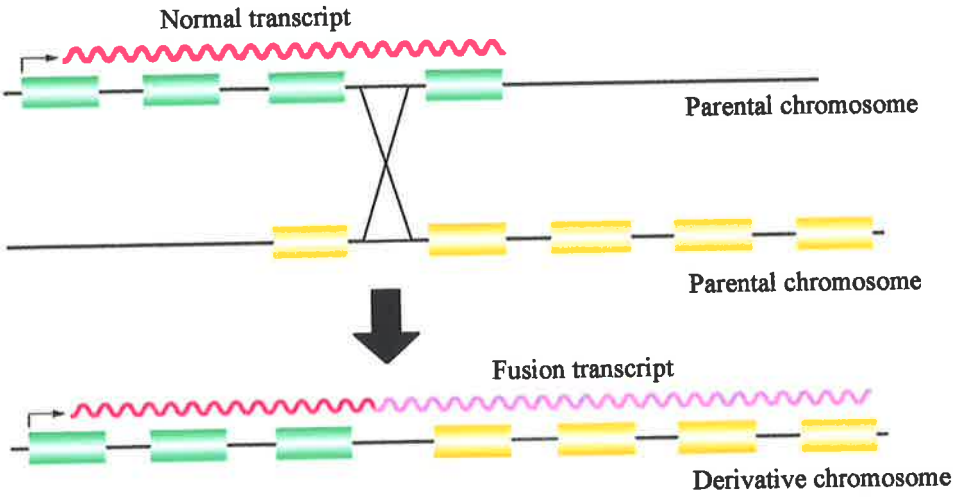


Figure 1.3. Formation of Fusion Gene by Chromosome Translocation.

Schematic illustration of the formation of a fusion gene by translocation. The top panel shows a normal chromosome encoding a normal gene, represented in green, which is normally transcribed, with the mRNA represented in red. The middle panel shows another normal chromosome encoding a normal gene, represented in yellow. In this case, the second gene is not normally transcribed, although this is not always the case. When translocation occurs, *via* the recombination indicated by the cross between the top two panels, within an intron of each gene, the result is the derivative chromosome shown in the lower panel, encoding the fusion gene. The fusion gene is transcribed under the control of the regulatory region of the first (green) gene, and the fusion transcript, shown in red and pink, is translated to produce a fusion protein with oncogenic function. Figure adapted from Griffiths *et al* 2000.

chromosome recombination or simply a true reflection of the importance of these genes in leukaemia remains to be determined (reviewed in Bohlander 2000). It is likely that both factors contribute to the frequency with which these genes are involved in translocation in leukaemia. For example, *MLL* is often involved in translocations due to site specific cleavage following topoisomerase II treatment (Aplan *et al* 1996), and *MLL* fusion genes have also been shown to directly cause leukaemia in transgenic mice, demonstrating the importance of *MLL* in leukemogenesis (Dobson *et al* 1999).

1.2.3 Mechanisms of Translocation

The mechanisms that give rise to translocations are most easily identified in cases of therapy-related leukaemia, which arise subsequent to treatment with chemotherapy or radiotherapy for a pre-existing cancer. Chemotherapeutic agents (for example, DNA topoisomerase II inhibitors (such as etoposide) and alkylating agents (such as cyclophosphamide)) are cytotoxic and target cells which are actively dividing. This is intended to target cancerous cells, but as haematopoietic cells are among the more actively dividing cells in the body, they are also affected. The mechanism of cytotoxicity exploited by each of these treatments involves double strand DNA cleavage, which is the first step required for recombination (or translocation). In support of this mechanism, specific DNA sites which are uniquely sensitive to double strand DNA cleavage have been identified within the *MLL* gene (Stanulla *et al* 2001) and *AML1* gene (Stanulla *et al* 1997). The most frequent translocations associated with therapy-related leukaemia include 11q23 rearrangements (Bloomfield *et al*

2002), which result in *MLL* fusion genes (reviewed in Ayton and Cleary 2001), the inv(16)(p13q22) (Andersen *et al* 2002), which results in the formation of the *CBFβ/MYH11* fusion gene (Marlton *et al* 1995), the t(15;17)(q22;q11) (Andersen *et al* 2002), which as already mentioned results in the expression of the *RARα-PML* and *PML-RARα* fusion genes (Goddard *et al* 1991) and 21q22 rearrangements (Slovak *et al* 2002), which result in *AML1* fusion genes (reviewed in Richkind *et al* 2000). These rearrangements also occur in *de novo* leukaemia, and so are not specific to therapy-related leukaemias. These translocations must therefore also be caused by mechanisms other than induction by therapy in *de novo* leukaemia.

It is more difficult to understand the mechanism of translocation in *de novo* myeloid leukaemias as patients have not been exposed to mutation-inducing therapy. It is generally thought that the initiating step is a double stranded DNA break (Richardson and Jasin 2000). The likelihood is that pre-existing deficiencies in DNA repair mechanisms, caused by other epigenetic or genetic changes in the cell, prevent the accurate repair of double-stranded DNA breaks that may occur during normal DNA replication, thus resulting in translocation. Many genes, including *ATM*, a gene implicated in leukaemia and other cancers, and *BRCA2*, a gene frequently inactivated in breast cancer, have been implicated in genomic instability leading to gross chromosomal rearrangements (Yu *et al* 2000, Boultwood 2001). *ATM* and *BRCA2* have both been shown to be involved in the repair of double stranded DNA breaks (Sharan *et al* 1997, Pandita 2002).

In lymphoid malignancies, translocations most frequently involve the immunoglobulin loci, which in normal B and T cells are involved in DNA rearrangement to produce a wide variety of antibodies and T cell receptors responsible for the immune response (reviewed in Bassing *et al* 2002). Illegitimate recombination (that is, recombination between two sequences which are not intended to be recombined) involving these loci leads to chromosomal translocation. As already described, translocations involving *Ig* loci lead to upregulation of the translocation partner genes, which include *cyclin D1* as described in Section 1.2.2.1.

1.3 Approaches to Breakpoint Identification

The first reciprocal translocation to be identified as such was the t(8;21)(q22;q22) in an AML M2 patient (Rowley 1973a). Shortly after this came the discovery that the Philadelphia chromosome, previously thought to be the result of a deletion of part of chromosome 22, was in fact a derivative of the t(9;22)(q34;q11) (Rowley 1973b). It was ten years later that the molecular consequences of the Philadelphia chromosome were revealed (de Klein *et al* 1982, Bartram *et al* 1983). This discovery was facilitated by the identification of the human homologue of the *c-abl* gene on 9q34 (Heisterkamp *et al* 1982). The approach used to identify *c-abl* disruption in t(9;22)(q34;q11) was Southern blotting, wherein genomic DNA from a cell carrying the translocation is restricted with a specific restriction enzyme and probed with a sequence thought to be near the breakpoint. The resultant bands were compared to those produced from material from a normal cell. If the probe is within the same restriction fragment as the breakpoint, aberrant bands are seen. This approach is still

used in breakpoint identification but in recent years, many new techniques and discoveries have advanced the process of mapping, cloning and identifying translocation breakpoints and breakpoint genes.

1.3.1 Candidate Gene Approach

The identification of *ABL* as one of the genes involved in the Philadelphia chromosome breakpoint (de Klein 1982) is an example of the candidate gene approach to identifying genes involved in translocations. Karyotypic analysis will give an approximate chromosomal location of the breakpoint (usually which chromosomal band is involved, and sometimes which sub-band). From this information, genes that have been mapped to the same region can be analysed to assess the most likely candidates from that group (usually genes already implicated in cancer, or implicated by their function). This approach is very dependent on the presence of a standout candidate (such as in the case of *ABL*, which was already known to be involved in leukaemia through studies of the Abelson murine leukaemia virus), and usually requires a degree of positional cloning beforehand to narrow the region of investigation (see Section 1.3.4). This approach therefore is enhanced by the development of other techniques allowing greater precision in defining the region of interest, and also by more thorough gene mapping of the chromosome region in question.

1.3.2 Fluorescence *in situ* Hybridisation (FISH)

The development in the mid-1980s of specific probes labelled with fluorochromes enabled direct probing of patient material, without the need to segregate the chromosomes by somatic cell hybridisation. This technique is substantially faster than Southern blotting, as all that is required is a metaphase spread of patient material and appropriate probes. The problem of available markers remains, as the cost of probing with a series of progressively closer markers in a step-wise fashion is quite expensive. This technique is best applied to confirming in new patients the molecular nature of translocations that have been determined in other patients by other means, rather than mapping novel breakpoints.

More recently, the development of full-chromosome paints and multicolour FISH allow the detection of all chromosomal abnormalities in a single experiment (reviewed in Kearney 1999), which is especially useful for detection of cryptic abnormalities which may be missed by standard karyotypic analysis. However, these techniques are not useful for fine breakpoint mapping, as they utilise markers with broad specificity (such as those specific for an entire chromosome).

1.3.3 Somatic Cell Hybridisation

Somatic cell hybridisation is a technique in which patient cells are fused with a rodent cell line. It is a powerful technique with applications to many fields of genetics (reviewed in Ringertz and Savage, 1976). For the purposes of translocation mapping

studies, somatic cell hybridisation is a useful means of segregating the human chromosomes so that they may be characterised separately. This is possible because when a human cell is fused with a rodent cell, the resultant hybrid cell then randomly discards the human chromosomes (while retaining the rodent chromosomes to ensure cell viability) until a stable configuration, usually containing one or few human chromosomes, is reached (Weiss and Green, 1967). Multiple hybrid cell lines can be constructed and screened for the retention of one translocation-derived chromosome without the retention of the other nor either of the two normal parental chromosomes. The construction of such a hybrid is illustrated in Figure 1.4. When such a cell line is found, it may be used in positional cloning analysis of the breakpoint.

1.3.4 Positional Cloning

Positional cloning can be performed when no information about the genes in the region of interest identified by karyotypic analysis is known. It is most frequently achieved by first segregating the derivative chromosomes *via* somatic cell hybridisation. Hybrid cell lines can then be used to map chromosomal markers from either parental chromosome to one or the other derivative chromosome, thus mapping the breakpoint with respect to the marker. The difficulty with this approach in the past was the lack of availability of reliably positioned markers, and the length of time required for a result to be achieved was almost prohibitive. However, once positional cloning had narrowed the region of interest, a candidate gene approach could be taken to expedite the process.

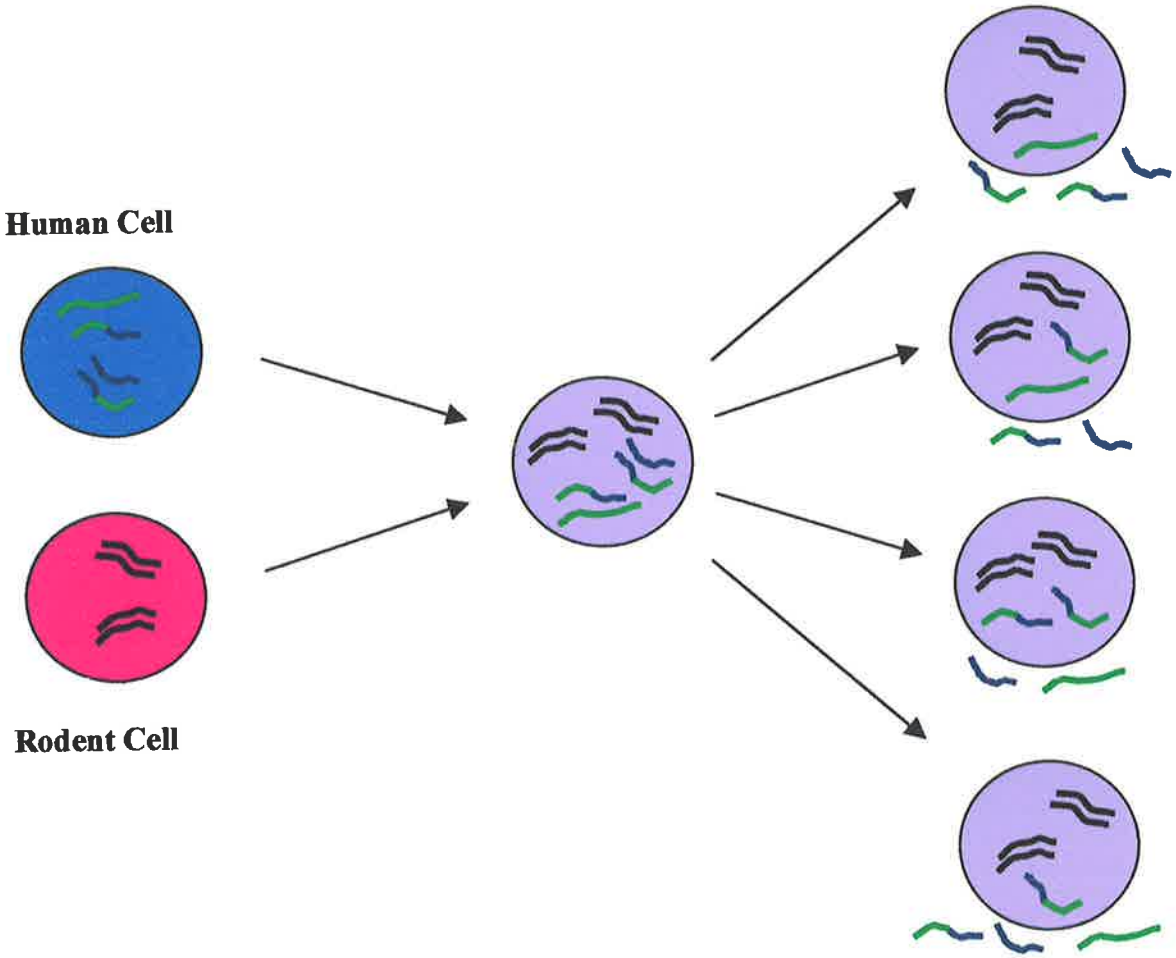


Figure 1.4. Schematic Representation of Construction of Somatic Cell Hybrids.

The leukemic patient cell on the left (in blue, containing parental and translocated chromosomes) is fused to a rodent cell (in pink, containing the normal rodent chromosome complement). The fusion results in a hybrid cell, shown in the middle of the diagram, which contains all of the human and rodent chromosomes. This hybrid randomly discards the human chromosomes until only a few are left. Possible outcomes of this are shown on the right of the figure, where discarded chromosomes are shown outside of the cell. Many hybrid colonies can be made and screened for retention of an informative combination of chromosomes (that is, only one translocated chromosome and neither of the normal parental chromosomes. Of the four hybrids shown, only the bottom hybrid is suitable for PCR analysis. The second hybrid from the top could be analysed using PCR markers for the blue chromosome but not the green chromosome.

1.3.5 Somatic Cell Hybrid PCR

Until recently, it had been almost impossible to take a completely systematic positional cloning approach to breakpoint identification. This was a result of the limited knowledge of the ordering of genes or markers along a chromosome. Without the benefit of this knowledge, information about one locus on a chromosome provides no information about any other locus, for the position of one with respect to the other was not known. This was a particular problem for mapping at the fine level, as markers could be positioned approximately by techniques such as *in situ* hybridisation (ISH) or, later, FISH, but such coarse mapping would still leave an area potentially containing thousands of genes, into which no further systematic progress could be made. Further progress then relied on a candidate gene having already been characterised and positioned within the region.

With the advent of the Human Genome Project and, since early 2001, the availability of near-complete scaffolds (a name used for large overlapping tracts of sequence) of sequence spanning the entire human genome, this situation has dramatically changed. Where before limits were imposed by the lack of ordered markers, now the entire genome, or any piece of it, is available for use as a marker and is precisely positioned with respect to every other piece. Therefore, a true "walking" approach can be adopted, where the assessment of each marker does provide information about other markers, and "steps" of any size can be taken in the process of narrowing the breakpoint region. The versatility and simplicity of this approach is enhanced by the use of PCR analysis rather than Southern analysis or FISH. This is principally because

of speed and ease of use, but is also significantly cheaper than performing the same number of steps with either of the other techniques. Since this systematic approach is purely positional, there is no need to determine whether a probe or marker contains the breakpoint, simply which side of the breakpoint it is on. In this scenario, PCR is favoured over Southern analysis and FISH for the smaller size of the respective markers or probes.

1.4 The Human Genomic Sequence

The work reported in this thesis was conducted over the years 1999-2002, and during that time the amount of sequence information available from the human genome resources has changed dramatically. References are made throughout the text to the different databases which were available at that particular time. These databases are described in detail in this section.

1.4.1 The GeneMap99 Genome Map

Prior to the existence of genomic sequence maps, the most informative source of data regarding genomic positional marker information was physical maps. The GeneMap99 radiation hybrid map (Deloukas *et al* 1998; see also http://www.ncbi.nlm.nih.gov/genemap/) is the most comprehensive of these. It was constructed by a large consortium of geneticists, using data obtained from typing two radiation hybrid panels for many markers. Radiation hybrids are distinct from somatic cell hybrids in that they are constructed from irradiated human DNA, such that only

small fragments of chromosomes remain intact. This allows determination of the presence or absence of a marker in a much smaller region than an entire chromosome (as would be the case with somatic cell hybrids), allowing much greater precision in mapping. Data derived from this map is useful for coarse mapping of breakpoints but has limited usefulness when compared to a full-sequence map.

1.4.2 The Human Genome Project

The Human Genome Project was launched by late 1990. The desire to obtain a complete human genomic sequence had arisen out of several observations, but most particularly that the generation of such a sequence would enable biomedical researchers to take a global view of the genome and allow more comprehensive analysis of gene function in genomic contexts. This vision is now being realised through the application of microarray expression studies, a technology that has grown out of the Human Genome Project.

The approach taken to sequencing the genome (reviewed in McPherson *et al* 2001) was termed hierarchical shotgun sequencing, and was based on the pre-existing technique of shotgun sequencing of smaller DNA molecules (such as bacterial genomes). The hierarchical component of the strategy is that the genome is first fragmented and cloned into a library; in most cases bacterial artificial chromosome (BAC; often referred to as clones in this thesis) libraries were used. These BACs were organised into a physical map, based on the presence of markers that had previously been mapped, often *via* physical maps such as GeneMap99, and BACs to be fully

sequenced were chosen so as to cover the entire genome. These selected BACs were then subjected to standard shotgun sequencing; that is, each BAC was restriction digested into fragments, appropriately sized for subcloning and direct sequencing. When these were sequenced to sufficient quality and redundancy, the sequence fragments were reassembled, to provide a full-length sequence of the BAC.

The final stage of the genome assembly, which is still ongoing, is the assembly of the BACs into full-length sequence of chromosomes. Each BAC was positioned in the genome by the presence of markers before it was sequenced, but assembly at the fine level has been difficult because of the imperfect accuracy of the initial marker mapping. This merely provided a framework, and it is not until sequence data is available that the BACs can be accurately and reliably positioned. To achieve this next step in the assembly, the International Human Genome Sequencing Consortium established a level of assembly termed contigs. In this context, contigs are composites of several adjacent, overlapping BACs, which have been assembled by BAC-end-matching. While BACs generally contain in the order of 0.1 - 0.5 Mb, contigs are generally 1 - 5 Mb. This process has given rise to the Draft Human Genomic Sequence which is available at the time of this writing (October 2002). Occasionally contigs are imprecisely assembled, and the distance between contigs and the sequence of these regions are still unknown, but on the whole the contig assembly now provides a high level of reliable coverage. This sequence data is available *via* the NCBI website (http://www.ncbi.nlm.nih.gov/).

1.4.3 Celera^{TM} Public Access Genome Database

A few months prior to the release of the Draft Human Genomic Sequence in 2000, the private biotechnology company Celera made their independent sequence database available to subscribers. This database was reportedly more thorough and complete than the NCBI assembly and was also better annotated with gene information. This database was unavailable to our department due to the cost of subscription, but early in 2001, a "freeze" of the database at the date of the publication of the genome assembly work performed at Celera (Venter *et al* 2001) was made freely available. This data was never updated but contained the entirety of Celera's sequence data, not including annotation, to that date. This thesis makes extensive use of that data, as well as of the NCBI assembly data, which continued to improve throughout 2001 and 2002.

1.5 Chromosome Band 1p36

Chromosome band 1p36 has long been known to be a frequent site of rearrangement in cancer, especially in neuroblastoma (Brodeur *et al* 1981, Lampert *et al* 1988) and leukaemia (Le Beau *et al* 1985, Bessho *et al* 1989). Deletions of 1p36, sometimes extending to 1p32, are frequent in many cancers (Mori *et al* 1998, reviewed in Knuutila *et al* 1999), as are both balanced and unbalanced translocations (reviewed in Olney *et al* 2002).

Several 1p36 translocations have been reported as recurrent in haematological malignancy. A summary of these is presented in Table 1.2. None of these translocations had been characterised at the molecular level prior to the beginning of the work described in this thesis.

Translocation	Disease	References
t(1;1)(p36;q25)	Diffuse Large B-cell Lymphoma	Dave *et al* 1999
inv(1)(p36;q21)	AML M2	van Limbergen *et al* 2002, Sait *et al* 2002
inv(1)(p36;q21)	CML	Alimena *et al* 1990, Secker-Walker *et al* 1995
t(1;3)(p36;q21)	MDS, AML	Moir *et al* 1984, Bloomfield *et al* 1985
t(1;7)(p36;q32)	MDS (Refractory Anemia with Excess Blasts)	Stefanescu *et al* 1994
t(1;7)(p36;q34)	AML M2	Specchia *et al* 1999
t(1;17)(p36;q21)	AML M3	Yamada *et al* 1983, Schwartz *et al* 1986
t(1;18)(p36;p11)	Diffuse Large B-cell Lymphoma	Speaks *et al* 1992 Dave *et al* 1999
t(1;9;22)(p36;q34;q11)	CML (involves *BCR/ABL*)	Dube *et al* 1989, Yehuda *et al* 1999
t(1;22)(p36;q11)	CML (involves *BCR/ABL*)	Dewald *et al* 1993, Nacheva *et al* 2000

Table 1.2. Balanced Recurrent 1p36 Translocations in Haematological Malignancy. The list was compiled by searching the Mitelman Database of Chromosome Aberrations in Cancer (http://cgap.nci.nih.gov/Chromosomes/Mitelman) and the literature.

1.6 Translocations Investigated in this Thesis

1.6.1 t(1;3)(p36;q21)

Chapter 3 details the investigation of the t(1;3)(p36;q21) translocation of Patient 1. This is a recurrent translocation which was included in Table 1.2. The t(1;3)(p36;q21) occurs in MDS and AML, and is generally accompanied by an elevated platelet count and dysmegakaryopoiesis (Bloomfield *et al* 1985, Grigg *et al* 1993). At the commencement of this study no work had been published toward the molecular characterisation of this translocation, although previous work in our laboratory had shown the breakpoint of a cytogenetically similar translocation in a different patient to lie between 1p36.2 and 1p36.33 by FISH (see Section 1.3.2) analysis (Varga *et al* 2001). Also prior to this study, hybrid cell lines had been constructed by fusion of Patient 1 cells with the PG19 murine fibroblast cell line (Dobrovic, unpublished). These hybrids had not been analysed and were stored in liquid nitrogen.

1.6.2 ins(12;1)(p13;p36p21)

Chapter 4 details an investigation of the rearrangement in Patient 2, which was an inverted insertion of 1p21 - 1p36 into 12p13 (ins(12;1)(p13;p36p21)). Previous work in our laboratory conducted using FISH analysis had shown the breakpoint in this rearrangement to be between 1p36.2 and 1p36.33 (Varga *et al* 2001).

1.6.3 t(1;3)(p36;p21)

Chapter 5 describes the investigation of the t(1;3)(p36;p21) translocation of Patient 3. The rearrangement in this patient was originally thought to be t(1;3)(p36;p23), but the mapping analysis determined the presence of a breakpoint at 3p21. The t(1;3)(p36;p21) had not been reported in the literature at the commencement of the study presented in this thesis, nor had any analysis been performed on this translocation in our laboratory.

1.7 Aims

The study detailed in this thesis investigated the molecular consequences of the 1p36 translocations of three AML patients detailed in Section 1.6. The aims of the investigation, with respect to each of the three translocations, were:

1) To map the 1p36 breakpoints, and to thus determine whether the breakpoints in the three patients are clustered in a common region, or whether there are multiple translocation target genes at 1p36;

2) To characterise the molecular effect of the translocations by investigation of changes to expression or involvement in fusions of any genes in the breakpoint regions which may be affected.

CHAPTER 2

Materials

and

Methods

CHAPTER 2

Materials and Methods

2.1 Materials

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

2.1.1 Enzymes

Enzymes for general DNA manipulations were purchased from Roche (Basel, Switzerland), Progen Industries (Brisbane, Australia), Fermentas (Vilnius, Lithuania) or New England Biolabs (Beverly, USA). HotStarTaq polymerase was purchased from Qiagen (Venlo, Netherlands). T4 DNA ligase was purchased from Promega (Madison, USA). Proteinase K was purchased from Merck (Whitehouse Station, USA).

2.1.2 Buffers and Solutions

General solutions and buffers were made up to the specified concentration with MilliQ water. pH was assessed where appropriate with either an electronic pH meter or by use of pH strips. Solutions were sterilised either by autoclaving (103 kPa, 121 °C for 20 min), or by filtering through a 0.22 µm membrane according to the method given in Sambrook *et al* (1989).

Solutions used for PCR were made with Ultra Pure Water (UPW; Fisher Biotec, Perth, Australia) which is certified DNase and RNase free.

dNTP stock: Initially, 40 mM stock solutions of dATP, dCTP, dGTP and dTTP (10 mM each) were prepared in UPW and the pH adjusted to between pH 7.0 - 8.0 with 1 M Tris. Later, pre-made dNTP solutions were purchased from Promega.

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

2 x Loading Buffer for RNA: 500 µl of de-ionised formamide, 100 µl of 10 x MOPS, 167 µl 37% formaldehyde, 133 µl UPW, 100 µl glycerol, 3 µl of 10 mg/ml ethidium bromide, 0.025 % bromophenol blue and 0.025 % xylene. Stored at - 20 °C.

20 % Sodium Dodecyl Sulphate (SDS): 20 g of sodium dodecyl sulfate (Sigma, St Louis, USA) per 100 ml of dH_2O.

20 x SSC: 3 M NaCl and 0.3 M sodium citrate, pH 7.2.

1 x TAE: 40 mM Tris base, 20 mM NaAc, 2 mM EDTA, adjusted to pH 7.8 with glacial acetic acid.

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

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

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

2.1.3 Media

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

MacConkey Agar Plates with selective antibiotics and colour selection: Red-white selection media was prepared by dissolving 52 g MacConkey (Oxoid, Basingstoke, England) powder in 1 L MilliQ water. The solution was autoclaved and allowed to cool to 50 °C before the addition of 1 ml of 100 mg/ml ampicillin. The solution was then poured into 85 mm petri dishes and allowed to set. Plates were used immediately or stored at 4 °C for a maximum of 1 month.

Dulbecco's Modified Eagle Medium (DMEM) (Gibco, Invitrogen, Carlsbad, USA) cell culture media was made according to manufacturer's instructions. For standard cell culture, 10% fetal calf serum (FCS) (CSL, Melbourne, Australia), 60 μg/ml penicillin (Sigma), 50 μg/ml streptomycin (Sigma) and 200 μg/ml L-glutamine (Sigma) were added. For somatic cell hybrid culture, HAT media (final concentration 100 μM hypoxanthine, 0.4 μM aminopterin, 16 μM thymidine, 3 μM glycine) was made by addition of 1 ml of 1 x HAT supplement (Gibco, Invitrogen) to 100 ml growth media.

2.2 Methods

2.2.1 Basic Nucleic Acid Isolation

2.2.1.1 Mononuclear Cell Isolation from Peripheral Blood

Blood samples were diluted to four times the original volume with phosphate buffered saline (PBS) in a 50 ml centrifuge tube. Approximately 8 ml of Ficoll-Paque (Pharmacia, Peapack, USA) was gently layered under the dilution using a pasteur pipette. The sample was then centrifuged at 570 x g for 20 min. Peripheral blood mononuclear cells (PBMNC) were removed from the interphase using a pasteur pipette in a total volume of 2 - 5 ml. These were placed in a 10 ml centrifuge tube and diluted to 10 ml with PBS. A cell count was performed before the cells were collected by centrifugation and then resuspended in 1 ml PBS. The suspension was transferred to a 1.5 ml Eppendorf tube and centrifuged again. The supernatant was removed prior to nucleic acid extraction using the method outlined in Section 2.2.1.3.

2.2.1.2 Thawing of Frozen Samples

Bone marrow or PBMNC samples frozen in liquid nitrogen storage were thawed by warming the ampoule in a beaker of warm water (approximately 37 °C) with agitation. Once thawed the sample was transferred to an Eppendorf tube and the cells collected by centrifugation. The supernatant was removed prior to nucleic acid extraction.

2.2.1.3 DNA Isolation from Cell Suspensions

DNA isolation from cell pellets was performed according to the method of Miller *et al* (1998). Approximately 2 x 10^7 cells were resuspended in 500 µl of TES buffer. To this, 30 µl of 10 mg/ml Proteinase K solution was added and the suspension was mixed before the addition of 30 µl of 20 % SDS, further mixing, and incubation at 37 °C overnight. An equal volume (560 µl) of 3 M NaCl was added. The sample was shaken vigorously and placed on ice for 10 min. The sample was centrifuged at 13000 x g for 10 min and 500 µl of supernatant removed to each of two fresh tubes. Two volumes (1 ml) of absolute ethanol was added to each tube. The samples were mixed by inversion and centrifuged at 13000 x g for 10 min. The supernatant was poured off and the remaining DNA pellet washed in 70 % ethanol. The tubes were centrifuged briefly at 13000 x g and the supernatant was removed. The pellets were air dried for one hour before being resuspended in an appropriate volume of 1 x TE buffer. DNA solutions were stored at 4°C.

2.2.1.4 RNA Isolation from Cell Suspensions

Approximately 10^7 cells were resuspended in 1 ml of Trizol (Gibco, Invitrogen) or Tri-Reagent (Sigma). RNA was extracted according to the manufacturer's protocol. RNA pellets were dissolved in UPW.

2.2.1.5 Plasmid DNA Isolation

Bacterial colonies were inoculated into 3 ml LB and incubated with shaking at 37 °C overnight. Plasmid minipreps of overnight cultures were performed using 1 ml of the overnight culture and the QIAprep Spin Miniprep Kit (Qiagen), used according to the manufacturer's instructions. Two µl of the final elution containing the plasmid DNA was used as the template in sequencing reactions.

2.2.2 Basic Nucleic Acid Manipulation

2.2.2.1 DNA Precipitation

To the solution of genomic or plasmid DNA, 1/10 of the initial volume of 3 M NaAc was added and mixed. Twice the volume of absolute ethanol was then added and the solution was mixed. The mixture was placed on ice or at - 20 °C for 10 min and then centrifuged at 13000 g for 10 min. The supernatant was removed and the DNA pellet washed with 70 % ethanol before being centrifuged at 13000 g for another 10 min. The supernatant was removed and the pellet air-dried briefly before resuspension in an appropriate volume of 1 x TE or UPW. DNA solutions were stored at 4°C.

2.2.2.2 DNA Sequencing

The ABI Prism Big Dye Terminator cycle sequencing kit (version 2 or version 3) (Amersham, Amersham, England) was used according to the manufacturer's instructions, using a PTC-100 thermal cycler. Cycling conditions were 96 °C for 30 s, an appropriate annealing temperature dependent on sequencing primer for 30 s, 60 °C for 4 min for 25 cycles. Reactions were stored at 4 °C until ready to purify. Purification was performed by precipitation with 75 % isopropanol. The supernatant was removed and the samples were dried for analysis by the Department of Molecular Pathology, IMVS, Adelaide. DNA sequence analysis was carried out using the Sequencher (Genecodes, Ann Arbor, USA) analysis package.

2.2.2.3 DNA/RNA Quantitation

DNA or RNA solutions were diluted 1:50 in UPW for spectrometry at 260 nm on a DU 650 Spectrophotometer (Beckman-Coulter, Fullerton, USA).

2.2.2.4 DNA Electrophoresis

Agarose gels of 0.8 % - 2.0 % were made by boiling the appropriate amount of agarose in 0.5 x TBE buffer. Electrophoresis was performed in 0.5 x TBE buffer at 100 V. Ethidium bromide (EtBr) was added to the gel and the buffer to a final concentration of 0.1 μg/ml. After electrophoresis, gels were photographed under UV light and analysed using the Kodak 1D Gel Documentation System (Kodak, Rochester, USA).

2.2.2.5 RNA Electrophoresis

Agarose gels of 0.8 % - 1.2 % were made by boiling the appropriate amount of agarose in a final concentration of 1 x MOPS buffer (Sigma). After the agarose had cooled to 50°C, formaldehyde was added to a final concentration of 1.2 M before pouring. Electrophoresis was performed in 1 x MOPS buffer at 35 - 50 V. After electrophoresis, gels were photographed under UV light and analysed using the Kodak 1D Gel Documentation System.

2.2.2.6 Restriction Endonuclease Digestion of DNA

Genomic DNA was digested with 10 U of restriction enzyme per µg of DNA in a total volume of 40 µl of appropriate 1 x reaction buffer as supplied by the manufacturer. PCR products were digested by removing a 10 µl aliquot from the PCR reaction tube and digesting in a total volume of 20 µl. One times Bovine Serum Albumin (BSA) (New England Biolabs) was also added unless already present in the restriction enzyme buffer.

2.2.2.7 DNA Ligation

Restricted DNA was ligated using T4 DNA ligase (Promega). The reaction mix was prepared according to the manufacturer's instructions and incubated in a 1 L room temperature water bath which was cooled gradually overnight to 4 °C.

2.2.3 Polymerase Chain Reaction (PCR)

2.2.3.1 Standard PCR

Standard PCR reactions were carried out in 0.5 ml tubes. Each reaction contained a final concentration of 0.2 µM of each oligonucleotide primer, 200 µM of each dNTP, 0.5 U of HotStarTaq polymerase, 1 x HotStar PCR buffer, 2.0 mM $MgCl_2$ and 100 ng of template DNA. The reaction was made up to a final volume of 50 µl for genomic PCR or 25 µl for RT-PCR using UPW and cycled in a PTC-100 or PTC-200 Programmable Thermal Cycler (MJ Research Inc, Boston, USA). The polymerase was activated as per manufacturer's instructions by incubation at 95 °C for 15 min, and reactions were then cycled for 30 - 45 cycles of denaturation at 94 °C for 45 s, annealing at 60 °C - 68 °C (dependent on primers) for 1 min and extension at 72 °C for 1 min. Touchdown PCR was frequently performed when optimising a new pair of primers. Touchdown PCR consisted of the same initial activation step, 10 cycles of 95 °C for 45 s, 65 °C - 68 °C minus 0.5 °C per cycle for 1 min and 72 °C for 1 min, followed by 20 - 35 cycles of 95 °C for 45 s, annealing at 60 °C - 63 °C for 1 min and 72 °C for 1 min.

2.2.3.2 Long Template PCR

PCR Amplification of large products or products of unknown size was performed using the Expand Long Template PCR System (Roche) in thin wall 0.2 ml PCR tubes. The manufacturer's protocol was followed, using the base concentration of 1.5 mM $MgCl_2$ (Buffer 3 as provided) in a final volume of 50 µl. Cycling conditions were

chosen in accordance with manufacturer's instructions, primers used and expected size of products. Generally, an initial step of 94 °C for 2 min was performed, followed by 10 cycles of denaturation at 94 °C for 10 s, annealing at 60 °C - 68 °C (dependent on the primers) for 30 s, and extension at 68 °C for 4 - 10 min (dependent on the expected product length). This was followed by 25 - 35 cycles of denaturation at 94 °C for 10 s, annealing at 60 °C - 68 °C for 30 s, and extension at 68 °C for 4 - 10 min plus 20 s per cycle. A final extension step of 7 min at 68 °C was also performed.

2.2.3.3 Reverse Transcription (RT) using MMLV

For standard RT, a primer mix containing 1 µg of template RNA, 2 µl of random hexamers (250 ng/µl) and UPW to a final volume of 12 µl was made in a 0.5 ml tube. The tube was incubated at 70 °C for 10 min and then placed on ice. After 3 min, an RT mix containing 4 µl of 5 x first strand cDNA synthesis buffer (Gibco, Invitrogen), 2 µl of 0.1 M DTT, 1 µl of 40 mM dNTPs, 1 µl of UPW and 1 µl of 200 U/µl MMLV-Reverse transcriptase (Gibco, Invitrogen) was added. An RT mix without the MMLV enzyme was always included as a negative control. These mixtures were then incubated at 37 °C for 60 min. UPW was added to a final volume of 80 µl. cDNA samples were stored at 4 °C. From each newly synthesised cDNA mixture, 2 µl was seeded into each subsequent 25 µl PCR reaction.

2.2.3.4 Reverse Transcription (RT) using Superscript II

RT using Superscript II (Gibco, Invitrogen) used for generating cDNA for 3' RACE, 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, Invitrogen). All reagents, with the exception of UPW, were supplied by the manufacturer. Briefly, 1 µg of total RNA was resuspended in a final volume of 24 µl in a 0.5 ml tube using UPW. After the addition of 1 µl adapter primer (AP), the solution was mixed gently and collected by brief centrifugation. Each sample was then heated to 70 °C for 10 min then transferred immediately to 50 °C. For each sample, a reaction mix was then prepared which contained: 7.5 µl of UPW, 5 µl of 10 x PCR buffer, 5 µl of 25 mM $MgCl_2$, 2.5 µl of 10 mM dNTP mix and 5 µl of 0.1 M DTT. This was prepared as a master mix and prewarmed to 42 °C. To each RNA sample, 25 µl of the above mix was added and this was immediately followed by the addition of 1 µl (200 U) of Superscript II. Incubation at 50 °C was continued for 50 min then terminated by incubating the tubes at 70 °C for 15 min. Finally, 1 µl of RNase H was added and the tubes were incubated for 20 min at 37 °C. Two µl of the reverse transcription reaction was used as template for subsequent PCR.

2.2.3.5 DNase I treatment of RNA

Where genomic DNA contamination could lead to misleading results from RT-PCR, RNA was treated with RNase free DNase I (Gibco, Invitrogen) prior to reverse transcription. For every 1 µg of RNA, 1 x DNase I reaction buffer (Gibco, Invitrogen), 1 U DNase I Amplification Grade (Gibco, Invitrogen) enzyme and UPW to a final volume of 10 µl was added. The mixture was incubated at 37 °C for 15 min before the addition of 1 µl of 25 mM EDTA and incubation at 65 °C for 10 min. The 10 µl mix was added directly to the primer mix for reverse transcription.

2.2.3.6 Procedure for 3' RACE

One µg aliquots of poly(A)$^+$ selected RNA (see Section 2.2.4) were reverse transcribed using Superscript II and the AP primer from the 3' RACE kit (Gibco, Invitrogen). The Expand Long Template RT-PCR System (Roche) was used in all subsequent PCR amplifications. The reverse transcription product was amplified with an appropriate specific forward primer and the Abridged Universal Amplification Primer (AUAP) (Gibco, Invitrogen) which was added after 10 cycles of PCR had been performed to enrich the reaction for the specific product. PCR was performed with an initial step at 94 °C for 2 min, followed by 10 cycles of 95 °C for 30 s, 65 - 68 °C (dependent on primer) for 30 s and 68 °C for 5 min. The reactions were removed from the thermal cycler to ice and allowed to cool before the addition of the AUAP primer to the reactions. The reactions were mixed and briefly centrifuged before being returned to the thermal cycler. One min was allowed for the reactions to return to temperature before resuming cycling for 45 cycles of 95 °C for 30 s, 65 - 68 °C (dependent on primer) for 30 s and 68 °C for 5 min. Second round RACE PCR was performed using a nested specific forward primer, the AUAP primer, and identical cycling conditions as for the first round PCR. Two µl of a 1:50 dilution of the primary reaction was used as template for this secondary RACE PCR.

2.2.3.7 Inverse PCR

Inverse PCR is a technique used to amplify regions of unknown sequence which are adjacent to regions of known sequence. Primers are designed to the known sequence, but are oriented so that they will extend away from each other, rather than towards

each other as in standard PCR. These primers are used to amplify a circularised template, prepared by restriction with an enzyme which is known to cut on one side of one primer binding site but not in between the two primer binding sites, followed by ligation. The amplification thus occurs over the unknown region of sequence containing the restriction site (see Figure 2.1).

One µg of genomic DNA was digested with an appropriate restriction enzyme, under conditions appropriate to that enzyme. Restricted DNA was precipitated and resuspended in 20 µl UPW. Ten µl of this was added to a 40 µl ligation reaction mix, consisting of 34.7 µl UPW, 0.3 µl (120 CELU or 1.8 Weiss units) T4 DNA ligase and 5 µl of 10 x ligation buffer. Ligation was performed in a 1 L room temperature water bath which was allowed to gradually cool to 4 °C overnight. The following day the reaction mix was ethanol precipitated and resuspended in 20 µl UPW. Two µl of this resuspension were used in the subsequent PCR reaction. This was performed using the Expand Long Template PCR System (Roche). Reaction mixes were made according to the manufacturer's protocol. Primers were designed specifically for each inverse PCR. Long Template PCR was performed with an initial step at 94 °C for 2 min, followed by 10 cycles of 95 °C for 30 s, 61 - 65 °C (dependent on primers) for 30 s and 68 °C for 10 min. This was followed by 25 cycles of 95 °C for 30 s, 61 - 65 °C (dependent on primers) for 30 s and 68 °C for 10 min plus 20 s per cycle. Second round inverse PCR was performed using two nested primers and 2 µl of a 1:50 dilution of the first round PCR. Conditions were otherwise identical to first round inverse PCR.

Figure 2.1. Schematic of Inverse PCR Design.

The three panels illustrate the relative arrangement of the PCR primers, restriction enzyme sites and the translocation breakpoint in each of three states.

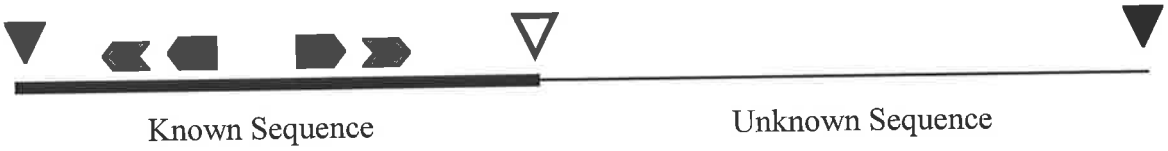
In all cases, the thick black line represents known sequence and the thin black line represents unknown sequence. Other features are indicated by symbols as shown in the key.

A) Patient Chromosome. Once the approximate position of the breakpoint is known, primers are designed, oriented as shown, as close to the breakpoint region as is feasible. Restriction enzymes are chosen such that there is a recognition sequence as near as possible on the far side of the primer furthest from the breakpoint, but no restriction site between the same primer and the breakpoint. The position of the restriction site closest to the breakpoint on the unknown sequence is of course not known. On the normal untranslocated chromosome, the position of this site is usually known, and the expected size of this normal inverse PCR product can be predicted.

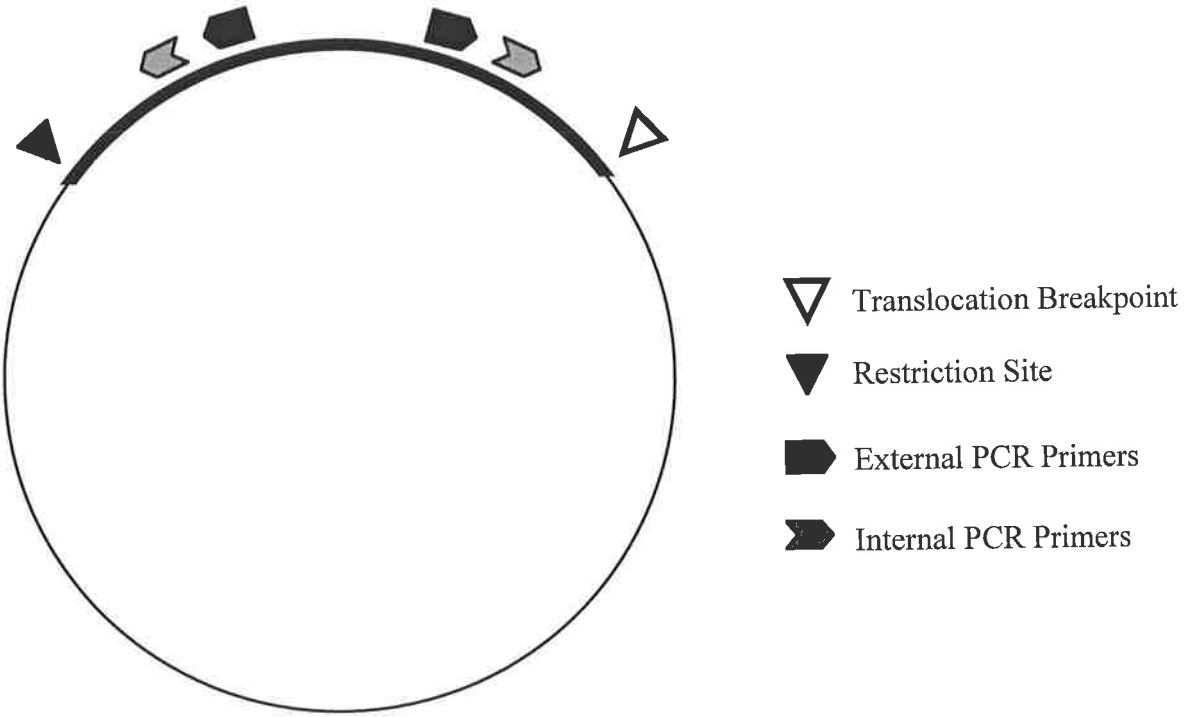
B) Restricted and Circularised Template. After restriction and ligation, the template is circularised such that the primers will amplify across the unknown sequence.

C) Inverse PCR Product. The amplified product is linear and contains known sequence at either end but unknown sequence in the middle. To determine the identity of the unknown sequence, sequencing is usually performed using a primer that will sequence across the restriction enzyme recognition site rather than the breakpoint, as the position of the breakpoint is not precisely known.

A)


Known Sequence
Unknown Sequence

B)


Translocation Breakpoint
Restriction Site
External PCR Primers
Internal PCR Primers

C)

2.2.3.8 Direct Purification of PCR Products

PCR product was purified directly from the reaction mix using the Concert Rapid PCR Purification System (Gibco, Invitrogen) or the QIAquick PCR Purification Kit (Qiagen). The manufacturers' instructions were followed for all purifications.

2.2.3.9 PCR Product Purification from Agarose Gels

For size purification of a specific PCR product from an amplification reaction generating more than one product, the reaction mix was electrophoresed through a low percentage (0.8 % - 1.2 %) standard agarose or low melting point agarose gel. Following visualisation of the gel under UV light, the band of interest was excised from the gel with a sterile scalpel blade. The agarose slice was transferred to a 1.5 ml Eppendorf tube. Purification of the PCR product from the agarose was performed using the Wizard PCR Preps DNA Purification System (Promega) or the QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer's directions.

2.2.3.10 PCR Product Cloning

On occasion, longer PCR products proved difficult to gel purify and sequence directly. In these instances, PCR reactions were purified as in Section 2.2.3.8 and ligated into pGEM®-T vector using the pGEM®-T Vector System II kit (Promega). Ligated products were transformed into JM-109 competent cells (Promega) by heat shock, and these cells were plated on red-white colour selection agar plates. White colonies were screened by PCR to ensure that they contained the correct size insert,

and then inoculated into 1 ml LB media. Cells were cultured overnight, following which plasmid preps were prepared.

2.2.3.11 PCR Primer Sequences

Primer sequences were designed using web-based and local programs, including Primer3 (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi), the Amplify package (Bill Engels, University of Wisconsin) and the Oligonucleotide Properties Calculator (http://www.basic.nwu.edu/biotools/oligocalc.html). Primers were purchased from Geneworks (Adelaide, Australia). Primers were resuspended in UPW to give a final stock concentration of 50 nM and stored at - 20 °C. Working dilutions of 5 nM were made and stored at - 20 °C or 4 °C. Oligonucleotide sequences of primers used in this thesis are presented in Table 2.1.

2.2.4 Northern Analysis

DIG-labelled RNA probes were prepared according to the manufacturer's instructions. RT-PCR products for use in the transcription reaction to generate the probe were themselves generated by standard RT-PCR using gene-specific primers with a T7 promoter sequence (ggatcctaatacgactcactatagggagg) at the 5' end of the reverse (anti-sense) primer. The product was purified as described in Section 2.2.3.8. The purified product was then used as template in the transcription reaction to generate DIG-labelled RNA probe, using reagents from the Strip-EZ™ RNA T7 kit (Ambion, Austin, USA), according to the manufacturer's instructions.

3'GATA2f1	cttggtgaagagaaccaaatgcag		425F1Af1	ctggcaggataccctctgctaaaa
3'GATA2r1	aaagtctctgctgaccaatgttcc		425F1Ar1	gatacactctttggccaccgacaa
333E3f1	cccatcatttgccttatcaagtgg		425F1Zf1	gaagaggccaaggacatcagaact
333E3r1	ggcagaatccctcaagagaaagg		425F1Zr1	agggttaggcacctacctctgaaa
38EG178f1	ctcctcgtcttccacttttcagag		425F2Af1	gtttgagtaattcccagagagacac
38EG178r1	tcagataaatcccagcaacagctcc		425F2Ar1	ctgggtcttaattttccttgacagc
38EG183f1	acaccgcacatttatcatctcacc		425F2Zf1	ctgagtagttctgggctgtttctc
38EG183f2	tcgaatcaggatggagcttgtcag		425F2Zr1	tccacttagtctccaaggactcca
38EG183f3	ggactcagtcctgctccatgagtc		425F3Af1	agcaaacccctgttgtggaactaga
38EG183r1	acttctgccaaaaacctcaagagc		425F3Ar1	ctagacacgagaaacacatccgaca
38EG183r2	attctgcacaggtggaggttcc		425F3Zf1	ttcctccagcccataaacagctaa
38EG183r3	gatcgcattccttttttatggctgc		425F3Zr1	cagacagccaatttcaacactgct
38EG183r4	gctttgagagtctgtctgcaagtc		493AP15f1	accttgtggtgacagcaatctttc
38EG188f1	gaaatgcttcccttcaggagagac		493AP15r1	acatgacacgagatcatccttacc
38EG188r1	ggatgaagctcttagatccccaag		493AP16f1	ctttcctatgggaagcatgaggac
38EG188r2	gcacttggccacttgtgtatgtgc		493AP16f2	gcggagatgattcttgctgcaaag
38EG189f1	gtacttggggaaagtggaagatcc		493AP16f3	tgaatcagcctacagccataccac
38EG189f2	cttgggtgacacttgcagaggtg		493AP16r1	gaatcatctccgcagtgagcatag
38EG190f1	acctctccctcctctgcattactg		493AP16r2	tttggacaggatccacagcaagtc
38EG190f2	attctgttttgaaggggggagctcc		493AP16r3	ctcatgcttcccataggaaagcag
38EG203f1	ctgggaataggaccaaatagggtg		493AP17f1	gagagttatccagattgtgtgtag
38EG203r1	ctctgcatctgggttctccttc		493AP18f1	tgaactcaggtatctgcccacctc
38EG210f1	caacatgtgtctacaccaccatcg		493AP19f1	tggctgggcacagtgactcatgc
38EG210r1	cctgtcatatccacaatcatgctcc		493AP20f1	ggaatttggcccagataaagtgtgg
38EG211f1	gaagtgttgtcaccccagaagttg		493AP20f2	ttctgaaaagtgaaaaagaggtgcg
38EG211f2	agagactctggaagcacctctttg		493AP20r1	ttcttgacttctcccaaacagtggg
38EG211f3	cgttgccttggctagattgttacc		493AP23f1	tccttctgcacttgtggaatatcg
38EG211r1	ggtaacaatctagccaaggcaacg		493AP23r1	aggctcatgaatttccctagaagc
38EG211r2	agtaaccctcaaagatggggacag		493AP31f1	gataccttatgggggcaatcagac
38EG211r3	tgtgatcaggccacgctcgaaag		493AP31r1	gattgggaggagaagggtattcac
38EG212f1	tagcttcaagcagatgtgggtcac		493AP61f1	accaaaaccctctgctatccagtc
38EG212r1	gacccttgtttagaatgcggactc		493AP61r1	tatttgtgtgtctgtggctgctcc
38EG213f1	taactctttacgccatggtccttc		5'LOC131656f1	cagatcctgagaaaagtgactagc
38EG213r1	tgcatgcacttggtaaccacatag		5'LOC131656r1	tcaaatcctcatctgctgagaacc
38EG217f1	gagaattggttctcctgagactgg		5'LOC90670f1	agggagagcaccagaacataaacc
38EG217r1	ccctagtgtggcagtttttgtgtc		5'LOC90670r1	aaggaatctcttgaggctgaatgc
38EG239f1	tctccaagctcctaagcatctcac		763B130r1	tcaccgcactcactccccagc
38EG239r1	ggtggatttcttccagaagatgcc		763B130r2	ggtttctcatcgcaggtccatcag
38EG357f1	acttctgcagatgagaagctctgg		95H11f1	atgctgtagcacgctttcacagc
38EG357r1	tgtcacaagccttccaagtagagg		95H11f2	cccactctagtagtttgaacaaccc
38EG40f1	acgctgctaccgtaacttcatcag		95H11r1	ccaacatgtagctgccaacaactg
38EG40r1	ttcacacctagcaatgcctgtagc		95H11r2	gttgaaacaagagctagacgagcac
38EG76f1	gttctcagactttcttccctgcac		95H12r1	ccttgagcctactaaggcttttcc
38EG76r1	cagggaaaagatcaactctcagtgg		95H12r2	aagctccagaccactttctcatgc

Table 2.1. Primer Sequences. Continued over the page.

AMPD2f1	ctcctcgtggtacagatggaaagtc		ITGA9f1	ctgcctgtgttctcttctttccac
AMPD2r1	gtgcctgacatcctagacctgtg		ITGA9r1	aactggcctggaagtctaggatg
ARHGEF16f1	tacttccacccctacatcgcctac		KIAA0342f1	cagaagctttctgtgccttcagag
ARHGEF16r1	gaacaggaaaaggtagcacgttgg		KIAA0342r1	aaggtgctcagggttgaatagacc
CHL1f1	tggtggtccctcacaaagagactg		LOC152440f1	tgtctgaggaactagcgattcagc
CHL1r1	cagtccaactcctacgaagctctg		LOC152440r1	attgcctccttccttctgtaaggc
D1S2661f1	agaggggaagaaaagaaggaagga		LRRFIP2f1	cacagggccctaaatacaattctgc
D1S2661r1	gagcctcagattcttcattggtga		LRRFIP2r1	atatgagaggacaaggccttccag
D1S468 f1	aaaccatgaattgctgataattaaccg		PCAFf1	caaaaatctgcgaaacatctcagc
D1S468 r1	cacacttccctctccaaaattggg		PCAFr1	tatgcaatgtgggtcataatgtgc
D3S12f1	agaatgtagcttgggaaatccctg		RARBf1	gtcaggaattggttacccctgtc
D3S12r1	tccctagaccactcattgcttcc		RARBr1	tagggagctgggttctagtcac
DBVY250f1	agagagaaatgcgacgccaagtag		RBMS3f1	tatctgacgaacctaaggggtcac
DBVY250r1	catttccaactatactcccgaacg		RBMS3r1	caagttccaaggtaagccacactg
DBVY251f1	cttgggaacaagatttctctgagg		rEVI1f1	aagccctttatctgtgaggtctgc
DBVY251r1	ctctatggaaggtagtcccttagc		rEVI1r1	atcaaaggaggcctgtggtacaag
DBVY252f1	caccccagtcacacaggataaag		rGATA2f1	cgtcttcttcaatcacctcgactc
DBVY252r1	ggcctcttgtcttcccatttctc		rGATA2r2	tagggtcaggagacacttctttgg
DBVY253f1	agcttgggacacgtctttactgag		rGOLGA4f2	caggagcaggaagatcttgaactg
DBVY253r1	cttagctaaacaagcctggcagtg		rGOLGA4f3	ccatacggatgtctcactctttgg
DBVY255f1	ctcccacagacagacagaaagtg		rGOLGA4r2	agtctgcaggtcccttacttttgc
DBVY255r1	aggaggagatgtatccgagtgtg		rGOLGA4r3	aggcagtctgcaacttatctgtgg
DBVY260f1	ctgcagtattaccacctgtggac		rGOLGAf4	gggagtttaatacacagctggcac
DBVY260r1	acagagaaggaaagccaaaaccag		rMEL1f1	agagaccatgacagagaagctgga
DBVY270f1	atgctctggctcttgtctctctg		rMEL1f2	actgtgcaggcaggctaagaacc
DBVY270r1	attaaggccacctgagacagtcc		rMEL1f3	tagtgtgtggctgcttctggactc
DBVY280f1	cagtctggtccttcttcccttag		rMEL1r1	ctcgtctaaaagtgcgtggttgtc
DBVY280r1	gaggctgagatggagaaggaaatc		rMEL1r2	tgagcacaccagtctatccattcc
DBVY290f1	agggtggaatggaactaaatgtgc		RPN1f1	tgaaggccttttacagatgtcagc
DBVY290r1	tggattgagatgatcgaagactgg		RPN1r1	atgagaagctcatctcaggaaagc
ENTPD3f1	gctaagtgtgaaagggaatgtgcc		rTELf4	tcctgatctctctcgctgtgagac
ENTPD3r1	ctctttcagttccctccacactc		rTELr3	tcgaggcactggaacatgaagtgg
FBXL2f1	gcaaggaaatgtctgagagtttgg		TGFBR2f1	ttcctccattgacaaagtcactcc
FBXL2r1	caaggtccaacagaattcctctcc		TGFBR2r1	tacctgagaggaagattgcagagg
FHITf1	gccaattgaaagccatagtgacag		TNFR2a1	ccactccccaccttcaattcctgg
FHITr1	gcagggatcttgcattctagtgag		TNFR2s1	ggcaaggctacactggtttcccc
GOLGA4f1	cagctaataaactggcacccttcc		TP73a1	acagaggtgaggcaggtctcccg
GOLGA4r1	tgcacttaccagaactcctctcag		TP73s1	ctctggtcctgcctgctcaccc
GR6f1	tcaaatggtagtgggtgggtacag			
GR6r1	ccattgccctattctagcatcctc			

Table 2.1. Primer Sequences. Continued from previous page.

Two μg of total RNA from PBMNC was poly(A)$^+$ selected using the Oligotex mRNA Kit (Qiagen) and electrophoresed on a denaturing 0.8 - 1.0 % agarose / 1.2 M formaldehyde gel, blotted onto Brightstar plus positively-charged nylon membrane (Ambion) according to the manufacturer's protocol and fixed to the membrane using a UV crosslinker (Stratalinker 1800, Stratagene, La Jolla, USA) with 120,000 μJ/ cm^2 of UV irradiation according to the manufacturer's instructions.

Membranes were pre-hybridised in approximately 10 ml PerfectHyb solution (Sigma) at 68 °C for 30 min, after which time 1 μl of DIG-labelled RNA probe was added. The probe was hybridised to the membrane at 68 °C overnight. Membranes were then washed to a final stringency of 0.5 x SSC, 0.1 % SDS at 68 °C. Detection of the presence of the DIG-labelled probe was performed using the DIG Luminescent Detection Kit for Nucleic Acids (Roche). Detection of DIG-labelled probe was performed according to the manufacturer's instructions. Membranes were exposed to hyperfilm ECL (Amersham) at room temperature for varied lengths of time and developed using Kodak GBX developer and fixer.

2.2.5 Cell Culture

2.2.5.1 Thawing of Cells for Cell Culture

Cells frozen in liquid nitrogen storage were thawed rapidly by gentle agitation in a beaker of warm water (approximately 37 °C). Once thawed, the cells were transferred to a 10 ml centrifuge tube. DMEM containing 10 % fetal calf serum (FCS, CSL) was added at a gradually increasing rate, initially dropwise, up to a total volume of 5 ml. An aliquot of the cell suspension was used to quantitate the cell number. The remainder of the suspension was pelleted at 400 x g. The supernatant was discarded and the cells resuspended in a small volume of DMEM containing 10 % FCS. An appropriate amount of this suspension was seeded into cell culture flasks. DMEM containing 10 % FCS was added.

2.2.5.2 Maintenance of Cells in Culture

Cells were cultured at 37 °C in a 5 % CO_2 environment. Media was aspirated and replaced every few days as required. When cells had grown sufficiently to require splitting, media was poured off and cells were washed in PBS. Sufficient volume of 0.25 % trypsin (Sigma) was added to just cover the bottom surface of the flask, and then incubated for 2 - 5 min at 37 °C. The media was then returned to the flask and agitated to remove all cells from the flask surface. An aliquot of the cell suspension was used to quantitate the cell density. The remainder of the suspension was pelleted at 400 x g. The supernatant was discarded and the cells resuspended in a small volume of DMEM containing 10 % FCS. An appropriate amount of this suspension was

seeded into cell culture flasks. DMEM containing 10 % FCS was added and the cells returned to a 37 °C, 5 % CO_2 environment.

2.2.5.3 Freezing of Cells

Cells were frozen for storage by pelleting no more than 10^7 cells and resuspending the pellet in approximately 1 ml of growth medium. To this, 10 % DMSO was added dropwise and with the tube constantly agitating. Cells were stored frozen in liquid nitrogen.

2.2.5.4 Somatic Cell Hybridisation

Murine cell lines were cultured as described in Section 2.2.5.2. Stored patient samples were thawed as described in Section 2.2.5.1. Cell counts were performed on each of the two cell populations. Following counting, $2x10^6$ cells from each population were pooled in a 10 ml tube. A 100 µl aliquot of this pooled population was removed for a no hybridisation growth control and plated in normal growth media (10 % FCS in DMEM). The remaining cells were centrifuged at 1200 rpm for 5 min. The supernatant was removed and the cell pellet was resuspended in 10 ml 37 °C serum free DMEM. This resuspension was then centrifuged at 1200 rpm for 10 min and all media was then carefully aspirated. The pellet was dislodged by flicking the tube and then, very slowly, 0.3 ml 50 % Polyethylene Glycol (PEG) 1500 (Boehringer, Ingelheim, Germany), which had been prewarmed to 37 °C, was added by pipetting down the edge of the tube and swirling the tube to gently resuspend the cells. When resuspension was complete, cells were incubated in PEG for 8 mins. After this time, 8

mls of serum free DMEM, prewarmed to 37°C, was added and mixed into the PEG solution. The mixture was centrifuged at 1200 rpm for 5 mins, and all media was aspirated. The pellet was very gently resuspended in 10 ml 37 °C serum free growth medium. This was centrifuged at 1200 rpm for 5 min. The media was removed and the cell pellet resuspended in 10 ml room temperature 10% FCS DMEM. This was divided evenly between the wells of three 6 well cell culture plates containing 10% FCS growth medium.

Cells were incubated at 37 °C with 5% CO_2 for 24 h. After this recovery period, the media was aspirated and replaced with 10% FCS DMEM containing HAT supplement. The media was thereafter replaced with fresh 10% FCS DMEM containing HAT supplement every three days. Colonies were seen after about three weeks. These were trypsinised within sterile plastic cloning cylinders and removed to separate plates, where they were expanded until they were ready for DNA preparation and PCR screening.

CHAPTER 3

The

(1;3)(p36;q21)

Translocation

CHAPTER 3

t(1;3)(p36;q21)

3.1 t(1;3)(p36;q21) in AML and MDS

3.1.1 Background

The t(1;3)(p36;q21) is a recurrent translocation which occurs in a subset of MDS and myeloid leukaemia patients (Moir *et al* 1984, Bloomfield *et al* 1985, Cambrin *et al* 1986, Welborn *et al* 1987). In AML, the translocation occurs most frequently in FAB subtypes M1 and M4 (Shimizu *et al* 2000). AML patients bearing this translocation have a clinically distinct phenotype. Patients bearing the t(1;3)(p36;q21) typically have a normal or increased platelet count, in contrast to the majority of AML patients, who present with a decreased platelet count (Grigg *et al* 1993). In addition, these patients show morphologic abnormalities of the megakaryocytic, erythroid and granulocytic haematopoietic lineages (known as trilineage dysplasia), with the dysmegakaryopoiesis being the most apparent (Bloomfield *et al* 1985, Najfeld *et al* 1988). Patients with this translocation generally respond poorly to treatment and have a poor prognosis (Najfeld *et al* 1988).

The clinical characteristics of increased platelet count and dysmegakaryopoiesis associated with t(1;3)(p36;q21) are shared by AML or MDS patients with the 3q21q26

syndrome (Bitter *et al* 1985, Pintado *et al* 1985, Fonatsch *et al* 1994, Secker-Walker *et al* 1995, Testoni *et al* 1999). This syndrome consists of translocations t(3;3)(q21;q26) or inv(3)(q21q26), which have similar breakpoints at both 3q21 and 3q26. These chromosomal rearrangements have been shown to result in transcriptional upregulation of the *EVI1* gene located at 3q26 (Morishita *et al* 1992, reviewed in Nucifora 1997).

3.1.1.1 The *MDS1-EVI1* Gene in Leukaemia

The *MDS1-EVI1* gene is encoded by a complex locus. *EVI1* was first identified as a target of upregulation by proviral insertion in murine myeloid leukaemia (Morishita *et al* 1988, Mucenski *et al* 1988). *MDS1* was first identified as a fusion partner of *AML1* in the t(3;21)(q26;q22) in AML (Nucifora *et al* 1994). The same study that identified the *AML1-MDS1* fusion also revealed an *AML1-EVI1* fusion and a complex *AML1-MDS1-EVI1* fusion transcript in the same patient (Nucifora *et al* 1994). The *MDS1* locus has been mapped approximately 200 kb telomeric of the *EVI1* locus at 3q26, and both genes are expressed in normal tissues as independent genes (Fears *et al* 1996). Surprisingly, it was found that *MDS1-EVI1* "fusion" transcripts were expressed in normal individuals, in addition to *EVI1* and *MDS1* transcripts (Fears *et al* 1996). The fusion is formed by a donor splice site from within the *MDS1* coding region joining to the *EVI1* exon 2 acceptor splice site. Until this discovery, researchers had been puzzled that exon 2 of *EVI1* is so highly conserved across species, yet the *EVI1* translation start site is in exon 3. *MDS1-EVI1* contains a PR domain, formed by the distal sequence of *MDS1* and the exon 2 and

proximal exon 3 sequence of *EVI1* (Fears *et al* 1996). PR domains are discussed in Section 3.1.1.2.

The 3q21q26 syndrome occurs in approximately 2% (combined) of all AML FAB subtypes except M3 (Testoni *et al* 1999). Generally, the 3q26 breakpoint is 5' of *EVI1*, between *MDS1* and *EVI1*, in the t(3;3)(q21;q26), and 3' of *EVI1* in the inv(3)(q21q26) (Fears *et al* 1996). The positions of the breakpoints at 3q21 vary quite substantially. There appears to be two breakpoint cluster regions (BCRs), named BCR-T (for telomeric BCR) and BCR-C (for centromeric BCR) separated by approximately 20 kb and spanning a total distance of approximately 100 kb. Breakpoints for each of the t(3;3)(q21;q26), the inv(3)(q21q26) and the t(1;3)(p36;q21) occur in both BCR-T and BCR-C (reviewed in Wieser 2002) (see Figure 3.21). The size of these breakpoint regions is consistent with the upregulation of *MEL1* or *EVI1* by *RPN1* enhancer elements, as it is not unknown for transcriptional enhancer elements to act over such large distances (Cory *et al* 1985, Graham and Adams 1986).

The common understanding of the molecular consequences of these rearrangements is that the *EVI1* gene is transcriptionally upregulated, as a result of juxtaposition to a putative enhancer which normally regulates the expression of the constitutively expressed gene *RPN1* at 3q21 (Suzukawa *et al* 1994). However, this explanation is unsatisfactory for a number of reasons. The most significant of these is that a substantial minority of patients with 3q21q26 syndrome do not express *EVI1* at elevated levels (Soderholm *et al* 1997, Langabeer *et al* 2001). Similarly, significant numbers of patients without 3q26

rearrangements also express *EVI1* at elevated levels, and yet these patients do not display the clinical characteristics of 3q21q26 syndrome (Russell *et al* 1994, Langabeer *et al* 2001). Therefore, *EVI1* expression is neither necessary nor sufficient to establish these clinical features.

3.1.1.2 PRDM Genes in Oncogenesis

The PR domain (PRDM) genes are a family of Kruppel-like zinc finger genes defined by the presence of a PR domain, named for *PRDI-BF1* (*BLIMP1*) and *RIZ*, the first two genes in which the domain was identified. The PR domain is related to the SET (*Su(var)3-9*, *Enhancer-of-zeste*, *Trithorax*) domain, a highly conserved domain which is reportedly involved in chromatin structural regulation *via* a protein methyltransferase function (Rea *et al* 2000). The PR domain has been shown to have a similar function (Huang *et al* 1998, Kouzarides 2002). Almost all members of the PRDM family contain the standard C2-H2 zinc finger DNA-binding motif, whereas most SET domain-containing genes do not have apparent DNA binding motifs. One of the few SET domain-containing genes which does bind DNA is *MLL*, which is located at 11q23 and is involved in normal haematopoietic development (Ernst *et al* 2002). *MLL* is also involved in numerous fusion genes as a result of chromosome translocations in leukaemia (reviewed in Dimartino and Cleary 1999, Anderson *et al* 2001).

A feature of the PRDM genes is their ability to code for two different protein products with apparently opposing functions. This phenomenon is best studied in the 1p36 gene

RIZ. This gene produces two transcripts from distinct promoters. These transcripts are identical in all respects but one: the *RIZ1* transcript contains the PR domain whereas the *RIZ2* transcript does not (Liu *et al* 1997). Both transcripts are expressed in normal cells of all types examined, but in cancerous cells (including breast, liver, bone and colon cancer as well as melanoma and neuroblastoma) the *RIZ1* transcript is downregulated while the *RIZ2* transcript is maintained at a normal level of expression (He *et al* 1998, Jiang *et al* 1999, Steele-Perkins *et al* 2001). Forced expression of *RIZ1* can cause G2/M cell cycle arrest or apoptosis in a number of cancer cell lines, lending support to the idea that it functions as a tumour suppressor (He *et al* 1998, Jiang *et al* 1999, Chadwick *et al* 2000). Mice specifically deficient for *RIZ1* (that is, with normal *RIZ2* expression) are prone to tumour formation, demonstrating a causal link between *RIZ1* inactivation and tumour formation (Steele-Perkins *et al* 2001). *MDS1-EVI1* is another example of this phenomenon. The full-length *MDS1-EVI1* transcript does contain a PR domain, whereas the *EVI1* transcript does not. Although the different functions of the variants are debated (Morishita *et al* 1995, Wimmer *et al* 1998, Izutsu *et al* 2001), the general consensus is that MDS1-EVI1 is an activator of transcription while EVI1 is a repressor (Nucifora 1997).

3.1.2 Patient Information

Patient 1 was a 73 year old woman who presented in December 1985 with pancytopenia. Her platelet count was 77 x 10^9/L, and bone marrow examination led to a diagnosis of refractory anemia with ringed sideroblasts. The patient remained asymptomatic until November 1987, when the platelet count had increased (116 x 10^9/L). By March 1988, the

platelet count was 206 x 10^9/L and bone marrow aspiration diagnosed AML M1 with 60% blasts, hypolobular megakaryocytes and no Auer rods or cytoplasmic granulation. The patient remained voluntarily untreated until her death in August 1989.

Patient 1 was the subject of a previously published study (Marsden *et al* 1992). This study was not concerned with breakpoint characterisation but rather with the loss of expression of ABO antigens during the course of the leukaemia. No work on the molecular characterisation of the translocation carried by Patient 1 was performed prior to the onset of the present study.

Cytogenetic study of Patient 1 bone marrow material was first performed in March 1988. Examination of 21 cells showed 2 of normal karyotype, 13 showing 46, XX, t(1;3)(p36;q21) and 6 that showed 46, XX, t(1;3)(p36;q21), t(14;17)(q32;q21). Cytogenetic analysis was repeated in April 1989, when all 15 cells examined showed 46, XX, t(1;3)(p36;q21), t(14;17)(q32;q21). This more complex karyotype is shown in Figure 3.1.

3.2 Somatic Cell Hybrid Analysis

Somatic cell hybrid lines were constructed by fusion of the PG19 murine melanoma cell line with bone marrow cells from Patient 1 (Dobrovic, unpublished). Many colonies were obtained, and twenty-eight of these from different plates were isolated and cultured.

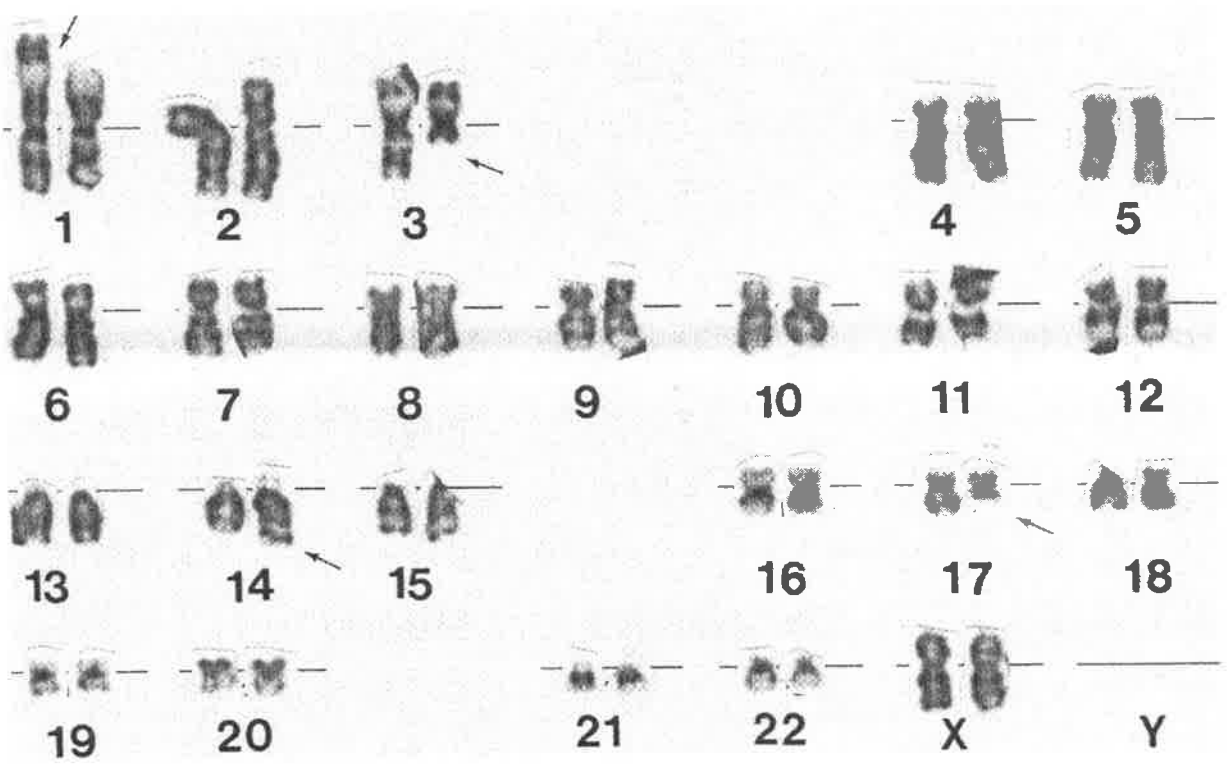


Figure 3.1. Patient 1 Karyotype.

The aberrant karyotype of the leukemic blasts of Patient 1, showing the t(1;3)(p36;q21) which is the subject of investigation in this chapter, and also a t(14;17)(q32;q21). Aberrant chromosomes are arrowed. Reproduced from Marsden *et al* (1992).

Genomic DNA was prepared from each hybrid cell line in order to assess their human chromosome content.

3.2.1 Chromosome 1 Breakpoint Analysis

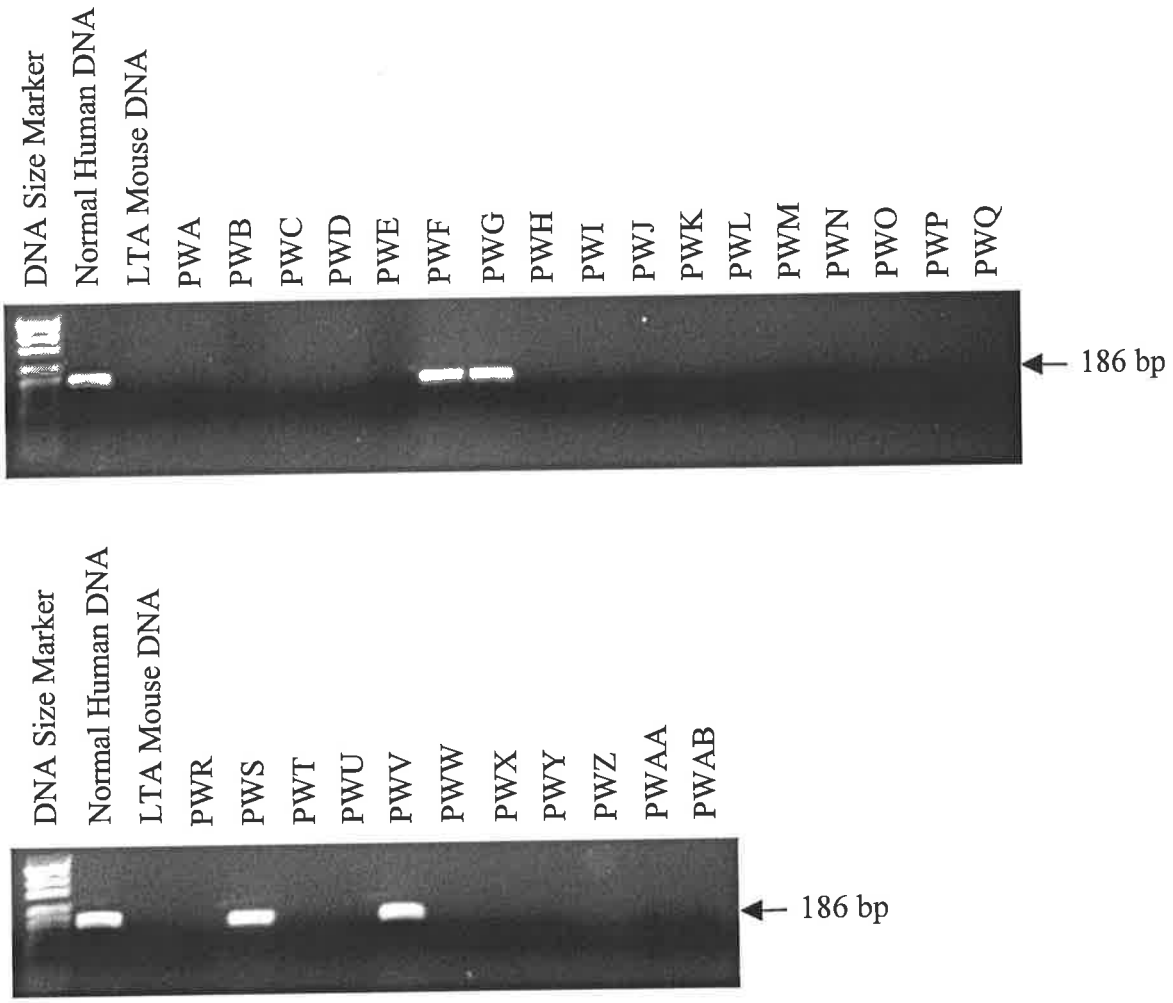
To identify a hybrid cell line which retained one of the der 1 and der 3 translocated chromosomes and neither of the normal chromosome 1 or normal chromosome 3, initial screening of the hybrid cell lines utilised 1p36 markers designed to regions flanking the genes *TP73* (telomeric marker) and *TNFR2* (centromeric marker). Specific PCR for each of these markers was performed on each of the 28 hybrid cell lines. This screen produced marker patterns which apparently indicated only hybrids which retained a normal chromosome 1 (positive for both markers) or no chromosome 1 material (negative for both markers) (see Figure 3.2).

Soon after this result, the publication of a study by Shimizu *et al* (2000) reported the localisation of the t(1;3)(p36;q21) breakpoints in their four patients to within a 90 kb region at 1p36.3. Although *TP73*, which was used as the telomeric marker in the initial screen performed on Patient 1 somatic cell hybrids, is also at 1p36.3, the screen could not confidently have been said to include the region described in the report by Shimizu *et al* (2000). This report included oligonucleotide primer sequences which were designed to amplify a portion of a 1p36 region which, when used as a probe in Southern hybridisation experiments on material derived from patients who carried the t(1;3)(p36;q21), was shown

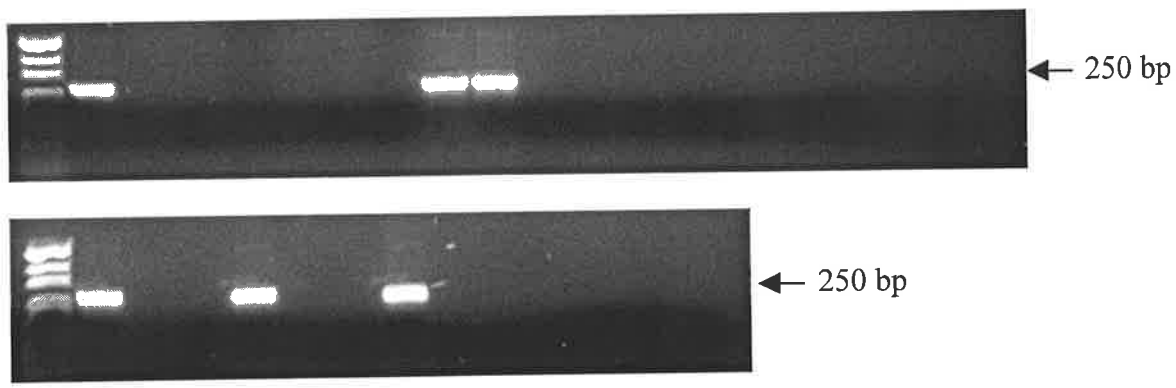
Figure 3.2. Chromosome 1 Marker Screen of PW Hybrid Cell Lines.

Results of testing for presence or absence of chromosome 1 markers in the panel of somatic cell hybrids generated from Patient 1 bone marrow material. *TP73* is at the telomeric end of 1p36. *TNFR2* is at the centromeric boundary of 1p36 and therefore centromeric of the presumed breakpoint. Lanes are loaded as indicated; the gel in panel B is loaded in the same order as the gel in panel A. The DNA size marker used was pUC19/*Hpa*II. A sample of DNA from a normal human individual was used as a positive control. A sample of DNA from the murine LTA cell line was used as a negative control. Hybrids are named PWA - PWAB. The expected size of the product is indicated on the right. Only four of the hybrids, PWF, PWG, PWS and PWV, appear to retain any chromosome 1 material, and all appear to retain the entire normal chromosome 1. They are therefore presumed not informative for chromosome 1 PCR walking.

A) *TP73* Screen



B) *TNFR2* Screen



to be telomeric to the breakpoint in all cases (Shimizu *et al* 2000). This region was therefore considered a useful telomeric locus to use to rescreen the 28 hybrid cell lines.

A search of the Genbank database using the BLASTN algorithm with the primer sequences reported by Shimizu *et al* (2000) identified a 184 kb clone designated RP11-425F18 (accession AC020709.3). At the time RP11-425F18 was not fully sequenced and was described in three unordered fragments. Primer pairs were designed to either end of each of these fragments, and one of these pairs, 425F3Z, was used as the telomeric marker for a repeat screening of the hybrids. The 425F3Z results were correlated with the results of the previous screen. The four hybrids which were found to contain chromosome 1 material in the previous screen, PWF, PWG, PWS and PWV, were also positive for 425F3Z. This indicates that all four hybrids do indeed contain a normal chromosome 1. This screen also identified a hybrid line, PWN, which retained 425F3Z but was negative for the markers *TP73* and *TNFR2* (see Figure 3.3). This pattern indicates that PWN contained the translocated der 3 chromosome, containing a short telomeric fragment of chromosome 1, and that the breakpoint lay between *TP73* and 425F3Z. Analysis of the PWN hybrid with the remaining markers designed to either end of each of the three fragments of RP11-425F18 showed that each of these markers was positive, indicating that the breakpoint lay outside of, and centromeric to, this clone.

Shortly after this time the public access Celera database became available, and searches of this database using the BLASTN algorithm revealed a genomic scaffold named GA_x2HTBKR38EG which contained the marker 425F3Z but none of the other five

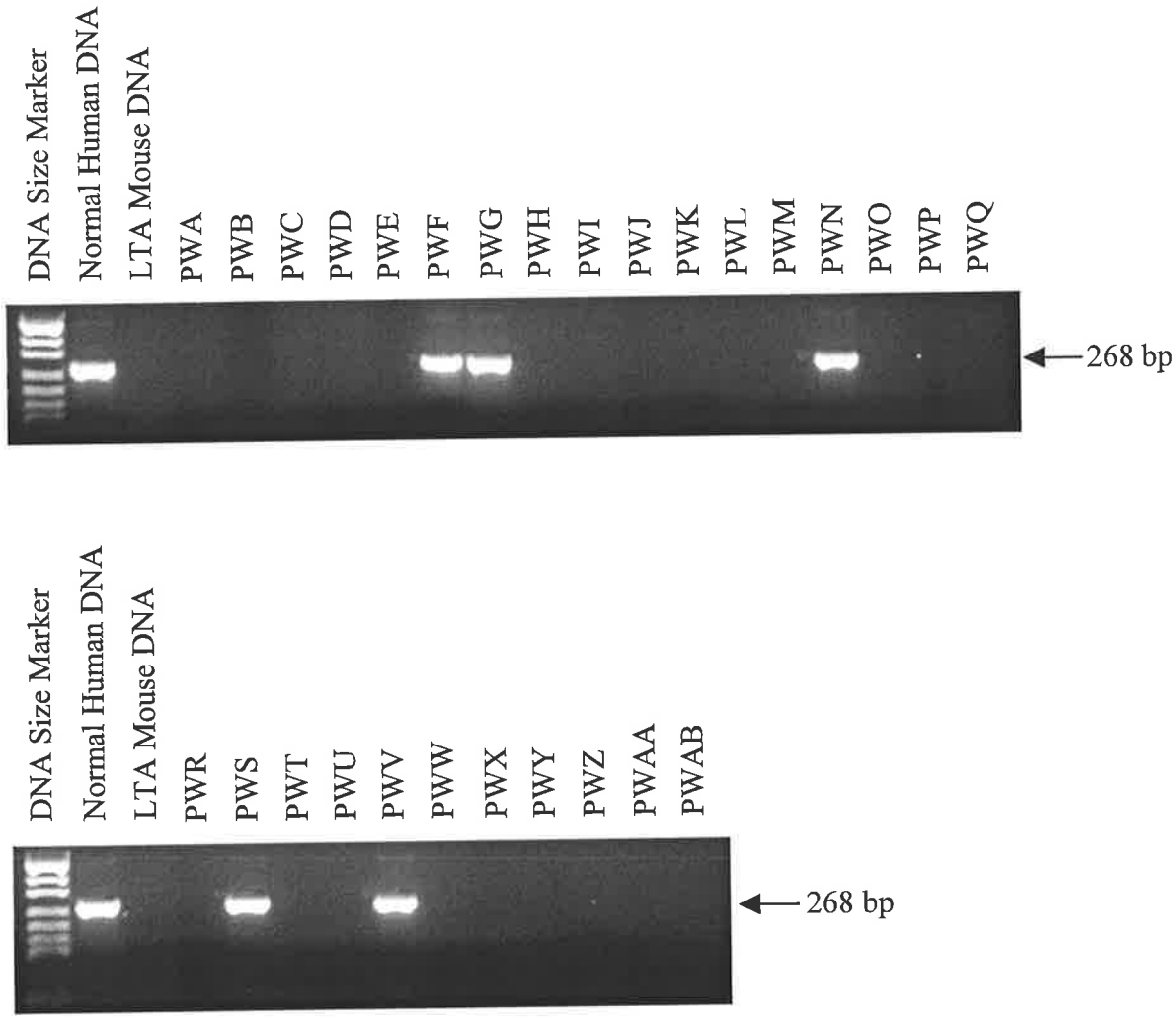


Figure 3.3. 425F3Z Screen of PW Hybrid Cell Lines.

Results of testing for presence or absence of chromosome 1 telomeric marker 425F3Z in the panel of somatic cell hybrids generated from Patient 1 bone marrow material. 425F3Z is further telomeric than *TP73* and so is a better telomeric marker. Gels are loaded in the same order as the gels in Figure 3.2. The four hybrids which were thought to contain normal chromosome 1, PWF, PWG, PWS and PWV (see Figure 3.2), retain 425F3Z also, in agreement with their retention of a complete chromosome 1. PWN, previously thought not to contain any chromosome 1 material, is positive for 425F3Z, indicating that it contains the telomeric portion of chromosome 1 and therefore is likely to contain the der 3 translocated chromosome. PWN is therefore informative for chromosome 1 PCR walking.

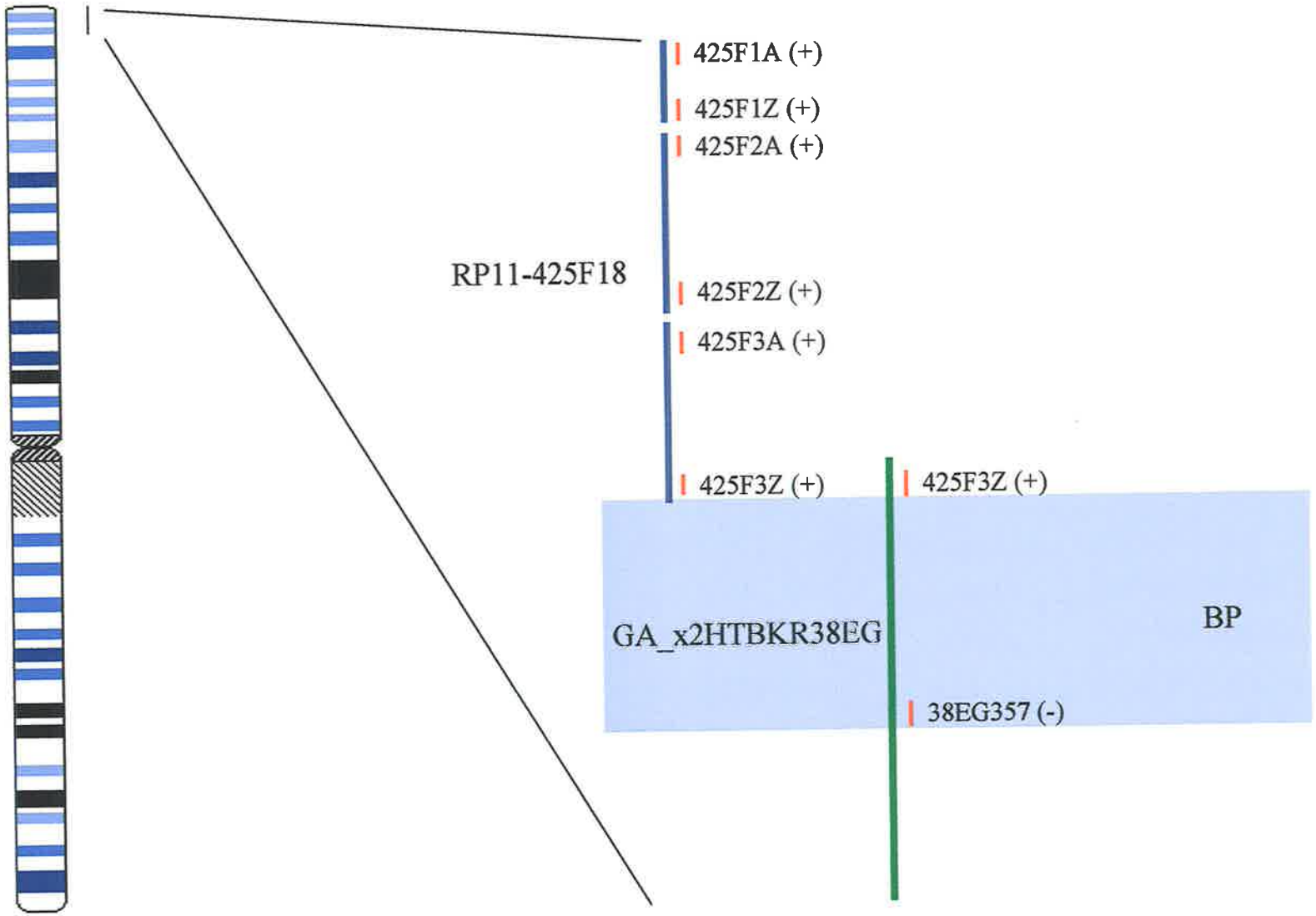
markers designed to RP11-425F18. As both RP11-425F18 and GA_x2HTBKR38EG contained 425F3Z, they were known to overlap, but their orientation with respect to the telomere was not known at this time.

To address this, a marker named 38EG357 was designed to the scaffold at a position 348 kb from 425F3Z. When this marker was tested for retention in the PWN hybrid cell line, it was found that the marker had been lost, demonstrating that the breakpoint lay between 425F3Z and 38EG357, and therefore was positioned on GA_x2HTBKR38EG. This also demonstrated the orientation of the RP11-425F18 clone and the GA_x2HTBKR38EG scaffold with respect to the telomere, as the positive marker must be on the telomeric side of the breakpoint (see Figure 3.4). Further testing with other markers designed to GA_x2HTBKR38EG (see Figure 3.5) on both sides of the breakpoint reaffirmed that the pattern of markers was consistent with the hybrid containing the der 3 chromosome and also narrowed the breakpoint region to between two markers located within 1 kb of each other, 38EG211#1 (positive, telomeric marker) and 38EG211#2 (negative, centromeric marker).

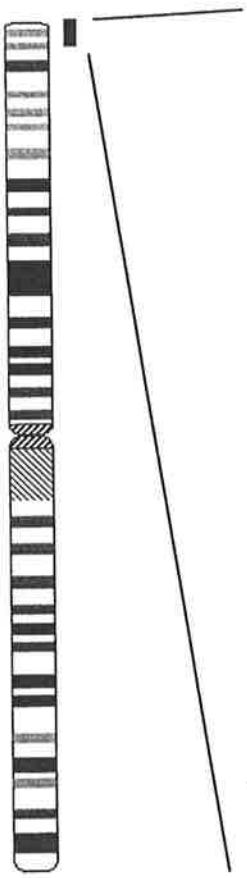
During the progress of the research described thus far in this chapter, the molecular characterisation of four t(1;3)(p36;q21) breakpoints was published (Mochizuki *et al* 2000). This report suggested that the t(1;3)(p36;q21) acted to upregulate a novel gene found on 1p36, *MDS1-EVI1-Like-1* (*MEL1*), by juxtaposing it with the transcriptional enhancers of *RPN1* in a similar way to that in which *EVI1* is upregulated by the t(3;3)(q21;q26) or inv(3)(q21q26). There was, however, no accompanying deposition of the nucleotide

Figure 3.4 Orientation of RP11-425F18 and GA_x2HTBKR38EG

The figure illustrates the portion of 1p36 covered by the NCBI BAC clone RP11-425F18 (shown in blue) and the Celera genomic scaffold GA_x2HTBKR38EG (shown in green). The positions of PCR markers used in somatic cell hybrid analysis are shown in red. Alongside the name of each marker is a + (indicating presence of the marker in the hybrid cell line PWN, and therefore presence of the marker on the der 3 translocated chromosome) or a - (indicating absence of the marker from PWN and the der 3 chromosome). The orientation of the clone and the scaffold with respect to each other was determined by the fact that 425F3Z was the only marker that the two shared in common, but their orientation with respect to the telomere was not known until the marker 38EG357 was determined to be absent from PWN. Because PWN contains the der 3 translocated chromosome, the negative markers from chromosome 1 are more centromeric than the positive markers. As 38EG357 is absent from PWN and 425F3Z is present, the translocation breakpoint (BP) must lie between these two markers on GA_ x2HTBKR38EG. This region is shown as a pale blue box. Note that the region covered by the marker 425F3Z is not included in this region as it is a positive marker and therefore the entire PCR product is known to be intact within PWN. 38EG357 is included in the breakpoint region, however, as it is absent from PWN and the breakpoint could possibly occur within the PCR product. The figure is not to scale.


425F1A (+)
425F1Z (+)
425F2A (+)
RP11-425F18
425F2Z (+)
425F3A (+)
425F3Z (+)
425F3Z (+)
GA_x2HTBKR38EG
BP
38EG357 (-)

sequence of the *MEL1* transcript into Genbank (AB078876 was not deposited until January 2002). This made it difficult to position *MEL1* with respect to the breakpoint position already identified in Patient 1.



Marker	Position on Scaffold (kb)	Presence in PWN
425F3Z	9	+
333E3	198	+
38EG203	203	+
38EG210	210	+
38EG211#1	211	+
38EG211#2	211.5	-
38EG212	212	-
38EG213	213	-
38EG217	217	-
38EG239	239	-
38EG357	357	-

Figure 3.5. Markers used in Chromosome 1 Breakpoint Analysis of PWN. The table shows the location of each marker on the genomic scaffold GA_x2HTBKR38EG and the result obtained on analysis of PWN for each marker. + indicates that the PWN retained the marker. - indicates the marker was not present in PWN. The region displayed covers the breakpoint region illustrated in Figure 3.4.

The Mochizuki *et al* (2000) report did include a protein sequence alignment, including a protein sequence for MEL1, in their Figure 2A. *MEL1* was reported to be a novel member of the transcription factor family of PR-domain genes (PRDM), and a search in the Genbank database for known PRDM genes produced a transcript for a gene known as *PRDM16* (NM_022114) which mapped to 1p36.3. A comparison of the protein sequences of PRDM16 (NP_071397) and that reported by Mochizuki *et al* (2000) for MEL1 showed that they were almost identical, notably so in the less-conserved regions where the two proteins differed from the MDS1-EVI1 protein (see Figure 3.6). The major variation between the reported MEL1 protein sequence and NP_071397 was the presence of an extra 19 amino acids in NP_071397, which was later determined by this study to be due to an mRNA splice variant (see Section 3.4.1.1).

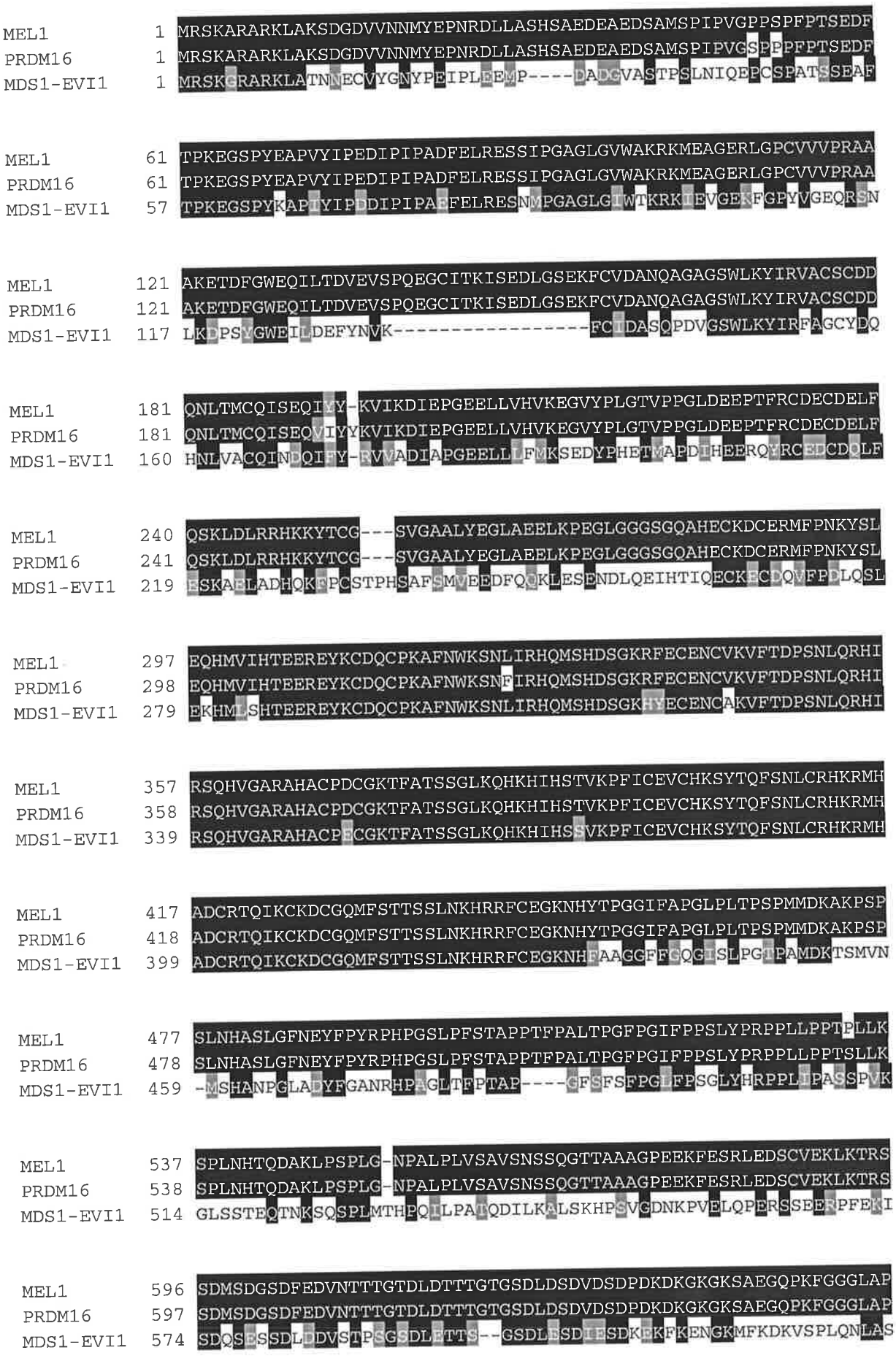
The use of BLASTN analysis of the nucleotide sequence of *PRDM16* against the public Celera database allowed the positioning of the start of the *PRDM16* transcript within the Celera genomic scaffold GA_x2HTBKR38EG. This was the same scaffold to which the Patient 1 1p36 breakpoint had already been localised, and so a strong indication that *PRDM16* was *MEL1*, and also that the translocation carried by Patient 1 was the same as those that had been mapped in the study by Mochizuki *et al* (2000). *PRDM16* and *MEL1* were later confirmed by further database entries (XM_010556) to be the same gene. The positional relationship between the chromosome, contigs, genes and markers is shown in Figure 3.7.

Figure 3.6. Alignment of MEL1 Protein Sequence with PRDM16 and MDS1-EVI1 Protein Sequences

This alignment was created with the multiple sequence alignment algorithm CLUSTALW, available through http://www2.ebi.ac.uk/clustalw/, and formatted using the Boxshade server at http://www.ch.embnet.org/software/BOX_form.html.

The input sequences were: MEL1 (as reported in Mochizuki *et al* (2000) Figure 2A); PRDM16 (NP_071397) and MDS1-EVI1 (constructed from an *MDS1* mRNA sequence, accession NM_004991, and an *AML1 -EVI1* fusion mRNA sequence, accession S69002, which was then translated using the ExPaSy Translate tool at http://kr.expasy.org/tools/dna.html).

It is clear from the alignment that the MEL1 and PRDM16 sequences are much more closely related to each other than either is to MDS1-EVI1. The degree of similarity was high enough to proceed under the assumption that PRDM16 was in fact MEL1, an assumption that was later confirmed by new database entries (XM_010556). The major difference between the two is a string of 19 amino acids present in PRDM16 (and MDS1-EVI1) but absent from MEL1 (PRDM aa 1233-1251). This sequence discrepancy was later resolved; see Section 3.4.1.1 for details.


MEL1 1 MRSKARARKLAKSDGDVVNNMYEPNRDLLASHSAEDEAEDSAMSPIPVGPPSPFPTSEDF
PRDM16 1 MRSKARARKLAKSDGDVVNNMYEPNRDLLASHSAEDEAEDSAMSPIPVGSPPPFPTSEDF
MDS1-EVI1 1 MRSKGRARKLATNNECVYGNYPEIPLEEMP----DADGVASTPSLNIQEPCSPATSSEAF
MEL1 61 TPKEGSPYEAPVYIPEDIPIPADFELRESSIPGAGLGVWAKRKMEAGERLGPCVVVPRAA
PRDM16 61 TPKEGSPYEAPVYIPEDIPIPADFELRESSIPGAGLGVWAKRKMEAGERLGPCVVVPRAA
MDS1-EVI1 57 TPKEGSPYKAPIYIPDDIPIPAEFELRESNMPGAGLGIWTKRKIEVGEKFGPYVGEQRSN
MEL1 121 AKETDFGWEQILTDVEVSPQEGCITKISEDLGSEKFCVDANQAGAGSWLKYIRVACSCDD
PRDM16 121 AKETDFGWEQILTDVEVSPQEGCITKISEDLGSEKFCVDANQAGAGSWLKYIRVACSCDD
MDS1-EVI1 117 LKDPSYGWEILDEFYNVK----------FCIDASQPDVGSWLKYIRFAGCYDQ
MEL1 181 QNLTMCQISEQIYY-KVIKDIEPGEELLVHVKEGVYPLGTVPPGLDEEPTFRCDECDELF
PRDM16 181 QNLTMCQISEQVIYYKVIKDIEPGEELLVHVKEGVYPLGTVPPGLDEEPTFRCDECDELF
MDS1-EVI1 160 HNLVACQINDQIFY-RVVADIAPGEELLLFMKSEDYPHETMAPDIHEERQYRCEDCDQLF
MEL1 240 QSKLDLRRHKKYTCG---SVGAALYEGLAEELKPEGLGGGSGQAHECKDCERMFPNKYSL
PRDM16 241 QSKLDLRRHKKYTCG---SVGAALYEGLAEELKPEGLGGGSGQAHECKDCERMFPNKYSL
MDS1-EVI1 219 ESKAELADHQKTPCSTPHSAFSMVEEDFQQKLESENDLQEIHTIQECKECDQVFPDLQSL
MEL1 297 EQHMVIHTEEREYKCDQCPKAFNWKSNLIRHQMSHDSGKRFECENCVKVFTDPSNLQRHI
PRDM16 298 EQHMVIHTEEREYKCDQCPKAFNWKSNFIRHQMSHDSGKRFECENCVKVFTDPSNLQRHI
MDS1-EVI1 279 EKHMLSHTEEREYKCDQCPKAFNWKSNLIRHQMSHDSGKHYECENCAKVFTDPSNLQRHI
MEL1 357 RSQHVGARAHACPDCGKTFATSSGLKQHKHIHSTVKPFICEVCHKSYTQFSNLCRHKRMH
PRDM16 358 RSQHVGARAHACPDCGKTFATSSGLKQHKHIHSTVKPFICEVCHKSYTQFSNLCRHKRMH
MDS1-EVI1 339 RSQHVGARAHACPECGKTFATSSGLKQHKHIHSSVKPFICEVCHKSYTQFSNLCRHKRMH
MEL1 417 ADCRTQIKCKDCGQMFSTTSSLNKHRRFCEGKNHYTPGGIFAPGLPLTPSPMMDKAKPSP
PRDM16 418 ADCRTQIKCKDCGQMFSTTSSLNKHRRFCEGKNHYTPGGIFAPGLPLTPSPMMDKAKPSP
MDS1-EVI1 399 ADCRTQIKCKDCGQMFSTTSSLNKHRRFCEGKNHFAAGGFFGQGISLPGTPAMDKTSMVN
MEL1 477 SLNHASLGFNEYFPYRPHPGSLPFSTAPPTFPALTPGFPGIFPPSLYPRPPLLPPTPLLK
PRDM16 478 SLNHASLGFNEYFPYRPHPGSLPFSTAPPTFPALTPGFPGIFPPSLYPRPPLLPPTSLLK
MDS1-EVI1 459 -MSHANPGLADYFGANRHPAGLTFPTAP----GFSFSFPGLFPSGLYHRPPLIPASSPVK
MEL1 537 SPLNHTQDAKLPSPLG-NPALPLVSAVSNSSQGTTAAAGPEEKFESRLEDSCVEKLKTRS
PRDM16 538 SPLNHTQDAKLPSPLG-NPALPLVSAVSNSSQGTTAAAGPEEKFESRLEDSCVEKLKTRS
MDS1-EVI1 514 GLSSTEQTNKSQSPLMTHPQILPATQDILKALSKHPSVGDNKPVELQPERSSEERPFEKI
MEL1 596 SDMSDGSDFEDVNTTTGTDLDTTTGTGSDLDSDVDSDPDKDKGKGKSAEGQPKFGGGLAP
PRDM16 597 SDMSDGSDFEDVNTTTGTDLDTTTGTGSDLDSDVDSDPDKDKGKGKSAEGQPKFGGGLAP
MDS1-EVI1 574 SDQSESSDLDDVSTPSGSDLETTS--GSDLESDIESDKEKFKENGKMFKDKVSPLQNLAS

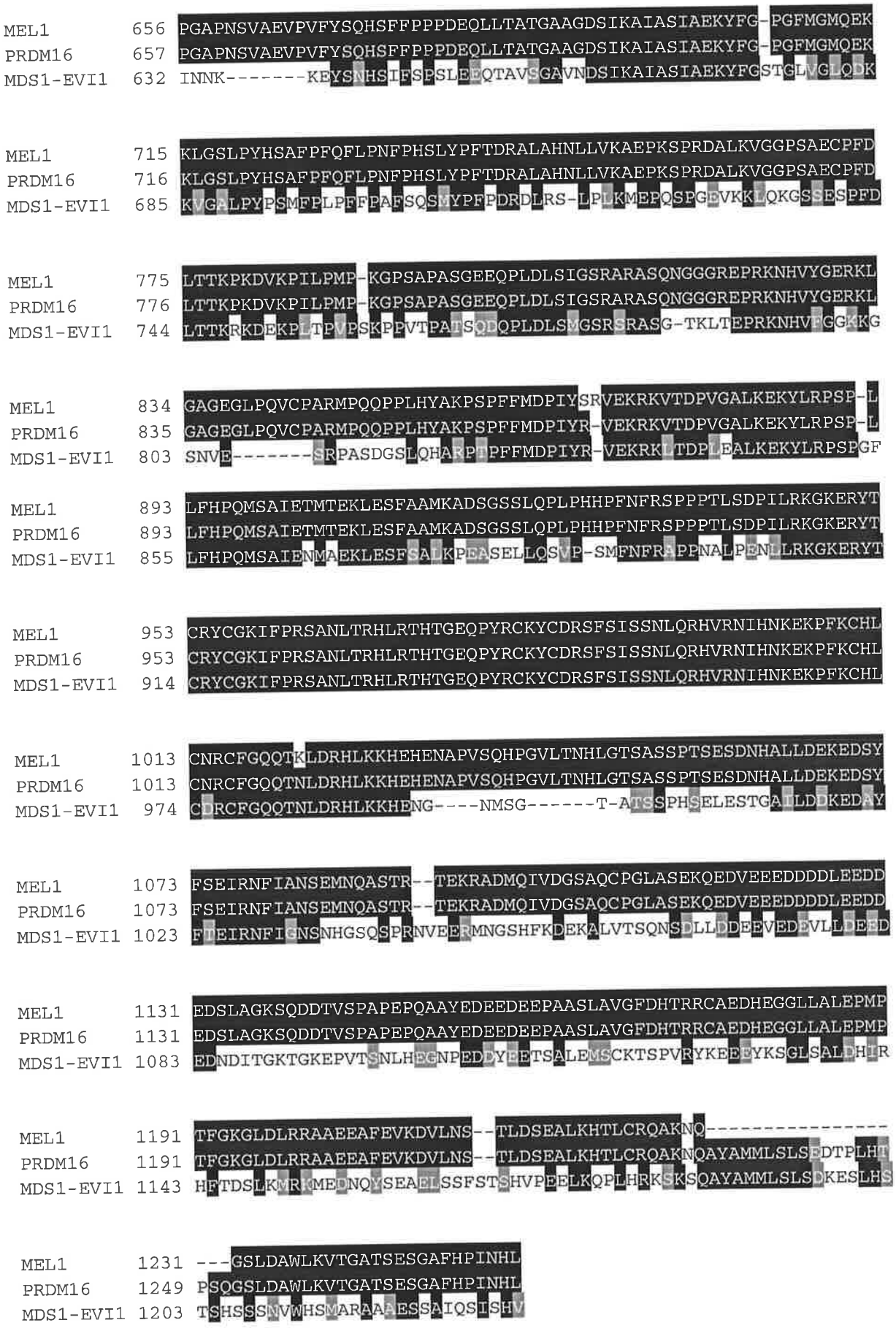
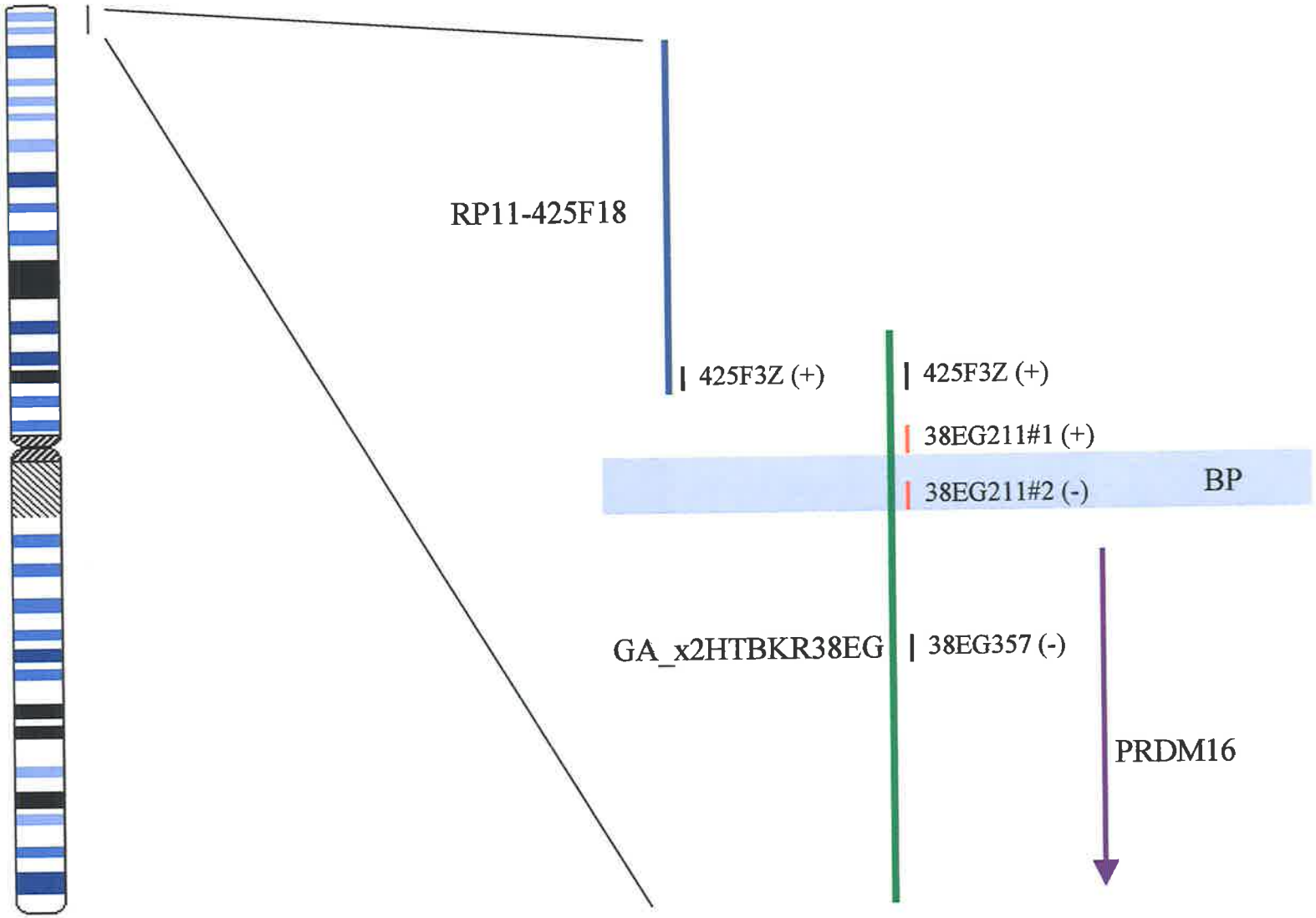

MEL1 656 PGAPNSVAEVPVFYSQHSFFPPPDEQLLTATGAAGDSIKAIASIAEKYFG-PGFMGMQEK
PRDM16 657 PGAPNSVAEVPVFYSQHSFFPPPDEQLLTATGAAGDSIKAIASIAEKYFG-PGFMGMQEK
MDS1-EVI1 632 INNK-------KEYSNHSIFSPSLEEQTAVSGAVNDSIKAIASIAEKYFGSTGLVGLQDK
MEL1 715 KLGSLPYHSAFPFQFLPNFPHSLYPFTDRALAHNLLVKAEPKSPRDALKVGGPSAECPFD
PRDM16 716 KLGSLPYHSAFPFQFLPNFPHSLYPFTDRALAHNLLVKAEPKSPRDALKVGGPSAECPFD
MDS1-EVI1 685 KVGALPYPSMFPLPFFPAFSQSMYPFPDRDLRS-LPLKMEPQSPGEVKKLQKGSSESPFD
MEL1 775 LTTKPKDVKPILPMP-KGPSAPASGEEQPLDLSIGSRARASQNGGGREPRKNHVYGERKL
PRDM16 776 LTTKPKDVKPILPMP-KGPSAPASGEEQPLDLSIGSRARASQNGGGREPRKNHVYGERKL
MDS1-EVI1 744 LTTKRKDEKPLTPVPSKPPVTPATSQDQPLDLSMGSRSRASG-TKLTEPRKNHVFGGKKG
MEL1 834 GAGEGLPQVCPARMPQQPPLHYAKPSPFFMDPIYSRVEKRKVTDPVGALKEKYLRPSP-L
PRDM16 835 GAGEGLPQVCPARMPQQPPLHYAKPSPFFMDPIYR-VEKRKVTDPVGALKEKYLRPSP-L
MDS1-EVI1 803 SNVE-------SRPASDGSLQHARPTPFFMDPIYR-VEKRKLTDPLEALKEKYLRPSPGF
MEL1 893 LFHPQMSAIETMTEKLESFAAMKADSGSSLQPLPHHPFNFRSPPPTLSDPILRKGKERYT
PRDM16 893 LFHPQMSAIETMTEKLESFAAMKADSGSSLQPLPHHPFNFRSPPPTLSDPILRKGKERYT
MDS1-EVI1 855 LFHPQMSAIENMAEKLESFSALKPEASELLQSVP-SMFNFRAPPNALPENLLRKGKERYT
MEL1 953 CRYCGKIFPRSANLTRHLRTHTGEQPYRCKYCDRSFSISSNLQRHVRNIHNKEKPFKCHL
PRDM16 953 CRYCGKIFPRSANLTRHLRTHTGEQPYRCKYCDRSFSISSNLQRHVRNIHNKEKPFKCHL
MDS1-EVI1 914 CRYCGKIFPRSANLTRHLRTHTGEQPYRCKYCDRSFSISSNLQRHVRNIHNKEKPFKCHL
MEL1 1013 CNRCFGQQTKLDRHLKKHEHENAPVSQHPGVLTNHLGTSASSPTSESDNHALLDEKEDSY
PRDM16 1013 CNRCFGQQTNLDRHLKKHEHENAPVSQHPGVLTNHLGTSASSPTSESDNHALLDEKEDSY
MDS1-EVI1 974 CDRCFGQQTNLDRHLKKHENG----NMSG------T-ATSSPHSELESTGAILDDKEDAY
MEL1 1073 FSEIRNFIANSEMNQASTR--TEKRADMQIVDGSAQCPGLASEKQEDVEEEDDDDLEEDD
PRDM16 1073 FSEIRNFIANSEMNQASTR--TEKRADMQIVDGSAQCPGLASEKQEDVEEEDDDDLEEDD
MDS1-EVI1 1023 FTEIRNFIGNSNHGSQSPRNVEERMNGSHFKDEKALVTSQNSDLLDDEEVEDEVLLDEED
MEL1 1131 EDSLAGKSQDDTVSPAPEPQAAYEDEEDEEPAASLAVGFDHTRRCAEDHEGGLLALEPMP
PRDM16 1131 EDSLAGKSQDDTVSPAPEPQAAYEDEEDEEPAASLAVGFDHTRRCAEDHEGGLLALEPMP
MDS1-EVI1 1083 EDNDITGKTGKEPVTSNLHEGNPEDDYEETSALEMSCKTSPVRYKEEEYKSGLSALDHIR
MEL1 1191 TFGKGLDLRRAAEEAFEVKDVLNS--TLDSEALKHTLCRQAKNQ----------
PRDM16 1191 TFGKGLDLRRAAEEAFEVKDVLNS--TLDSEALKHTLCRQAKNQAYAMMLSLSEDTPLHT
MDS1-EVI1 1143 HFTDSLKMRKMEDNQYSEAELSSFSTSHVPEELKQPLHRKSKSQAYAMMLSLSDKESLHS
MEL1 1231 ---GSLDAWLKVTGATSESGAFHPINHL
PRDM16 1249 PSQGSLDAWLKVTGATSESGAFHPINHL
MDS1-EVI1 1203 TSHSSSNVWHSMARAAAESSAIQSISHV

Figure 3.7. Orientation of RP11-425F18 and GA_x2HTBKR38EG.

The figure illustrates the position of the *PRDM16* or *MEL1* gene within the portion of 1p36 covered by the NCBI BAC clone RP11-425F18 (shown in blue) and the Celera genomic scaffold GA_x2HTBKR38EG (shown in green). The positions of the PCR markers 425F3Z and 38EG357 are shown in gray, and the positions of the PCR markers 38EG211#1 and 38EG211#2, which define the breakpoint region, are shown in red. The breakpoint region itself is highlighted as a pale blue box. Alongside the name of each marker is a + (indicating presence of the marker in the hybrid cell line PWN, and therefore presence of the marker on the der 3 translocated chromosome) or a - (indicating absence of the marker from PWN and the der 3 chromosome). The position and orientation of the *PRDM16* mRNA sequence, illustrated as a purple arrow, was determined with respect to GA_x2HTBKR38EG by BLASTN analysis. The breakpoint region defined by somatic cell hybrid analysis is approximately 43 kb 5' of the transcription start site of the gene. This is in accordance with expectations, based on the results of Mochizuki *et al* (2000). The figure is not to scale.

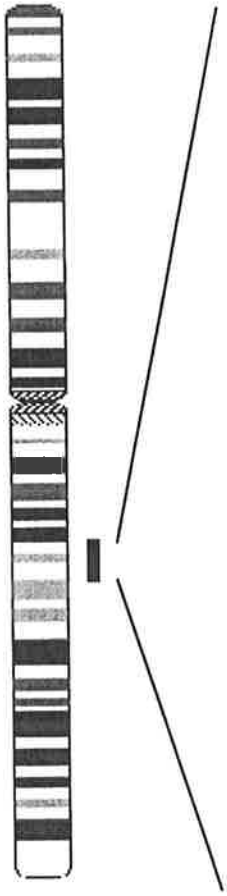

RP11-425F18
425F3Z (+)
425F3Z (+)
38EG211#1 (+)
38EG211#2 (-)
BP
GA_x2HTBKR38EG
38EG357 (-)
PRDM16

3.2.2 Chromosome 3 Breakpoint Analysis

Mochizuki *et al* (2000) also described the chromosome 3 breakpoint in the t(1;3)(p36;q21) as being approximately 90 to 130 kb downstream (centromeric) of the *RPNI* locus in their four patients. This information provided a good basis for localising the chromosome 3 breakpoint in Patient 1. Using the BLASTN algorithm and *RPN1* nucleotide sequence NM_002950 to search the public Celera database produced a genomic scaffold (GA_x2HTBL4DBVY) containing the 3' end of *RPN1* and 370 kb of downstream sequence. Primers were designed to a series of markers along this 370 kb region, and these were tested against the PWN hybrid to determine their position with respect to the breakpoint (see Figure 3.8). The pattern of marker retention and loss was consistent with the hybrid containing a der 3 translocated chromosome as expected. This process located the chromosome 3 breakpoint in Patient 1 between the markers DBVY253 (positive, centromeric marker) and DBVY255 (negative, telomeric marker). These markers are separated by a distance of 2 kb, and DBVY255, the closer of the two, is 123 kb downstream from the 3' end of *RPN1*.

3.3 Amplification and Sequencing of Breakpoints

To determine the sequence at the breakpoints, it was first necessary to amplify across the breakpoints. The identification of positive markers (that is, markers retained by the hybrid and therefore on the der 3 chromosome) at the chromosome 1 and chromosome 3



Marker	Position on Scaffold (kb)	Presence of PWN
3'*GATA2*	238	+
DBVY250	250	+
DBVY251	251	+
DBVY252	252	+
DBVY253	253	+
DBVY255	255	-
DBVY260	260	-
DBVY270	270	-
DBVY280	279	-
DBVY290	288	-
GR6	333	-
RPN1	378	-

Figure 3.8. Markers used in Chromosome 3 Breakpoint Analysis of PWN. The table shows the location of each marker on the genomic scaffold GA_x2HTBL4DBVY and the result obtained on analysis of PWN for each marker. + indicates that the PWN retained the marker. - indicates the marker was not present in PWN.

breakpoints made it possible to use the primers from those markers to amplify across the der 3 breakpoint. Also, the identification of negative markers at both breakpoints enabled the amplification of the der 1 breakpoint, assuming that there was no loss of material

associated with the translocation and that, therefore, all markers which were absent from the hybrid containing the der 3 chromosome were in fact present on the der 1 chromosome. Long distance PCR (Section 2.2.3.2) was performed on genomic DNA from Patient 1 and also from a normal human control, using the primers in combinations that would produce either a normal chromosome 1 and chromosome 3 product, or a translocated der 1 or der 3 product. The position and orientation of the markers and primers with respect to the telomere, centromere and breakpoint are illustrated in Figure 3.9. The results of this PCR are shown in Figure 3.10.

As illustrated in Figure 3.10, both the der 1 and der 3 breakpoints were successfully amplified, as were both normal chromosome products. The products amplified from the translocated chromosomes were sequenced and the sequence of both breakpoints, along with the Genbank sequences from 1p36 and 3q21 with which the breakpoints align, can be seen in Figure 3.11. As can be seen from this sequence alignment, chromosome 1 material is lost during the translocation and chromosome 3 material is duplicated. Using the numbering presented in Figure 3.11, chromosome 1 sequence corresponding to clone RP11-333E3 base pairs 30857 - 30891 (35 bp in total) is lost during translocation. Chromosome 3 sequence corresponding to RP11-475N22 base pairs 61280 - 61477 (198 bp in total) is duplicated during translocation. There is some uncertainty in these figures as the RP11-333E3 and RP11-475N22 sequences at the der 1 breakpoint overlap by 2 bp, obscuring the precise position of the original breaks.

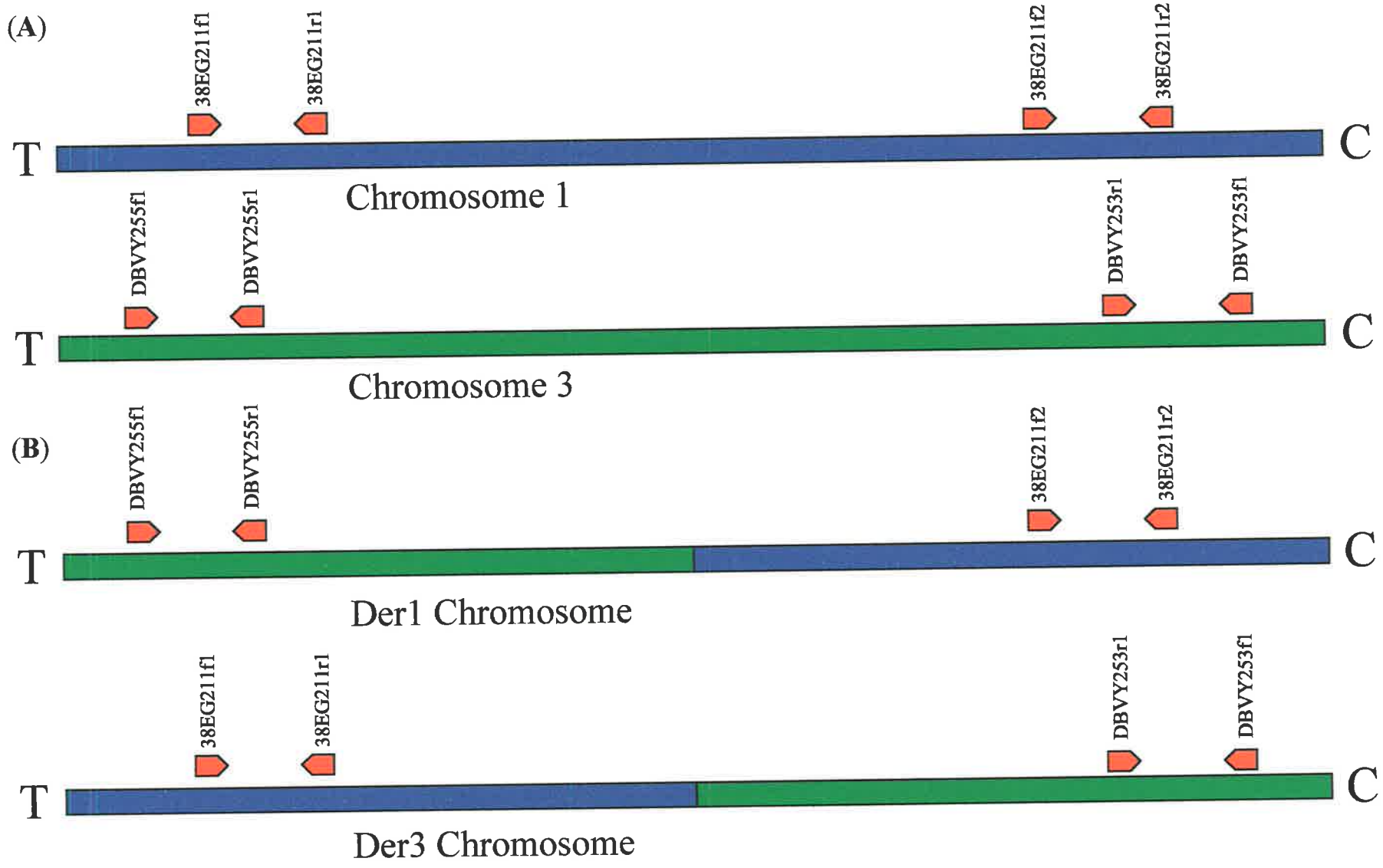
Figure 3.9. Primer combinations for breakpoint amplification.

Diagrammatic representation of positions of primers used to amplify across the breakpoint of Patient 1. Chromosome 1 derived material is shown in blue and Chromosome 3 derived material is shown in green. The breakpoints shown in Panel (B) are illustrative only, as the exact position of the breakpoints were unknown. Primers are denoted by red pentagons. "T" and "C" denote the telomeric and centromeric side of the breakpoint respectively. The figure is not to scale.

(A) Primer position and orientation as found on a normal chromosome.

(B) Primer position and orientation on Patient 1 translocated chromosome.

The primers shown are the closest positive (38EG211f1&r1 and DBVY253f1&r1) and negative (38EG211f2&r2 and DBVY255f1&r1) markers found by somatic cell hybrid analysis. Amplification using the external primers in the combinations shown in (b) (DBVY255f1 with 38EG211r2, and 38EG211f1 with DBVY253f1) on Patient 1 DNA will allow amplification of the breakpoint of each derivative chromosome.


(A)
38EG211f1
38EG211r1
38EG211f2
38EG211r2
T
C
Chromosome 1
DBVY255f1
DBVY255r1
DBVY253r1
DBVY253f1
T
C
Chromosome 3
(B)
DBVY255f1
DBVY255r1
38EG211f2
38EG211r2
T
C
Der1 Chromosome
38EG211f1
38EG211r1
DBVY253r1
DBVY253f1
T
C
Der3 Chromosome

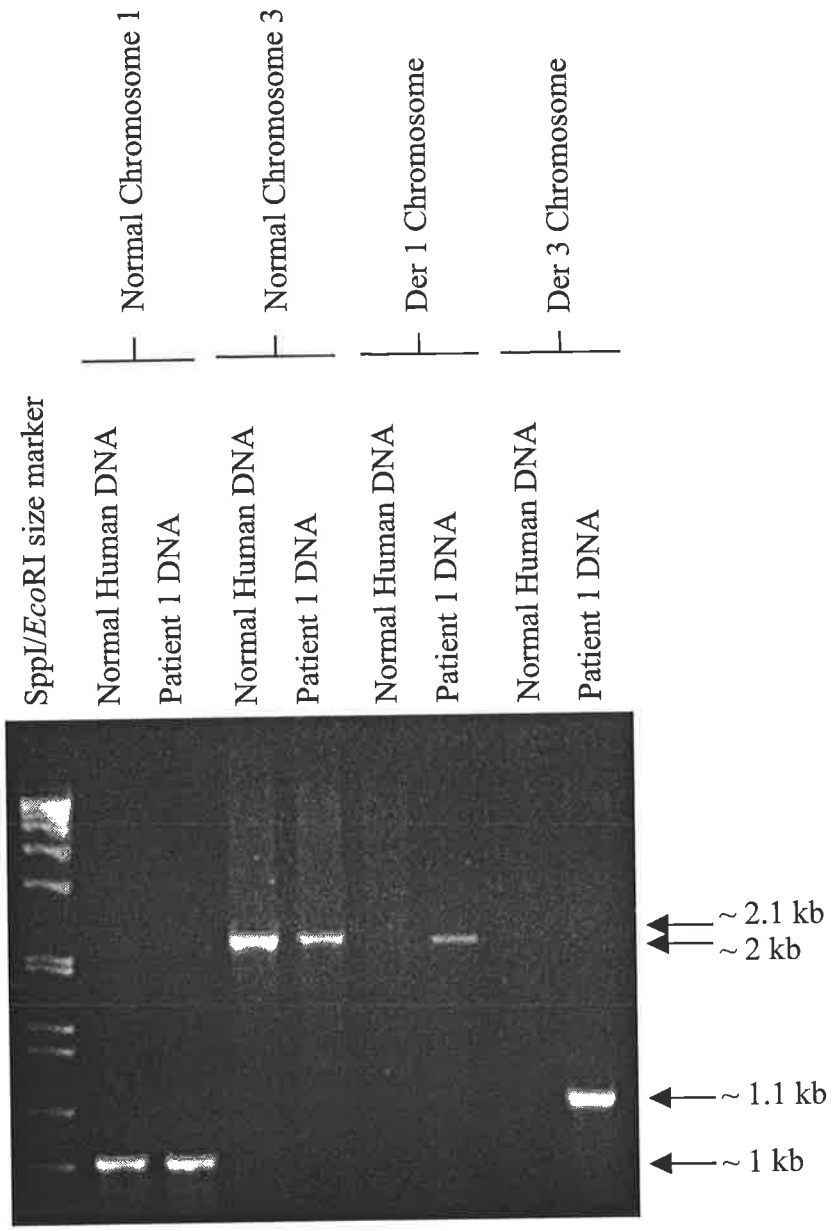


Figure 3.10. Long Template PCR Analysis of t(1;3)(p36;q21) Breakpoint in Patient 1.

The primers 38EG211f1 and 38EG211r2 were used to amplify across the breakpoint region on the normal chromosome 1, and the primers DBVY253f1 and DBVY255f1 were used to amplify across the breakpoint region on the normal chromosome 3. The primer combinations were then altered to amplify across the breakpoint on each translocated chromosome: 38EG211r2 and DBVY255f1 to amplify across the der 1 breakpoint, and 38EG211f1 and DBVY253f1 to amplify across the der 3 breakpoint.

Figure 3.11A. Sequence Alignment of the Patient 1 der 3 Breakpoint Sequence with Normal Chromosome 1 and 3 Sequences

The sequence presented in red was obtained by automated sequencing of the PCR product amplified from Patient 1 genomic DNA using the primers 38EG211f1 and DBVY253f1. The sequencing reaction was performed using the 38EG211f3 primer. This sequence is aligned with NCBI sequence data from clone RP11-333E3 (accession AL356984.13, from chromosome 1p36, represented in green) and clone RP11-475N22 (accession AC080005.24, from chromosome 3q21, represented in blue).

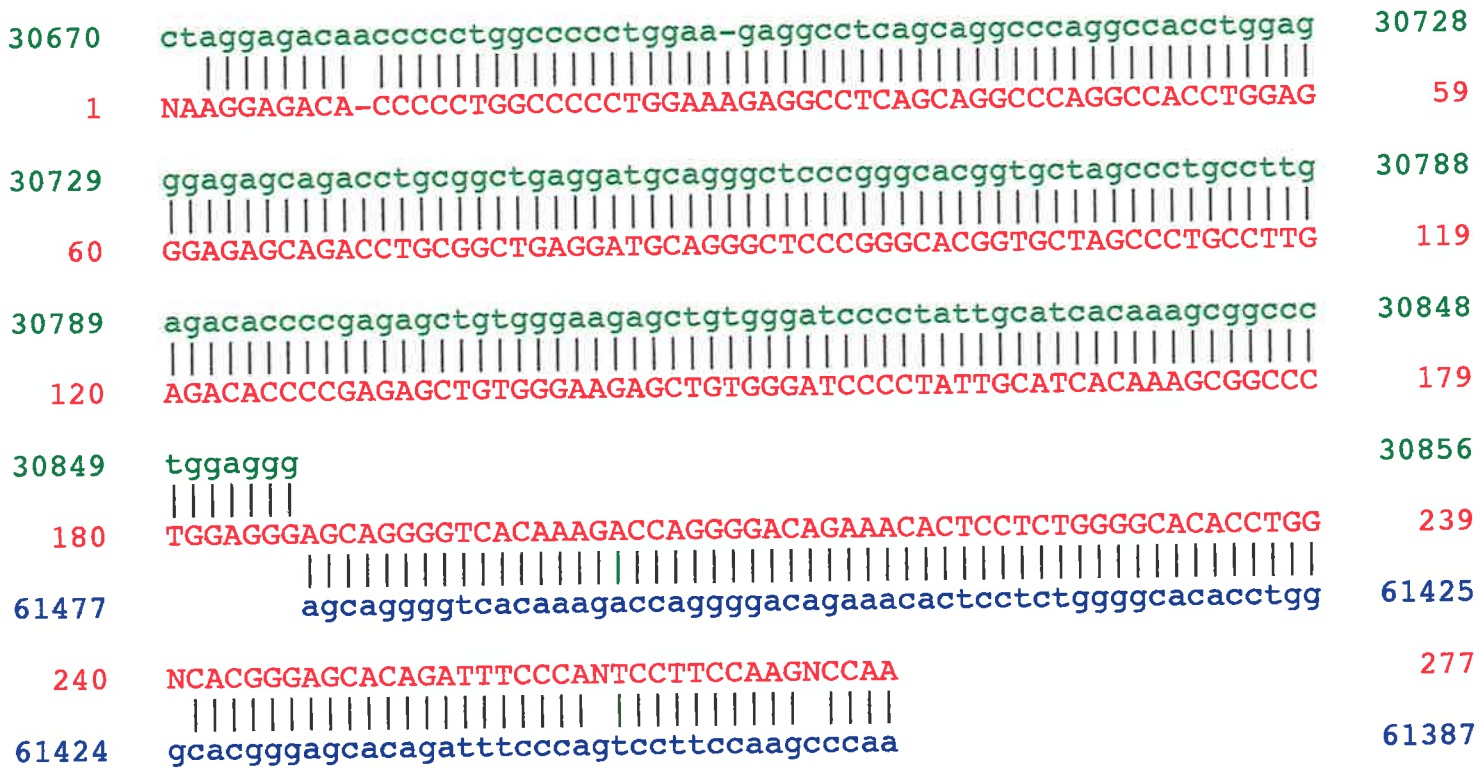
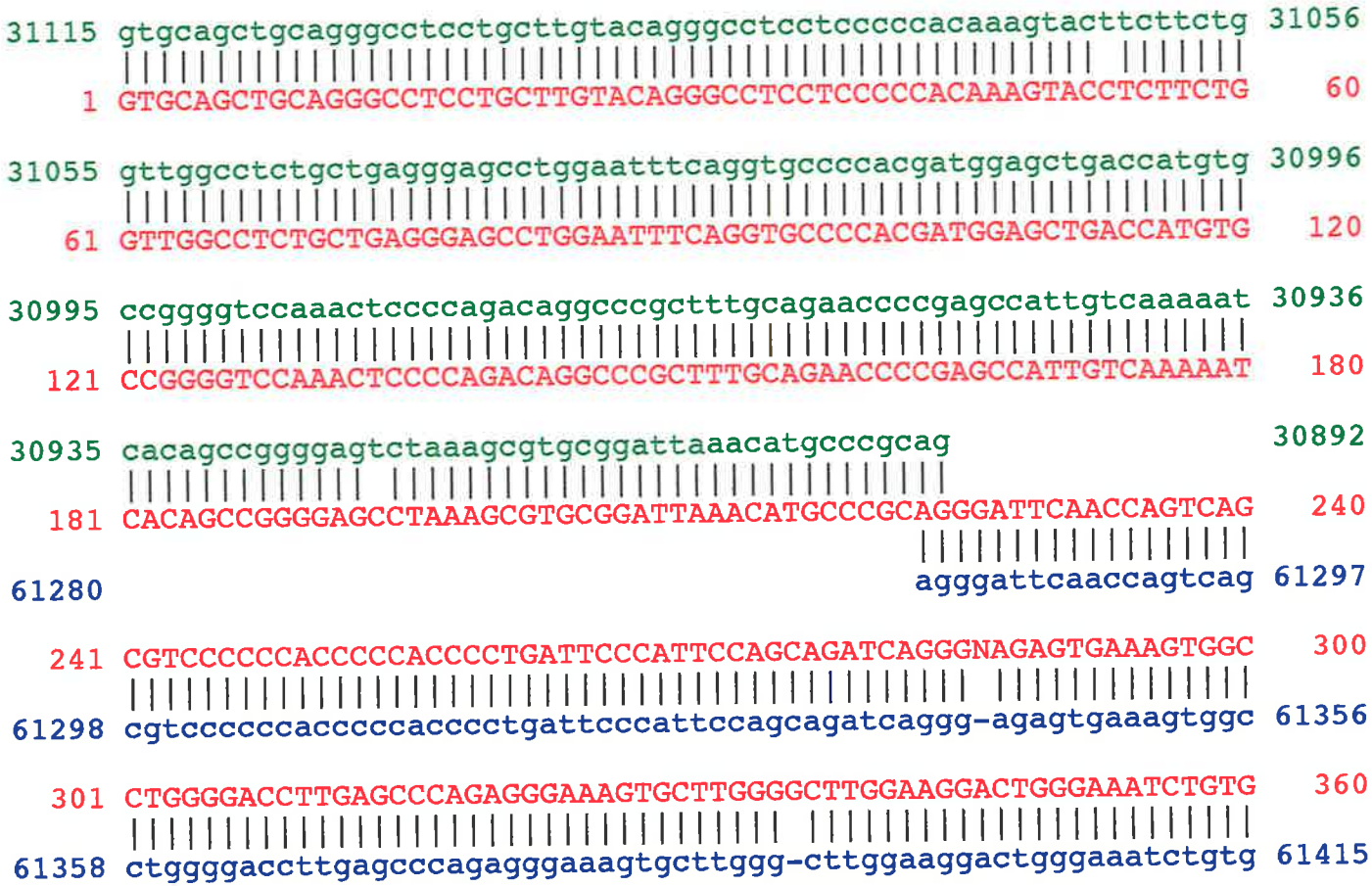

30670 ctaggagacaaccccctggcccctggaa-gaggcctcagcaggcccaggccacctggag 30728
1 NAAGGAGACA-CCCCCTGGCCCCCTGGAAAGAGGCCTCAGCAGGCCCAGGCCACCTGGAG 59
30729 ggagagcagacctgcggctgaggatgcagggctcccgggcacggtgctagccctgccttg 30788
60 GGAGAGCAGACCTGCGGCTGAGGATGCAGGGCTCCCGGGCACGGTGCTAGCCCTGCCTTG 119
30789 agacaccccgagagctgtgggaagagctgtgggatcccctattgcatcacaaagcggccc 30848
120 AGACACCCCGAGAGCTGTGGGAAGAGCTGTGGGATCCCCTATTGCATCACAAAGCGGCCC 179
30849 tggaggg 30856
180 TGGAGGGAGCAGGGGTCACAAAGACCAGGGGACAGAAACACTCCTCTGGGGCACACCTGG 239
61477 agcaggggtcacaaagaccaggggacagaaacactcctctggggcacacctgg 61425
240 NCACGGGAGCACAGATTTCCCANTCCTTCCAAGNCCAA 277
61424 gcacgggagcacagatttcccagtccttccaagcccaa 61387

Figure 3.11B. Sequence Alignment of the Patient 1 der 1 Breakpoint Sequence with Normal Chromosome 1 And 3 Sequences

The sequence presented in red was obtained by automated sequencing of the PCR product amplified from Patient 1 genomic DNA using the primers 38EG211f1 and DBVY253f1. The sequencing reaction was performed using the 38EG211r3 primer. This sequence is aligned with NCBI sequence data from clone RP11-333E3 (accession AL356984.13, from chromosome 1p36, represented in green) and clone RP11-475N22 (accession AC0800005.24, from chromosome 3q21, represented in blue).


31115 gtgcagctgcagggcctcctgcttgtacagggcctcctcccccacaaagtacttcttctg 31056
1 GTGCAGCTGCAGGGCCTCCTGCTTGTACAGGGCCTCCTCCCCCACAAAGTACCTCTTCTG 60
31055 gttggcctctgctgagggagcctggaatttcaggtgccccacgatggagctgaccatgtg 30996
61 GTTGGCCTCTGCTGAGGGAGCCTGGAATTTCAGGTGCCCCACGATGGAGCTGACCATGTG 120
30995 ccggggtccaaactccccagacaggcccgctttgcagaaccccgagccattgtcaaaaat 30936
121 CCGGGGTCCAAACTCCCCAGACAGGCCCGCTTTGCAGAACCCCGAGCCATTGTCAAAAAT 180
30935 cacagccggggagtctaaagcgtgcggattaaacatgcccgcag 30892
181 CACAGCCGGGGAGCCTAAAGCGTGCGGATTAAACATGCCCGCAGGGATTCAACCAGTCAG 240
61280 agggattcaaccagtcag 61297
241 CGTCCCCCCACCCCCACCCCTGATTCCCATTCCAGCAGATCAGGGNAGAGTGAAAGTGGC 300
61298 cgtccccccacccccacccctgattcccattccagcagatcaggg-agagtgaaagtggc 61356
301 CTGGGGACCTTGAGCCCAGAGGGAAAGTGCTTGGGGCTTGGAAGGACTGGGAAATCTGTG 360
61358 ctggggaccttgagcccagagggaaagtgcttggg-cttggaaggactgggaaatctgtg 61415

The breakpoint sequence data allowed precise positioning of the breakpoints. As already stated, the location of the 3q21 breakpoint, 124 kb 3' of *RPN1* as shown in Figure 3.12, is consistent with the findings of chromosome 3 breakpoints presented in Mochizuki *et al* (2000) (Figure 3.13), which range from approximately 80 kb to 130 kb 3' of *RPN1*. The map in Figure 3.12 also shows that the 1p36 breakpoint in Patient 1 is 43 kb telomeric of the start of *MEL1*. This breakpoint is therefore substantially closer to *MEL1* than any of the four breakpoints described in Mochizuki *et al* (2000), which range from approximately 100 kb to 220 kb upstream of *MEL1* (Figure 3.13). This estimate is based on the breakpoint map provided in Mochizuki *et al* (2000) as Figure 1. Given that the size of *MEL1* in this diagram is approximately 10 kb, whereas *MEL1* is now known to be approximately 250 kb, it is likely that the map is unreliable and therefore these estimates of breakpoint positions with respect to *MEL1* may be incorrect. Attempts to correlate the restriction sites present in the diagram with known sequence data proved unsuccessful. This may be due to the involvement of polymorphic restriction sites, but is more likely due to the incomplete characterisation of the *MEL1* gene sequence at the time the Mochizuki study was published, and therefore incorrect positioning of the start site of the *MEL1* transcript on the map presented in Mochizuki *et al* (2000).

3.4 Expression Analysis

To clarify the effect of the translocation on the transcription of genes located near the breakpoint, expression analyses were performed using RT-PCR. A second AML M1 patient bearing t(1;3)(p36;q21), Patient 4, was identified and used in expression analysis

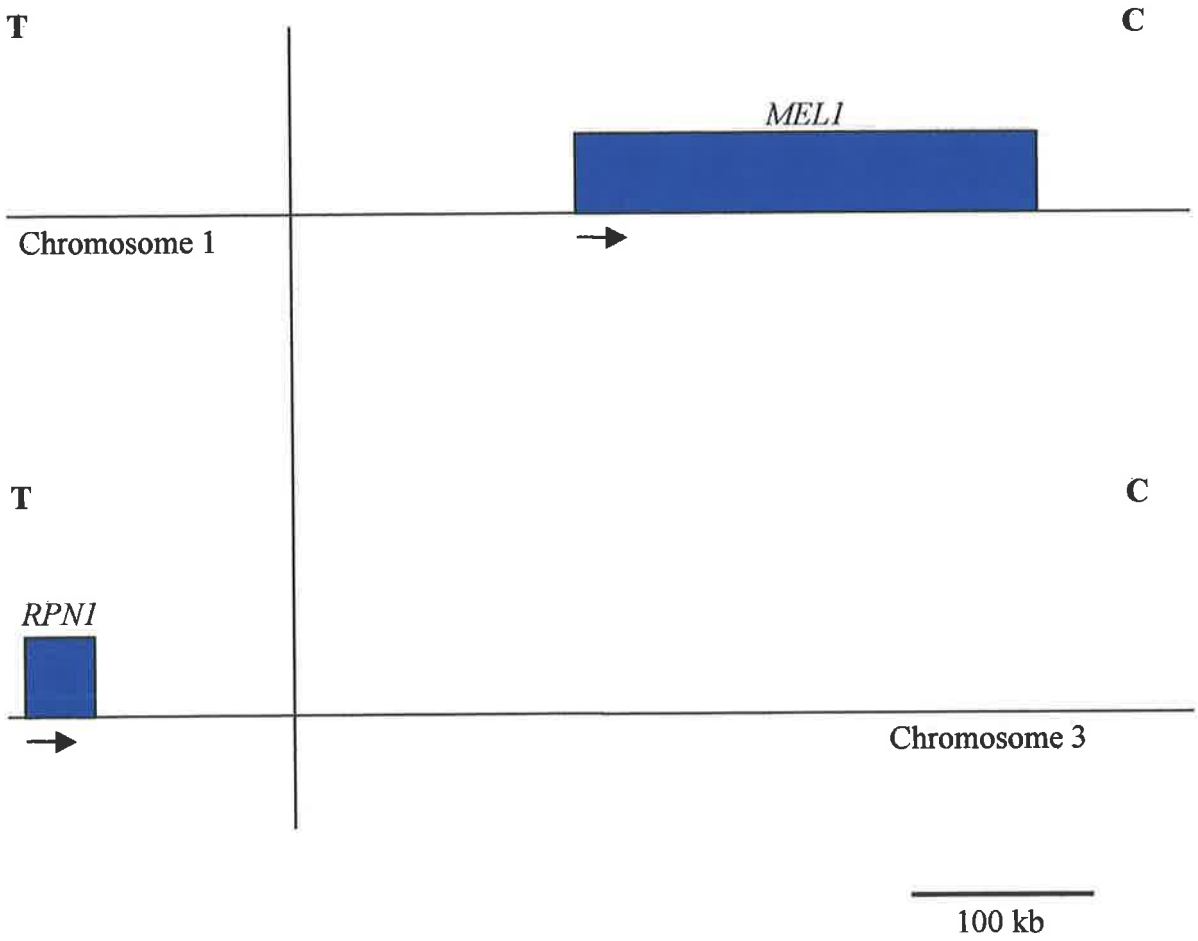


Figure 3.12. Map of Patient 1 Breakpoint with respect to *MEL1* and *RPN1*

The horizontal lines represent normal chromosome 1 (top) and normal chromosome 3 (bottom), and the vertical line passing through both represents the breakpoint. "T" and "C" indicate the telomeric and centromeric side of the breakpoint respectively. The blue boxes indicate the genes *MEL1* and *RPN1*. The genes are labelled above the appropriate boxes, and the arrows indicate the direction of transcription. The approximate scale is indicated by the 100 kb scale bar at lower left. All distances and other information are correct according to the NCBI genome assembly build 29, with the exception of the length of *MEL1*, which has been adjusted to accomodate the full-length mRNA, which features more 5' sequence.

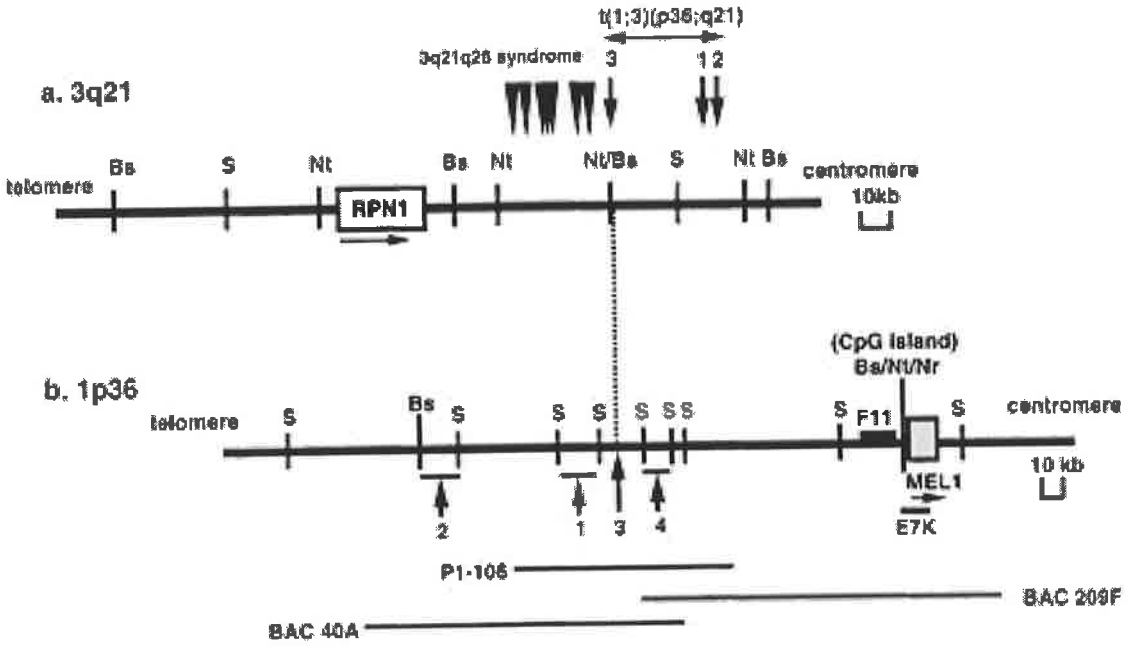


Figure 3.13. Map of t(1;3)(p36;q21) Breakpoints Reported in Mochizuki *et al* (2000).

This figure is reproduced from Mochizuki *et al* (2000) Figure 1. The caption is reproduced below.

Physical maps of the chromosomal breakpoints at 3q21 and 1p36 in t(1;3)-positive MDS/AML. (A) Mapping of the breakpoints at 3q21. Arrows and numbers indicate the breakpoints in cases 1 to 3 with t(1;3)(p36;q21), and arrowheads indicate the breakpoints in 7 AML cases with t(3;3)(q21;q26) or inv(3)(q21q26) (3q21q26 syndrome) previously reported. The position and orientation of the *RPN1* gene are indicated by a horizontal arrow. (B) Mapping of the breakpoints at 1p36. Arrows and numbers indicate the breakpoints in the 4 cases analyzed. The positions of the P1 phage clone (P1-106) and BAC clones (40A and 209F) are indicated below. A F11 cDNA fragment was isolated by the exon trapping method. The position and orientation of the *MEL1* gene are indicated by a horizontal arrow. Restriction sites are indicated by the following letters: B (*BssH*II), S (*Sfi*I), Nt (*Not*I), and Nr (*Nru*I).

as an additional source of confirmation that any effect on the expression of a gene was not specific to Patient 1 but a general outcome of the t(1;3)(p36;q21). However, there was insufficient material to make somatic cell hybrids from this patient and therefore no breakpoint mapping was performed on Patient 4.

3.4.1 *MEL1* Expression Analysis

Primers used for the initial *MEL1* RT-PCR expression analysis were rMEL1f1 and rMEL1r1. These primers produce a 501 bp product as expected. This product was sequenced in both directions and confirmed as a *MEL1* product (data not shown). In this initial analysis, a large number of PCR cycles (45) was performed to ensure amplification of any product present. Under these conditions, normal human bone marrow was found to be positive for *MEL1* expression, in contrast to the results of Mochizuki *et al* (2000). However, this expression was at a very low level (see Figure 3.14). RT-PCR of Patient 1 and Patient 4 material confirmed that *MEL1* is indeed upregulated in these cells, as seen in Figure 3.14. However, as also seen in Figure 3.14, analysis of several AML patients who do not carry the t(1;3)(p36;q21) translocation showed that a subset of them also expressed *MEL1*. This was an unexpected result, as Mochizuki *et al* (2000) reported that *MEL1* expression was specific to patients bearing the t(1;3)(p36;q21). However, the Mochizuki *et al* (2000) study only investigated t(1;3)(p36;q21)-positive patients and t(1;3)(p36;q21)-negative leukemic cell lines for *MEL1* expression, so could not rule out expression of *MEL1* in patients with t(1;3)(p36;q21)-negative AML.

Figure 3.14 *MEL1* Expression by RT-PCR Analysis.

Results from RT-PCR using the primer pair rMEL1f1 and rMEL1r1. The band (501 bp) is *MEL1* - specific. The loading order of the gel is provided below. Patients 1 and 4 are noted. No other patient carried the t(1;3)(p36;q21). *MEL1* expression is clearly present in patients with AML without this translocation. A low amount of expression is also visible in normal bone marrow.

1. AML M1 (Patient 1)

2. AML M4 (Patient 4)

3. AML M1

4. AML M4

5. AML M2

6. AML M1

7. AML M6

8. AML M2

9. AML M2

10. AML M6

11. AML M1

12. Normal Bone Marrow

13. H_2O

To determine whether the level of *MEL1* expression is similar between patients with the t(1;3)(p36;q21) translocation and those without, the *MEL1* RT-PCR was repeated at a lower cycle number, using different *MEL1* primers (rMEL1f2 and rMEL1r2) and the QuantumRNA 18S Internal Standards Kit (Ambion), to determine the relative level of expression by the use of 18S rRNA as an internal loading control. This pair of *MEL1* RT-PCR primers produce a 548 bp product as expected, but also a 491 bp product. The analysis of this smaller product is described in Section 3.4.1.1. Many AML patients were analysed using this kit, and the results are shown in Figure 3.15. Each of the four patients with 1p36 rearrangements expressed *MEL1*. Of the remaining patients, 1 out of 5 AML M1 patients and 4 out of 9 AML M2 patients expressed *MEL1*. These results suggest that *MEL1* is more frequently expressed in relatively immature leukaemias (AML M1 and M2). Examination of the clinical records of these patients revealed no correlation of non-translocation-associated *MEL1* expression with any particular chromosomal rearrangement, nor with CD34 positivity. The *MEL1* expression analysis was expanded to include 7 patients with MDS and 2 of these were also found to express *MEL1*.

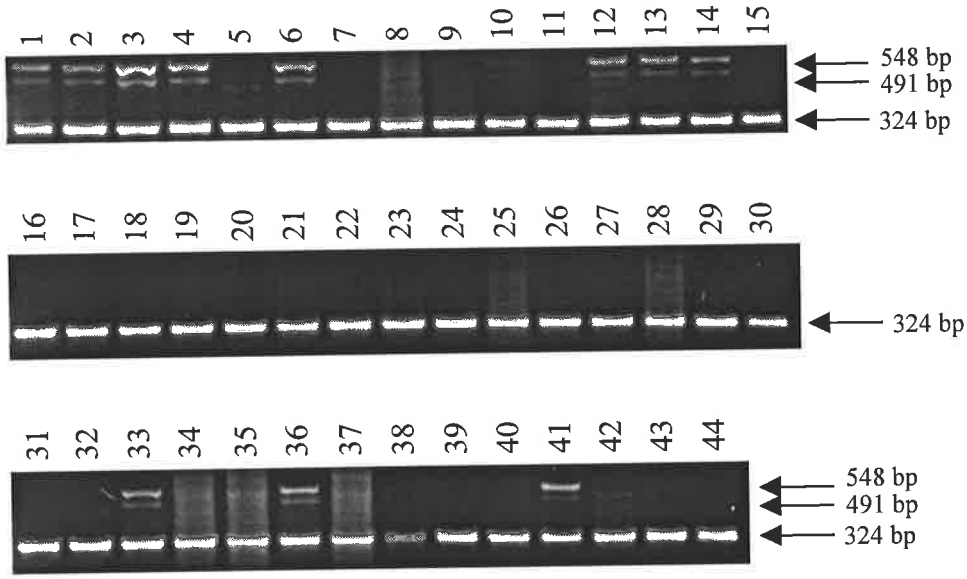
The patients not carrying the t(1;3)(p36;q21) but expressing *MEL1* at a similar level as those patients with the translocation do not bear the same clinical characteristics (dysmegakaryopoiesis and elevated or normal platelet count) as has been reported for patients with the t(1;3)(p36;q21). Therefore, it was concluded that *MEL1* expression alone cannot account for the unusual AML phenotype that results from the t(1;3)(p36;q21), and that there may be other molecular outcomes of the translocation that may account for the phenotype.

1. t(1;3)(p36;q21) AML M1 (Patient 4)
2. t(1;3)(p36;q21) AML M1 (Patient 1)
3. t(1;3)(p36;p21) AML (Patient 3)
4. ins(12;1)(p13;p36p21) AML M1 (Patient 2)
5. t(3;3)(q21;q26) AML NOS
6. inv(3)(q21q26) AML M2
7. inv(3)(q21q26) AML M2
8. AML M0/M1
9. AML M1
10. AML M1
11. AML M1
12. AML M1
13. AML M2
14. AML M2
15. AML M2

16. AML M2
17. AML M2
18. AML M2
19. AML M3
20. AML M3
21. AML M3
22. AML M3
23. AML M4
24. AML M4
25. AML M4
26. AML M4
27. AML M4 Eo
28. AML M4 Eo
29. AML M5
30. AML M6

31. AML M6
32. AML M6
33. AML M2
34. AML M4
35. MDS
36. MDS
37. MDS
38. MDS
39. MDS
40. MDS
41. MDS
42. Normal Bone Marrow
43. Normal Bone Marrow
44. Normal PBMNC

A)



B)

Lane	Major *MEL1* Band/18S	Minor *MEL1* Band/18S
Lane 1	0.253	0.053
Lane 2	0.288	0.068
Lane 3	0.893	0.162
Lane 4	0.530	0.128
Lane 6	0.483	0.099
Lane 12	0.410	0.075
Lane 13	0.419	0.097
Lane 14	0.211	0.079
Lane 33	0.248	0.080
Lane 36	0.276	0.075
Lane 41	0.429	0.017

Figure 3.15. Quantition of *MEL1* Expression

A) Results from RT-PCRs using the QuantumRNA 18S Internal Standards Kit (Ambion) and the primer pair rMEL1f2 and rMEL1r2. The two top bands (548 bp and 491 bp) are *MEL1*-specific. The two bands correspond to the the full length transcript and the splice variant detected by this primer combination (see Section 3.4.1.1 for details). The lower band (324 bp) is the Ambion 18S rRNA product, used to determine relative expression. The loading order of the gels is provided on the opposite page.

B) Quantitation of expression as determined using the Kodak 1D Gel Documentation System. Values given are the intensity of the gene-specific band expressed as a fraction of the intensity of the 18S rRNA control band.

3.4.1.1 Identification of a *MEL1* Splice Variant

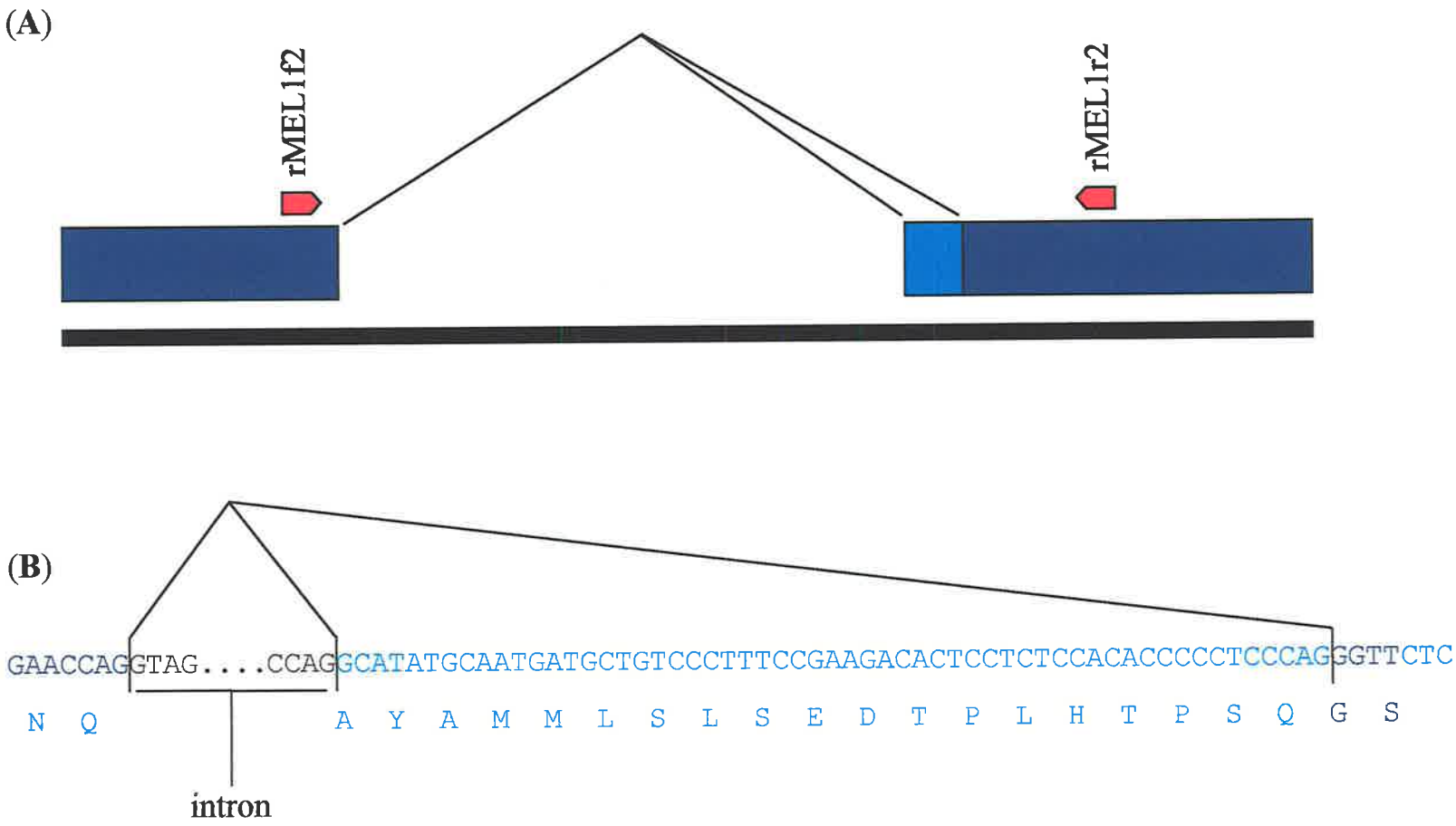
The *MEL1* RT-PCR primers also produce a slightly smaller band which, when sequenced, was found to be an alternately spliced variant of the *MEL1* transcript. The alternate form has not been previously reported, but transcripts conforming to both isoforms were found in the Genbank database (NM_022114 includes the extra sequence; AB078876 does not). Figure 3.16 details the variation in splicing. The shorter variant is missing 57 bp from the beginning of the final exon of *MEL1*. These 57 bp are within the coding region of the gene, and the removal of them removes the amino acid sequence AYAMMLSLSEDTPLHTPSQ from the MEL1 protein. This sequence occurs 25 amino acid residues prior to the stop codon. This splice variation explains the difference between the PRDM16 protein sequence (NP_071397) and the MEL1 protein sequence reported in Mochizuki *et al* (2000) and commented on earlier in Section 3.2.1 and Figure 3.6. It is unknown whether this splice variant is translated but to determine whether the biological activity of the splice variant may be different in any way, the amino acid sequence which differentiated the two versions of the proteins was investigated for functional protein domains by use of the ProDom database (http://prodes.toulouse.inra.fr/prodom/doc/prodom.html), and also for protein localisation signals using the PSORT II algorithm (http://psort.nibb.ac.jp/). Neither investigation suggested that the spliced out sequence was of functional significance. Whether the activity of the splice variant protein, if translated, is in any way different from activity of full-length MEL1 protein, is therefore also unknown.

Figure 3.16. *MEL1* Splice Variation.

Diagrammatical representation of the differences between the two variants of the *MEL1* transcript detected by RT-PCR using the primers rMEL1f2 and rMEL1r2.

(**A**) Genomic sequence is represented by the black bar, and the blue boxes represent exonic sequences of *MEL1*. The dark blue boxes are exons shared by the full length transcript and the shorter splice variant. The pale blue box is spliced in to the longer transcript and out of the shorter transcript. The position of the primers within the common exons are indicated by red arrows.

(**B**) mRNA and amino acid differences between the two products. The top line represents mRNA sequence and the bottom line represents amino acid sequence. Sequences common to both products are shown in dark blue. Sequences unique to the longer transcript are shown in pale blue. Intronic sequence is shown in black. Note that the consensus GT splice donor site and AG splice acceptor site are present within the intron.


(A)
rMEL1f2
rMEL1r2
(B)
GAACCAGGTAG....CCAGGCATATGCAATGATGCTGTCCCTTTCCGAAGACACTCCTCTCCACACCCCCTCCCAGGGTTCTC
N Q
A Y A M M L S L S E D T P L H T P S Q G S
intron

3.4.1.2 Identification of Other Genes in the Breakpoint Region

Given that expression of *MEL1* was demonstrated to be insufficient for induction of the specific clinical characteristics exhibited by t(1;3)(p36;q21)-positive patients, consideration was given to other genes that lay around the breakpoint, and whether upregulation or other altered expression of these genes could be more specifically associated with the t(1;3)(p36;q21) translocation. Genes on both 1p36 and 3q21 were considered. The NCBI Human Genome Map View web site (http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/maps.cgi) was used to identify genes that were located near either breakpoint and so may be involved in the molecular effects of the translocation. Figure 3.17 shows the location of the three genes, *ARPM2*, *ARHGEF16* and *GATA2*, chosen for expression analysis because of their proximity to the breakpoint in Patient 1.

3.4.2 *ARPM2* Expression Analysis

The transcription start site of the *ARPM2* gene is 8 bp from the 1p36 breakpoint on the der 1 chromosome. It was therefore possible that the breakpoint may disrupt the promoter and other regulatory sequences of this gene, and so *ARPM2* expression was analysed by RT-PCR. *ARPM2* is encoded by a single exon, as determined by BLASTN analysis of the transcript sequence (NM_080431) and confirmed by the NCBI Mapview website. Any analysis by RT-PCR therefore has the potential to be complicated by the amplification of a genomic product. For this reason, only DNase-treated cDNA preparations were used in the

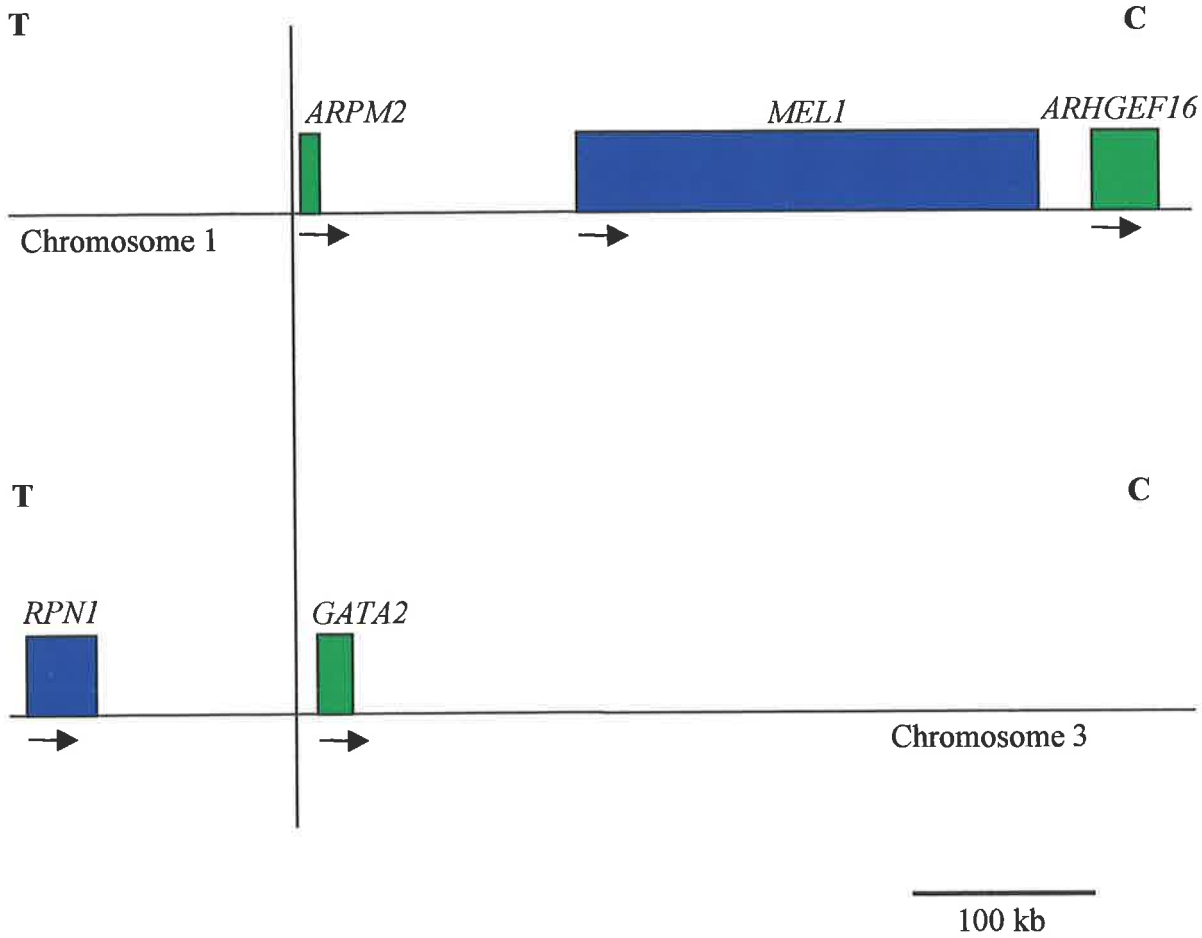


Figure 3.17. Position of *ARPM2*, *ARHGEF16* and *GATA2* with respect to the Patient 1 Breakpoint

The horizontal lines represent normal chromosome 1 (top) and normal chromosome 3 (bottom), and the vertical line passing through both represents the breakpoint. "T" and "C" indicate the telomeric and centromeric side of the breakpoint respectively. The blue boxes indicate the genes *MEL1* and *RPN1*, which have been discussed and positioned previously. The green boxes represent the genes *ARPM2*, *ARHGEF16* and *GATA2*. All genes are labelled above the appropriate boxes, and the arrows indicate the direction of transcription. The approximate scale is indicated by the 100 kb scale bar at lower right. All distances and other information are correct according to the NCBI genome assembly build 29, with the exception of the length of *MEL1*, which has been adjusted to accommodate the full-length mRNA, which features more 5' sequence.

expression analysis of this gene. Harata *et al* (2001) reported the successful amplification of an *ARPM2*-specific product from every tissue tested, including colon. Attempts by the present study to amplify a product from several DNase-treated normal colon cDNAs using primers 38EG211f2 and 38EG211r2, which are specific for *ARPM2*, were unsuccessful. Products were amplified when non-DNase-treated cDNAs were used, indicating that the PCR conditions were adequate and that DNase-treatment was important. The quality of the DNase-treated cDNA samples was verified by amplification of several other gene products (results not shown). Testing a small panel of DNase-treated AML patient RNA samples by RT-PCR, including Patient 1 and Patient 4, also resulted in no amplification (results not shown), indicating that, whether or not *ARPM2* is expressed in other tissues, it is not differentially expressed in t(1;3)(p36;q21)-positive AML and t(1;3)(p36;q21)-negative AML. The gene was therefore not investigated further.

3.4.3 *ARHGEF16* Expression Analysis

Primers ARHGEF16f1 and ARHGEF16r1, which produce a 488 bp product, were used to investigate *ARHGEF16* expression by RT-PCR using the QuantumRNA 18S Internal Standards Kit (Ambion), in a series of AML patients including Patient 1 and Patient 4. The results of this RT-PCR are shown in Figure 3.18. Quantitation of expression showed substantial variability between patients, but not a clear pattern in terms of translocation-positive and -negative patients. Transcription of the gene is therefore not consistently affected by the translocation and there was also no apparent correlation between

A)

B)

Lane	*ARHGEF16*/18S
Lane 1	1.036
Lane 2	0.656
Lane 3	0.651
Lane 4	1.236
Lane 6	0.875
Lane 8	1.061
Lane 9	0.816

Figure 3.18. Quantition of *ARHGEF16* Expression.

A) Results from RT-PCRs using the QuantumRNA 18S Internal Standards Kit (Ambion) using the primer pair ARHGEF16f1 and ARHGEF16r1. The top band (488 bp) is *ARHGEF16*-specific. The lower band is the Ambion 18S rRNA product, used to determine relative expression. The loading order of the gels is provided below.

B) Quantitation of expression as determined using the Kodak 1D Gel Documentation System. Values given are the intensity of the gene-specific band expressed as a fraction of the intensity of the 18S rRNA control band.

1. t(1;3)(p36;q21) AML M1 (Patient 4)
2. t(1;3)(p36;q21) AML M1 (Patient 1)
3. AML M2
4. AML M4 Eo
5. AML M1
6. AML M2
7. AML M2
8. Normal Bone Marrow
9. Normal Bone Marrow

ARHGEF16 expression and *MEL1* expression. *ARHGEF16* expression was therefore not investigated further.

3.4.4 *GATA2* Expression Analysis

GATA2 expression analysis was performed by RT-PCR using the QuantumRNA 18S Internal Standards Kit (Ambion) and the gene-specific primers rGATA2f1 and rGATA2r2, which produce a 421 bp product. The full panel of patients examined for *MEL1* expression were examined for *GATA2* expression also. The results of these RT-PCRs are presented in Figure 3.19. Low levels of *GATA2* transcription were observed in normal bone marrow as expected (Orkin 1992). Transcription was increased in both Patient 1 and Patient 4, suggesting that there may be an effect of the t(1;3)(p36;q21) on *GATA2* expression. As *GATA2* is on 3q21, it was of interest to investigate whether there was a similar effect in patients with the t(3;3)(q21;q26) or inv(3)(q21q26), as these rearrangements have similar 3q21 breakpoints and phenotypic correlations. *GATA2* expression analysis of the t(3;3)(q21;q26) patient and the two inv(3)(q21q26) patients examined did indeed show strong upregulation of *GATA2* transcription.

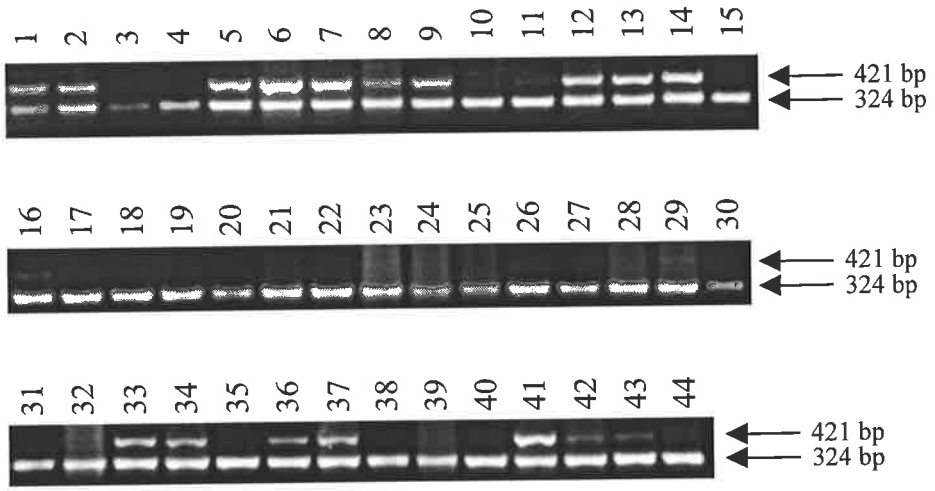
GATA2 transcription was also seen, however, in several t(1;3)(p36;q21)-negative AML patients. This phenomenon has been previously reported, and attributed to a leukemic expansion of haematopoietic progenitor cells which normally express *GATA2* (Tsuzuki *et al* 2000). The results presented here would appear to support this hypothesis, as those leukaemias that show *GATA2* expression are those that arise from haematopoietic cells

1. t(1;3)(p36;q21) AML M1 (Patient 4)
2. t(1;3)(p36;q21) AML M1 (Patient 1)
3. t(1;3)(p36;p21) AML (Patient 3)
4. ins(12;1)(p13;p36p21) AML M1 (Patient 2)
5. t(3;3)(q21;q26) AML NOS
6. inv(3)(q21q26) AML M2
7. inv(3)(q21q26) AML M2
8. AML M0/M1
9. AML M1
10. AML M1
11. AML M1
12. AML M1
13. AML M2
14. AML M2
15. AML M2

16. AML M2
17. AML M2
18. AML M2
19. AML M3
20. AML M3
21. AML M3
22. AML M3
23. AML M4
24. AML M4
25. AML M4
26. AML M4
27. AML M4 Eo
28. AML M4 Eo
29. AML M5
30. AML M6

31. AML M6
32. AML M6
33. AML M2
34. AML M4
35. MDS
36. MDS
37. MDS
38. MDS
39. MDS
40. MDS
41. MDS
42. Normal Bone Marrow
43. Normal Bone Marrow
44. Normal PBMNC

A)



B)

Lane	*GATA2*/18S
Lane 1	1.137
Lane 2	0.826
Lane 5	1.119
Lane 6	1.662
Lane 7	1.266
Lane 8	0.408
Lane 9	0.781
Lane 11	0.028
Lane 12	0.830
Lane 13	1.008
Lane 14	1.055
Lane 16	0.064
Lane 33	0.490
Lane 34	0.331
Lane 36	0.281
Lane 37	0.375
Lane 41	0.978
Lane 42	0.093
Lane 43	0.086

Figure 3.19 Quantition of *GATA2* Expression

A) Results from RT-PCRs using the QuantumRNA 18S Internal Standards Kit (Ambion) using the primer pair rGATA2f1 and rGATA2r2. The top band (421 bp) is *GATA2*-specific. The lower band (324 bp) is the Ambion 18S rRNA product, used to determine relative expression. The loading order of the gels is provided on the opposite page and is the same as that of the gels shown in Figure 3.15.

B) Quantitation of expression as determined using the Kodak 1D Gel Documentation System. Values given are the intensity of the gene-specific band expressed as a fraction of the intensity of the 18S rRNA control band.

that are at earlier stages of differentiation, when *GATA2* is normally expressed (Orkin 1992). There was no apparent correlation between non-translocation associated GATA2 expression and any particular chromosomal rearrangement or CD34 positivity. There appeared to be a quantitative difference in the level of *GATA2* expression between patients with 3q21 breakpoints and those without, however, and quantitation was undertaken to confirm this. The results are presented in Figure 3.19B. There is not a clear distinction in expression levels between *GATA2*-expressing AML patients with and without 3q21 rearrangements, but there does appear to be a trend showing greater expression levels in patients with 3q21 breakpoints. *GATA2* dysregulation may, therefore, be a genuine effect of the t(1;3)(p36;q21), the t(3;3)(q21;q26) and the inv(3)(q21q26).

3.4.5 *EVI1* Expression Analysis

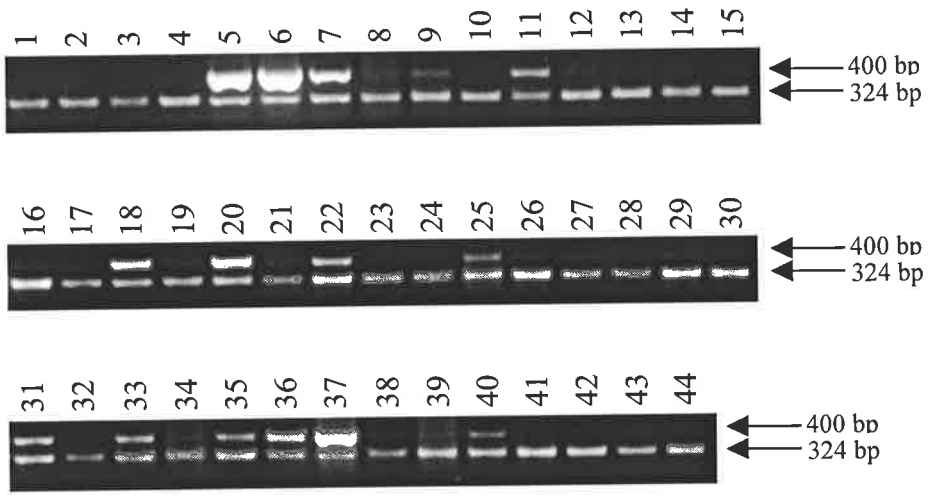
EVI1 expression was examined to confirm that transcription was seen in patients with t(3;3)(q21;q26) and inv(3)(q21q26), and also to confirm previous reports that transcription was not restricted to leukaemias with these rearrangements (Russell *et al* 1994, Langabeer *et al* 2001). Analysis was performed by RT-PCR using the QuantumRNA 18S Internal Standards Kit (Ambion) and the gene-specific primers rEVI1f1 and rEVI1r1, which amplify a 400 bp product. The results of this analysis are presented in Figure 3.20. *EVI1* was expressed as expected in the three patients with 3q26 rearrangements. It was also seen in a subset of the other AML and MDS patients. Unlike *MEL1* and *GATA2* expression in t(1;3)(p36;q21)-negative and 3q21 rearrangement-negative patients respectively, *EVI1*

1. t(1;3)(p36;q21) AML M1 (Patient 4)
2. t(1;3)(p36;q21) AML M1 (Patient 1)
3. t(1;3)(p36;p21) AML (Patient 3)
4. ins(12;1)(p13;p36p21) AML M1 (Patient 2)
5. t(3;3)(q21;q26) AML NOS
6. inv(3)(q21q26) AML M2
7. inv(3)(q21q26) AML M2
8. AML M0/M1
9. AML M1
10. AML M1
11. AML M1
12. AML M1
13. AML M2
14. AML M2
15. AML M2

16. AML M2
17. AML M2
18. AML M2
19. AML M3
20. AML M3
21. AML M3
22. AML M3
23. AML M4
24. AML M4
25. AML M4
26. AML M4
27. AML M4 Eo
28. AML M4 Eo
29. AML M5
30. AML M6

31. AML M6
32. AML M6
33. AML M2
34. AML M4
35. MDS
36. MDS
37. MDS
38. MDS
39. MDS
40. MDS
41. MDS
42. Normal Bone Marrow
43. Normal Bone Marrow
44. Normal PBMNC

A)



B)

Lane	*EVI1*/18S
Lane 5	3.804
Lane 6	3.508
Lane 7	1.761
Lane 8	0.260
Lane 9	0.371
Lane 11	1.017
Lane 18	2.059
Lane 20	2.026
Lane 22	0.566
Lane 25	0.366
Lane 31	0.769
Lane 33	0.734
Lane 35	0.544
Lane 36	1.497
Lane 37	2.992
Lane 40	0.432

Figure 3.20 Quantition of *EVI1* Expression

A) Results from RT-PCR using the QuantumRNA 18S Internal Standards Kit (Ambion) using the primer pair rEVIf1 and rEVIr1. The top band (400 bp) is *EVI1*-specific. The lower band (324 bp) is the Ambion 18S rRNA product, used to determine relative expression. The loading order of the gels is provided on the opposite page and is the same as that of the gels shown in Figure 3.15.

B) Quantitation of expression as determined using the Kodak 1D Gel Documentation System. Values given are the intensity of the gene-specific band expressed as a fraction of the intensity of the 18S rRNA control band.

expression in 3q26 rearrangement-negative patients did not appear to correlate with AML subtype. Expression levels were quantitated and this data is shown in Figure 3.20B.

3.4.6 Expression Analysis Summary

The results for the expression analysis of *MEL1*, *GATA2* and *EVI1* are presented in Table 3.1, to allow easy cross-referencing between genes and to illustrate those patients who express more than one of these genes.

For those patients who expressed either *GATA2*, *EVI1* or *MEL1*, patient records were checked to determine whether they displayed normal or elevated platelets, or any evidence of dysmegakaryopoiesis. Other than those patients with either t(1;3)(p36;q21), t(3;3)(q21;q26) or inv(3)(q21q26), all patients had platelet counts below the normal range, and no mention of dysmegakaryopoiesis was found. This was true also for patients who expressed *GATA2* and either *EVI1* or *MEL1*. This confirms that neither *GATA2*, *EVI1* nor *MEL1* expression can account for the clinical phenotype that is associated with these translocations, and nor can any combination of the expression of these genes. The effect of sufficiently high levels of *GATA2* expression remains a possible cause of the phenotype, however.

Patient	AML Subtype	Chromosome Rearrangement	*MEL1*	*EVI1*	*GATA2*
1 (Patient 4)	M1	t(1;3)(p36;q21)	+	-	+
2 (Patient 1)	M1	t(1;3)(p36;q21)	+	-	+
3 (Patient 3)	Indeterminate	t(1;3)(p36;p21)	+	-	-
4 (Patient 2)	M1	ins(12;1)(p13;p36p21)	+	-	-
5	NOS	t(3;3)(q21;q26)	-	+	+
6	M2	inv(3)(q21q26)	+	+	+
7	M2	inv(3)(q21q26)	-	+	+
8	M0/M1		-	+	+
9	M1		-	+	+
10	M1		-	-	-
11	M1		-	+	+
12	M1		+	-	+
13	M2		+	-	+
14	M2		+	-	+
15	M2		-	-	-
16	M2		-	-	+
17	M2		-	-	-
18	M2		-	+	-
19	M3		-	-	-
20	M3		-	+	-
21	M3		-	-	-
22	M3		-	+	-
23	M4		-	-	-
24	M4		-	-	-
25	M4		-	+	-
26	M4		-	-	-
27	M4 Eo		-	-	-
28	M4 Eo		-	-	-
29	M5		-	-	-
30	M6		-	-	-
31	M6		-	+	-
32	M6		-	-	-
33	M2		+	+	+
34	M4		-	-	+
35	MDS		-	+	-
36	MDS		+	+	+
37	MDS		-	+	+
38	MDS		-	-	-
39	MDS		-	-	-
40	MDS		-	+	-
41	MDS		+	-	+
42	Normal BM		-	-	+
43	Normal BM		-	-	+
44	Normal PB		-	-	-

Table 3.1. Summary of Expression Analysis. Results from Figures 3.15, 3.20 and 3.21 are summarised here. + indicates that expression was detected by RT-PCR. - indicates that expression was not detected. BM = bone marrow. PB = peripheral blood. NOS = no origin specified. Only chromosome rearrangements involving 1p36, 3q21 or 3q26 are specified.

3.5 Discussion

3.5.1 Position of Translocation Breakpoints in Patient 1

The 1p36 breakpoint in Patient 1 was determined to be 43 kb 5' of the *MEL1* gene. It is difficult to ascertain with any certainty the position of the 1p36 breakpoint in the patients presented in Mochizuki *et al* (2000) because of the probable misplacement of the *MEL1* gene on their map, which in turn is most likely due to the incomplete characterisation of *MEL1* at the time that the map was constructed. If this map is correct, then the breakpoint reported in Patient 1 is the closest breakpoint to *MEL1* reported thus far. It is more likely that the distances reported in Mochizuki are overestimated, and that the four Mochizuki patients have breakpoints in the same approximate region. Nevertheless, the map suggests that the four breakpoints range over an approximately 80 kb region, and so it seems likely that the 1p36 breakpoints in the t(1;3)(p36;q21) are clustered over quite a large region.

The 3q21 breakpoint in Patient 1 is in line with expectations according to previously mapped t(1;3)(p36;q21), t(3;3)(q21;q26) and inv(3)(q21q26) breakpoints, which, as previously mentioned, cluster in two regions, BCR-C and BCR-T (reviewed in Wieser 2002) (see Figure 3.21). The breakpoint found in Patient 1 is 6 kb from the 5' end of *GATA2*, and therefore fits into the centromeric end of BCR-C.

Figure 3.21. Breakpoint Cluster Regions at 3q21.

The figure shows the relative positions of RPN1 and GATA2 (represented by blue boxes with arrows indicating the direction of transcription) with respect to the breakpoint cluster regions BCR-C and BCR-T (represented by pink boxes). All known t(1;3)(p36;q21), t(3;3)(q21;q26) and inv(3)(q21q26) 3q21 breakpoints cluster in one of these two regions. C and T indicate the centromeric and telomeric end of the region shown respectively. The figure is approximately to scale. The figure was adapted from Wieser 2002.

3.5.2 *MEL1* Expression

3.5.2.1 Identification of a Novel *MEL1* Splice Variant

Examination of *MEL1* expression led to the identification of a *MEL1* splice variant. The two isoforms vary by the inclusion or exclusion of the amino acid sequence AYAMMLSLSEDTPLHTPSQ, 25 residues from the C terminus of the MEL1 protein (see Figure 3.16). It appears as though the variants are not differentially regulated at the transcriptional level, since the expression of one as detected by the *MEL1* RT-PCR is always accompanied by the expression of the other. However, whether both transcripts are translated and, if translated, whether both proteins have similar functions and function to the same level, is unknown. The NCBI database contains protein sequences of both variants (NP_071397 and XP_010556), but it is unclear as to whether these are conceptual translations (that is, based only on transcript sequences) or actual protein sequences. The biological relevance of the splice variation is therefore unknown. It is clear, however, that this splice variant does not encode a PR domain-negative version of the MEL1 protein, as the PR domain is encoded by the 5' end of the *MEL1* gene, and this splice variation occurs much further towards the 3' end.

3.5.2.2 Expression of *MEL1* in MDS and AML Patients without t(1;3)(p36;q21)

The data obtained in this study supports the notion that *MEL1* is activated by the occurrence of the t(1;3)(p36;q21), as it is expressed in both patients with this translocation

who were examined. However, the data also demonstrates that *MEL1* is expressed in cases of AML without the t(1;3)(p36;q21).

Of the six t(1;3)(p36;q21) patients examined for *MEL1* expression (two in the present study and four in Mochizuki *et al* (2000)), all six were found to express the gene at high levels. Conversely, of the thirty AML patients without the t(1;3)(p36;q21), not including Patients 2 and 3 who have other rearrangements of 1p36, only five were positive for *MEL1* expression. This correlation of *MEL1* expression with the occurence of the t(1;3)(p36;q21) supports the hypothesis that the t(1;3)(p36;q21) is responsible for the upregulation of *MEL1* in patients with this translocation. This upregulation is likely to be *via* the position effect of the juxtaposed *RPN1* enhancer sequences from 3q21, as proposed by Mochizuki *et al* (2000), and similar to the proposed mechanism of *EVI1* upregulation caused by the t(3;3)(q21;q26) and inv(3)(q21q26) (reviewed in Nucifora 1997).

There are two possible mechanisms to explain *MEL1* expression in patients without t(1;3)(p36;q21), One possibility is that *MEL1* is normally expressed by a subpopulation of haematopoietic progenitor cells and that, when these cells are the target cells of transformation for leukemogenesis, this expression is maintained and amplified merely as a by-product of this expansion. This possibility is suggested by the correlation between expression of *MEL1* in t(1;3)(p36;q21)-negative AML and early AML subtype, as this result fits with normal *MEL1* expression in a particular type of early haematopoietic cell. Four of the *MEL1*-expressing t(1;3)(p36;q21)-negative AML cases were diagnosed as AML M2, while the fifth case was AML M1. This suggests that *MEL1* may be normally

expressed by a moderately mature progenitor cell, and that these *MEL1*-expressing t(1;3)(p36;q21)-negative leukemias result from the expansion of a member of this population. This possibility is also supported by the finding that *MEL1* is expressed at low levels in normal bone marrow. *MEL1* expression was not seen in the analysis of normal bone marrow in the Mochizuki *et al* (2000) study, but experiments detecting weak *MEL1* expression by using higher RT-PCR cycle numbers in the present study showed that a signal was detectable (see Figure 3.14), suggesting that a small subset of haematopoietic progenitor cells may normally express *MEL1*. The hypothesis is also supported by the correlation of the *MEL1* expression findings with the results obtained for *GATA2* expression. Every non-t(1;3)(p36;q21) patient who expressed *MEL1* also expressed *GATA2*, which is known to be expressed normally in a subset of early haematopoietic cells (Persons *et al* 1999). It is also known that leukemic expansion of this population does occur, and results in leukemias which show apparent upregulation of *GATA2* expression (Tsuzuki et al (2000)).

There is a lack of correlation between non-translocation-associated *MEL1* expression and CD34 positivity (see Table 3.1), and the expansion model is therefore dependant on the existence of two populations of haematopoietic progenitors which normally express *MEL1*: one which is $CD34^+$ and one which is $CD34^-$. Whether this is the case is unknown, and *MEL1* expression analysis on flow-sorted bone marrow subpopulations would be of interest in assessing the credibility of this hypothesis. Sorted subpopulation expression analyses have been performed for *GATA2* expression, and these indicate that *GATA2* is expressed in both $CD34^+$ and $CD34^-$ fractions (Nakauchi et al, 1999). This is consistent with the progenitor expansion hypothesis as an explanation for *GATA2* expression in non-

translocation patients which, as observed in this study, also does not correlate with CD34 positivity. Therefore, as *GATA2*-expressing progenitor expansion can occur for $CD34^+$ and $CD34^-$ leukaemias, it may also be a possible explanation for non-translocation associated *MEL1* expression, especially given the correlation between non-translocation associated *GATA2* and *MEL1* expression. If *MEL1* expression in t(1;3)(p36;q21)-negative AML is due to this expansion of a normally *MEL1* expressing cell, then it is unlikely to contribute to leukemogenesis within this cell type because the expression is not abnormal within the context of the cell.

The second possible mechanism by which *MEL1* expression in patients without t(1;3)(p36;q21) occurs is that *MEL1* expression is aberrantly activated by some means other than juxtaposition to enhancer elements by translocation. While the present study offers no evidence as to the nature of this second mechanism, one possible mechanism is that *MEL1* is expressed as a result of activation by another regulatory protein, which is aberrantly expressed in those leukemic patients in which *MEL1* expression is seen. To date, no potential regulators of *MEL1* expression have been identified in the literature. Given the evidence discussed above, this mechanism is considered less likely than the model of leukemic expansion of a normally *MEL1* expressing progenitor. However, it is known that genes which are upregulated by leukemic chromosome translocation can also be upregulated by other mechanisms. Examples of this include *BCL1* and *MYC* (Falini and Mason 2002, Hecht and Aster 2000). *MEL1* expression *via* this mechanism would be more likely to contribute to leukemogenesis, as the expression is occurring in a cell type in which it is not normally expressed, thus having abnormal transforming effects.

3.5.2.3 Oncogenic Significance of *MEL1*

Although expression of *MEL1* is not exclusively linked to patients carrying the t(1;3)(p36;q21) translocation, it is aberrantly expressed in leukemic cells and therefore it is possibly involved in the development or progression of leukaemia. Indeed, the fact that it may be upregulated by two independent mechanisms may be seen as stronger evidence that *MEL1* is an important gene in leukemogenesis.

Other evidence for the importance of *MEL1* expression in leukemogenesis may be extrapolated from studies of the other members of the PRDM gene family, such as *MDS1-EVI1* and *RIZ*. *MDS1-EVI1* has long been known to be involved in leukaemia, *via* translocation-mediated upregulation and also aberrant expression caused by other mechanisms (reviewed in Nucifora 1997). The *RIZ* gene, located on 1p36, has also been implicated in oncogenic processes, especially by differential expression of its two isoforms (He *et al* 1998), a mechanism that probably also applies to *MDS1-EVI1* involvement in leukaemia (reviewed in Nucifora 1997) and which may also apply to *MEL1*. Mochizuki *et al* (2000) saw some evidence of expression of a shorter transcript of *MEL1* by Northern analysis, and it seems possible that this shorter transcript is generated by an internal transcription start site which occurs 3' of the *MEL1* PR domain. If such a transcript is expressed, then differential regulation of these two isoforms of *MEL1* may be an oncogenic mechanism in this gene as is the case for *MDS1-EVI1* and *RIZ*. More study is required to determine whether this mechanism also applies to *MEL1*.

3.5.2.4 *MEL1* Expression is Insufficient to Explain Clinical Observations

The fact that *MEL1* is expressed in patients who have neither the t(1;3)(p36;q21) nor the clinical phenotype associated with it (dysmegakaryopoiesis and elevated platelets) suggests that *MEL1* expression alone may be insufficient to create this phenotype. However, this depends on the mechanism by which *MEL1* expression arises, as expression of *MEL1* within a normally *MEL1*-expressing cell is unlikely to contribute to an aberrant phenotype (as may be the case in t(1;3)(p36;q21)-negative AML), whereas translocation-mediated expression in a cell which does not normally express *MEL1* may. Therefore, aberrant *MEL1* expression may be perfectly correlated with clinical phenotype, if *MEL1* expression in t(1;3)(p36;q21)-negative AML is not considered aberrant.

A review of the literature revealed that a similar phenomenon occurs involving *EVI1*. As previously mentioned, the *EVI1* gene at 3q26 is expressed in patients with either the t(3;3)(q21;q26) or the inv(3)(q21q26), and these patients share the clinical phenotype described for patients carrying the t(1;3)(p36;q21). Patients with one of the three rearrangements present with dysmegakaryopoiesis and elevated platelets. As *EVI1* and *MEL1* are related genes, it is tempting to suggest that the expression of either one of these genes can create the clinical phenotype witnessed in patients with either of the

translocations. However, *EVI1* is also expressed in AML patients with no rearrangement of chromosome 3q, and these patients lack the distinct clinical phenotype associated with the 3q rearrangements (Langabeer *et al* 2001). The expansion of a normally-expressing cell probably does not explain *EVI1* expression in 3q21q26-negative AML, as *EVI1* expression in 3q21q26-negative AML does not correlate with AML subtype. Also, there are reports of 3q21q26 AML which do not display *EVI1* expression, although it is not stated whether these patients display elevated platelet counts and dysmegakaryopoiesis (Soderholm *et al* 1997, Langabeer *et al* 1997). As the present study has shown that patients without the t(1;3)(p36;q21) who express *MEL1* also lack the clinical phenotype, it must be concluded that expression of either of these genes is insufficient to cause the clinical phenotype. As discussed above, however, it may be that inappropriate expression of *MEL1*, caused by the t(1;3)(p36;q21), causes the phenotype, while expression in other patients may not have a clinical effect because the expression is due to the expansion of a normally-expressing progenitor cell.

The focus of the present study shifted to examine other possible outcomes of the t(1;3)(p36;q21) translocation, such as aberrant expression of other genes near the breakpoint, which may account for this phenotype either alone or in combination with expression of *MEL1*.

3.5.3 *ARPM2* Expression Analysis

The *ARPM2* gene is located 8 bp from the 1p36 breakpoint in Patient 1, and was therefore a candidate for translocation-mediated transcriptional dysregulation. In investigating the expression pattern of *ARPM2* in normal tissues, Harata *et al* (2001) amplified a product corresponding to *ARPM2* from colon cDNA. Because *ARPM2* is a single-exon gene, it is impossible to design an RT-PCR for this gene that will allow the distinction of a genomic product from an mRNA product on the basis of size. For this reason, DNase-treated cDNA must be used in any RT-PCR investigating a single-exon gene. Harata *et al* (2001) do not state whether the cDNA used in their study was DNase-treated, although their RT negative controls are negative, indicating that they may have DNase-treated their samples. Using primers that matched a sequence deposited in Genbank by these authors (NM_080431) and DNase-treated colon cDNA, the present study was unable to replicate this result. Harata *et al* (2001) demonstrated that colon was one of the lower-expressing tissues examined. Possibly the RT-PCR employed in the present study was not as sensitive as that used in Harata *et al* (2001). This possibility seems unlikely, however, given that the primers and conditions employed by the present study amplify a product from genomic DNA controls and non-DNase-treated cDNA samples. Because the conditions of the PCR used in the present study are known to successfully amplify, however, it can be concluded that *ARPM2* is not differentially expressed in t(1;3)(p36;q21)-positive AML and t(1;3)(p36;q21)-negative AML. *ARPM2* is therefore of no further relevance to the present study.

3.5.4 *ARHGEF16* Expression Analysis

Although *ARHGEF16* has not been characterised in detail, by sequence homology it is thought to be a guanine exchange factor (GEF) for Rho proteins. Rho proteins are a branch of the Ras superfamily involved in numerous signal transduction pathways which regulate a variety of cellular responses (Ridley 2001, Wherlock and Mellor 2002). GEFs regulate the activation state of the Rho proteins by the addition of a guanine group (Overbeck *et al* 1995, Pan and Wessling-Resnick 1998). Members of the Rho family, and several of their guanine exchange factors (GEFs), have been implicated in oncogenic processes (reviewed in Jaffe and Hall 2002, Boettner and Van Aelst 2002). Examples of RhoGEFs thought to be involved in leukemogenesis include *VAV*, *LARG* and *CLG* (del Peso *et al* 1997, Esteve *et al* 1998, Booden *et al* 2002, Himmel *et al* 2002). *ARHGEF16* is 3' of *MEL1* on 1p36 and is oriented in the same direction, so it is reasonable to hypothesise that a general upregulation of transcription mediated by a long-range enhancer element as a result of the t(1;3)(p36;q21) translocation may affect *ARHGEF16* as well as *MEL1*.

The results of the expression analysis of *ARHGEF16* in t(1;3)(p36;q21) and non-t(1;3)(p36;q21) patients showed that transcription of the gene in leukaemia is not dependent on the presence of the translocation. The two patients with t(1;3)(p36;q21)-positive AML examined showed moderate and low levels of *ARHGEF16* expression respectively. Expression levels among t(1;3)(p36;q21)-negative AML patients varied quite markedly, and some patients did not express *ARHGEF16* at all, despite the fact that it was expressed at moderate levels in normal bone marrow. Therefore, any differences in

ARHGEF16 expression among AML patients are not linked to the t(1;3)(p36;q21). Of the patients with t(1;3)(p36;q21)-negative AML examined for *ARHGEF16* expression, only one was positive for *MEL1* expression. This patient was negative for *ARHGEF16* expression. Therefore, the mechanism responsible for upregulation of *MEL1* expression in patients with t(1;3)(p36;q21)-negative AML does not have a similar upregulation effect on *ARHGEF16* expression. It is possible that variation of expression of the gene does play a role in leukaemia given the variety of expression levels seen, but this role is not connected to the presence of the t(1;3)(p36;q21).

3.5.5 *GATA2* Expression Analysis

The transcription start site of *GATA2* is 6 kb from the 3q21 breakpoint identified in Patient 1, and was therefore a candidate for altered expression as a result of the t(1;3)(p36;q21). The GATA transcription factors are a highly related family of proteins which bind the DNA sequence motif GATA and have transactivation capacity (reviewed in Orkin 1998). *GATA1*, *GATA2* and *GATA3* are known to play important roles in haematopoiesis (Weiss and Orkin 1995), while *GATA4*, *GATA5* and *GATA6* are involved in non-haematopoietic developmental pathways (reviewed in Molkentin 2000).

3.5.5.1 Normal Function of *GATA2* in Haematopoiesis

GATA2 plays a critical role in the development of numerous non-haematopoietic organs (Dorfman *et al* 1992, Brewer *et al* 1995, Ma *et al* 1997, Zhou *et al* 1998), but this

discussion will be restricted to the role *GATA2* plays in haematopoiesis. *GATA2* is generally accepted as important in the proliferation and maintenance of immature multipotential haematopoietic progenitors and probably haematopoietic stem cells themselves. Its expression is seen in these progenitors and also in mast cells, megakaryocytes, erythrocytes, monocytes, neutrophils, eosinophils, basophils and early myeloid- and erythroid-specific progenitors (Nagai *et al* 1994, Tsai *et al* 1998, Yamaguchi *et al* 1998). Lack of *GATA2* expression in progenitor cells leads to failure to respond to haematopoietic cytokines, leading to the failure of these progenitor cells to proliferate or survive (Tsai *et al* 1994, Tsai *et al* 1997). *GATA2* expression decreases in the late stages of haematopoiesis, during differentiation (Cheng *et al* 1996).

The other GATA family members involved in haematopoietic regulation, *GATA1* and *GATA3*, have distinct patterns of expression. *GATA1* expression partially overlaps with *GATA2* expression, and is seen in erythrocytes, megakaryocytes, mast cells, eosinophils and basophils (Yamaguchi *et al* 1998). Its expression is considered essential for maturation of erythroid and megakaryocytic lineages and platelet formation (Tsai *et al* 1998, Fujiwara *et al* 1996, Pevny *et al* 1991, Pevny *et al* 1995, Shivdasani 1997). *GATA3* expression is mostly restricted to the T-lymphocyte lineage, where it is required for early differentiation (Ting *et al* 1996), but it is also seen in mast cells and eosinophils (Yamaguchi *et al* 1998).

3.5.5.2 *GATA2* Expression in Patients with 3q21 Rearrangements

Examination of the expression of *GATA2* in the panel of leukaemia patients demonstrated that it was expressed at high levels in both t(1;3)(p36;q21) patients examined, while it was expressed in normal bone marrow or blood at low levels, as expected due to the expression of *GATA2* in normal haematopoietic progenitor cells. As *GATA2* is located on 3q21, it is also a candidate for dysregulation in t(3;3)(q21;q26) and inv(3)(q21q26). Expression of *GATA2* was therefore examined in one t(3;3)(q21;q26) patient and two inv(3)(q21q26) patients. A high level of expression was found in all three cases. However, *GATA2* was also shown to be expressed in several AML and MDS patients without 3q26 rearrangements. These patients do not display the distinct clinical phenotype associated with the t(1;3)(p36;q21), t(3;3)(q21;q26) and inv(3)(q21q26). This shows that, as is the case for *MEL1* and *EVI1*, *GATA2* expression alone is not responsible for imparting this phenotype.

GATA2 expression in AML has been frequently reported (Nagai *et al* 1994, Minegishi *et al* 1997, Tsuzuki *et al* 2000). This may simply reflect the progenitor nature of the target cell of leukemic transformation (Tsuzuki *et al* 2000). The results presented here show that the level at which *GATA2* is expressed appears to be greater in patients bearing the t(1;3)(p36;q21), the t(3;3)(q21;q26) or the inv(3)(q21q26), however, suggesting that upregulation of *GATA2* expression is a genuine effect of translocations involving 3q21.

3.5.5.3 Co-expression of *GATA2* with *MEL1* or *EVI1*

As *GATA2* expression, like *MEL1* or *EVI1* expression, occurs in patients without chromosome rearrangement or the clinical phenotype of dysmegakaryopoiesis and elevated platelets, its expression alone is insufficient to establish this phenotype. However, the possibility that *GATA2* expression in combination with either *MEL1* or *EVI1* expression, which would be the expression pattern in most patients with either t(1;3)(p36;q21), t(3;3)(q21;q26) or inv(3)(q21q26), may account for this phenotype was further explored.

Correlation of the expression patterns reported in Sections 3.4.1, 3.4.4 and 3.4.5 (see Table 3.1) revealed some patients who expressed *GATA2* and either *MEL1* or *EVI1* (see Table 3.1). Examination of these patients' clinical records showed that all had decreased platelet counts, and there was no report of dysmegakaryopoiesis. It must be concluded, therefore, that *GATA2* expression in combination with either *MEL1* or *EVI1* expression is unable to account for the leukemic phenotype of patients with 3q21 rearrangements.

3.5.5.4 Mechanism of Dysregulation of *GATA2* Expression

GATA2 knockout mice fail to develop beyond mid-gestation due to haematopoietic failure (Tsai *et al* 1994). In the process of rescuing this phenotype in order to study other effects of the knockout, Zhou *et al* (1998) demonstrated that for normal haematopoietic expression of *GATA2* to be restored, sequences between 100 and 150 kb 5' of the *GATA2*

transcription start site were required. Although the existence and position of similar sequences has not been demonstrated in humans, it is reasonable to suppose that such a regulatory element exists. The 3q21 breakpoint in Patient 1 was mapped 6 kb 5' of *GATA2*, between the proposed regulatory sequences and the start of the gene. Therefore, the translocation may have the effect of physically separating the *GATA2* gene from a haematopoietic regulatory element. Indeed, this is the case for all known 3q21 breakpoints in either t(1;3)(p36;q21) (Mochizuki *et al* 2000) or t(3;3)(q21;q26) or inv(3)(q21q26) translocations (reviewed in Wieser 2002). Disruption of haematopoietic regulation of the *GATA2* gene may be a common feature of all patients with t(1;3)(p36;q21), t(3;3)(q21;q26) and inv(3)(q21q26). As these patients share a common leukemic phenotype and a common translocation breakpoint, it seems likely that the effect of the chromosome rearrangements at this breakpoint, namely transcriptional upregulation of *GATA2*, contributes to this phenotype.

3.5.5.5 GATA Family Member Overexpression Studies

Given that overexpression of *GATA2* is associated with the t(1;3)(p36;q21), t(3;3)(q21;q26) and inv(3)(q21q26), it is instructive to investigate the effects of overexpression of *GATA2* in haematopoietic cells. Persons *et al* (1999) conducted a study overexpressing *GATA2* in haematopoietic stem cells (HSCs) and found that this resulted in blockage of both proliferation and differentiation of these cells, although the cells did not apoptose. A block in differentiation fits well with the hypothesis that overexpression of *GATA2* has an oncogenic effect, although the block in proliferation does not.

Overexpression of *GATA2* in the erythroleukaemia cell line K562, which is more differentiated than HSCs, caused the cells to shift in phenotype toward the megakaryocyte lineage (Ikonomi *et al* 2000). Similar overexpression studies performed in primary erythroid progenitor cells caused a decrease in the rate of cellular proliferation and maturation but no lineage alteration (Ikonomi *et al* 2000).

Overexpression studies performed using other GATA family members have shown that *GATA1* overexpression in the murine myeloid leukaemia cell line M1 induced megakaryocytic and erythroid differentiation (Yamaguchi *et al* 1998), and *GATA3* expression in HSCs also induced megakaryocytic and erythroid differentiation (Chen and Zhang 2001), despite normal *GATA3* function being restricted to the lymphoid lineage. Visvader *et al* (1993) demonstrated that forced expression of any of *GATA1*, *GATA2* or *GATA3* in a murine myeloid cell line induced megakaryocytic differentiation. These studies suggest at least a degree of functional redundancy between the three proteins.

3.5.5.6 Functional Redundancy of GATA Family Genes

The fact that GATA proteins all recognise and bind the same DNA motif in gene promoters *in vivo* (Weiss *et al* 1997) suggests that they may be functionally redundant, and that the specific roles that each gene plays in haematopoietic regulation may be due to temporally and spatially specific expression patterns rather than functional specificity. This hypothesis is supported by several observations arising out of GATA gene family

overexpression studies (Visvader *et al* 1992, Chen and Zhang 2001, discussed above). Other studies have addressed this question directly, as outlined below.

Tsai *et al* (1998) constructed *GATA3* knock-in mice, such that the mice lacked *GATA1* but expressed *GATA3* under the control of the *GATA1* regulatory sequences. This resulted in partial rescue of the *GATA1* knock-out phenotype, but could not restore terminal differentiation of the megakaryocytic and erythroid lineages. The incompleteness of the phenotype could be due to differential regulation of *GATA1* and *GATA3* at the translational or post-translational level, or qualitative differences in the two proteins' transactivation capacity. While these mice were only partially rescued and still died prenatally due to haematopoietic failure, Takahashi *et al* (2000) performed similar experiments in which *GATA1* deficient mice were rescued by the expression of either *GATA2* or *GATA3* under *GATA1* transcriptional regulatory controls. In these experiments, the mice survived to adulthood. The only abnormality displayed in the adult mice was abnormal erythropoiesis. These results demonstrate partial functional redundancy of the GATA family proteins.

There is also evidence that the GATA family proteins do have distinct functions. There appear to be some differences in the DNA-binding sequence preferences of the GATA family members, with GATA2 and GATA3 also able to bind the DNA sequence AGATCTTA, while GATA1 cannot bind this motif (Ko and Engel 1993). Yamaguchi *et al* (1998) demonstrated that, although GATA1 and GATA2 could both bind the GATA motif in the *Granule Major Basic Protein* (*MBP*) gene promoter, only GATA1 caused

transactivation of the gene. The proposed mechanism for this difference is in the different abilities of the proteins to attract transcriptional cofactors. The GATA family proteins conduct a wide variety of protein-protein interactions, with each other (Crossley *et al* 1995) and other transcription factors and cofactors (reviewed in Cantor and Orkin 2001). The extent to which these interactions are specific to the GATA family member in question is unknown. Nevertheless, it seems clear that there is a degree of functional redundancy between these three GATA binding proteins, and that the timing and lineage-specificity of their expression is as important as the functional specificity of each protein.

The partial functional redundancy shared by the GATA family proteins is of interest when considering the upregulation of *GATA2* as a result of any of the t(1;3)(p36;q21), t(3;3)(q21;q26) or inv(3)(q21q26). The expression of *GATA2* which results from the rearrangement is ectopic, at least in leukaemias deriving from haematopoietic cells which have begun to differentiate and so have inactivated transcription of *GATA2*. However, the cellular context in which *GATA2* is ectopically expressed in these leukaemias is similar to that in which *GATA1* is normally expressed due to its involvement in the differentiation process. Perhaps *GATA2* expression, in performing part of the function normally performed by *GATA1*, affects the differentiated phenotype of the cells and so contributes to the abnormal morphology, especially dysmegakaryopoiesis, associated with translocation-bearing cells, as *GATA1* is involved in megakaryocytic determination (Matsumura and Kanakura, 2002).

For this reason, it is possible that the aberrant transcription of *GATA2* seen in patients carrying either t(1;3)(p36;q21) or t(3;3)(q21;q26) may see the GATA2 protein adopt part of the normal function of *GATA1* protein; that is, that it may promote differentiation of the cell along the megakaryocytic pathway. This may account for the phenotype observed in these patients of abnormal megakaryopoiesis (as megakaryopoiesis would be occurring in cell types in which it does not normally occur) and increased platelet counts (as more cells would be forced down the megakaryocytic lineage, which would ultimately produce more platelets). However, *GATA1* involvement in megakaryocytic development is complex. It appears to be a negative regulator of early megakaryocytic proliferation, while at later stages of development it is thought to promote differentiation (Shivdasani 2001). The degree to which *GATA2* can substitute for it is equally unclear, and it is also the case that *GATA2* expression overlaps with *GATA1* expression in normal megakaryocytic progenitors to an extent (Shivdasani 2001). Despite this complexity, it remains possible, even likely, that aberrant expression of *GATA2* in haematopoietic cells may promote dysmegakaryopoiesis.

It remains true that the most likely molecular explanation for the clinical phenotype associated with the 3q21 rearrangements in AML is due to the molecular effect which the translocation has on 3q21 genes. In the results presented in Section 3.4.4, there appeared to be a trend in which 3q21 patients expressed *GATA2* at a level higher than other *GATA2*-positive AML patients. This correlation was not perfect, but a larger sample may demonstrate this observation more conclusively. GATA2 protein may have a different effect at higher concentrations, as suggested by the overexpression studies previously discussed (Persons *et al* 1999, Ikonomi *et al* 2000). If this is the case, higher levels of

GATA2 expression caused by any of the t(1;3)(p36;q21), t(3;3)(q21;q26) or inv(3)(q21q26) could be responsible for dysmegakaryopoiesis and elevated platelet counts in patients with the translocation, but the lower levels expressed in non-translocation patients may be insufficient for this phenotype to develop.

3.5.6 Interactions between GATA family members and PRDM family members

As *MEL1* and *GATA2* are upregulated as a result of the t(1;3)(p36;q21), and *EVI1* and *GATA2* are upregulated as a result of the t(3;3)(q21;q26) or inv(3)(q21q26), it is of major interest that members of the two families to which the three genes belong, the GATA family and the PRDM family, appear to interact.

As transcription factors, both GATA proteins and PRDM proteins bind specific promoter sequences and influence transcription of the downstream genes. The GATA family bind the consensus DNA sequence WGATAR (where W is an A or T and R is an A or G) (Wall *et al* 1988, Evans *et al* 1988). EVI1, the best-characterised PRDM protein, has two DNA-binding domains distinguished by the number of Zn finger structures present. The seven-Zn finger domain binds the recognition sequence GAYAAGATAAGATAA (where Y is a C or a T) (D1-CONS) (Delwel *et al* 1993) while the three-Zn finger domain recognises GAAGATGAG (D2-CONS) (Morishita *et al* 1995). The WGATAR motif, therefore, is contained within the recognition sequence for the seven-Zn finger domain of EVI1, and so it is a possibility that there is competition for binding at EVI1 recognition sites between EVI1 and GATA family proteins. Indeed it has been demonstrated that GATA1 can

transactivate a promoter containing the D1-CONS element, and that EVI1 can repress this transactivation (Soderholm *et al* 1997, Perkins *et al* 1996). As EVI1 is generally believed to act as a repressor but MDS1-EVI1 is a transactivator, this is especially pertinent to the case of translocation-mediated upregulation of *EVI1*, because translocation patients generally do not express *MDS1-EVI1*, as the breakpoint lies between *MDS1* and *EVI1* (Soderholm *et al* 1997, Nucifora 1997).

However, it is difficult to interpret this interaction in terms of leukemogenesis. How can *GATA2* upregulation as a result of translocation contribute to leukemogenesis when the same translocation event causes upregulation of *EVI1*, which represses transactivation by the GATA2 protein? One answer to this is that EVI1 is more selective in its DNA-binding recognition sequence (that is, the required recognition sequence is longer), and so will only be able to bind a subset of GATA-element containing promoters. Indeed, there is no known GATA-responsive site which also conforms to the EVI1 recognition sequence (Perkins *et al* 1996). If this is the case, the competition for binding sites may be meaningless in the context of regulation of *in vivo* target gene expression.

A similar competitive binding situation may exist for *GATA2* and *MEL1* in the case of the t(1;3)(p36;q21). No DNA binding study of the MEL1 protein has been reported, however it seems likely that a similar recognition sequence to that of EVI1 will eventually be identified. If this is the case it will be significant to ascertain whether the MEL1 protein is capable of acting as a repressor of transcription as well as the transactivator that it is presumed to be. Unlike some translocations involving *EVI1*, the breakpoint in the

t(1;3)(p36;q21) does not interrupt the PR domain of *MEL1*, such that expression of full-length transcript, encoding PR domain-containing protein, is a possible outcome of the translocation. In order to answer these questions, DNA-binding and transactivation studies of *MEL1* must be performed, as must an analysis of the nature of *MEL1* transcripts resulting from the translocation (as well as in the normal cell).

There is some evidence to suggest another level of interaction between GATA family members and PRDM family members. Shapiro *et al* (1995) conducted a screen to identify binding partners of the GATA3 protein. They identified a novel protein for which they obtained only partial sequence (accession U23736). BLAST analysis of this sequence reveals that it is identical to the 3' sequence of RIZ. This therefore is evidence for direct protein-protein interaction between GATA3 and RIZ, and it is tempting to suggest that there may be interactions between other members of both families, including GATA2, EVI1 and MEL1. Whether this is indeed the case, and what the nature of these interactions might be, awaits further investigation.

3.6 Conclusions

The 1p36 translocation breakpoint in Patient 1 was mapped and shown to be 43 kb 5' and telomeric of the transcription start site of *MEL1*. This was similar to the findings of Mochizuki *et al* (2000) in four patients with cytogenetically identical translocations. The 3q21 breakpoint in Patient 1 was shown to be 124 kb 3' of *RPN1* and 6 kb 5' of *GATA2*. This is in accordance with other 3q21 breakpoints involved in t(1;3)(p36;q21), t(3;3)(q21;q26) or inv(3)(q21q26).

MEL1 expression was found to be expressed at very low levels in normal bone marrow, possibly representing expression of *MEL1* by a small subset of haematopoietic progenitor cells. This result was in contrast to the findings of Mochizuki *et al* (2000), who found that *MEL1* was not expressed in normal bone marrow.

MEL1 expression was upregulated in both patients with t(1;3)(p36;q21) examined, but it was also observed in several other AML and MDS patients without 1p36 chromosomal abnormalities. A correlation was observed between t(1;3)(p36;q21) and *MEL*1 expression, suggesting that there are two distinct mechanisms of *MEL*1 expression in AML.

One of these mechanisms is chromosome aberration, but there are two possible explanations for *MEL1* expression in patients without chromosome aberration. Firstly, there may be a second mechanism of *MEL1* upregulation, such as the overexpression of a

MEL1-activating transcription factor. Secondly, the perceived upregulation of *MEL1* may actually be the result of the expansion of a haematopoietic cell population which normally expresses *MEL1*. This hypothesis is supported by the fact that the AML patients who were found to express *MEL1* all had AML M1 or M2, derived from an early haematopoietic progenitor cell. Also, all non-t(1;3)(p36;q21) patients who expressed *MEL1* also expressed *GATA2*, which is known to be expressed in a subpopulation of early haematopoietic cells. *MEL1* expression was also seen in two patients with distinct 1p36 abnormalities. These patients were not included in the statistical analysis and are investigated in the following two chapters.

Expression analysis of other genes around the Patient 1 breakpoint demonstrated that *ARPM2* and *ARHGEF16* were not involved in a translocation-mediated leukemogenic mechanism. *GATA2* was found to be upregulated in both patients with t(1;3)(p36;q21), as well as patients with t(3;3)(q21;q26) and inv(3)(q21q26). Some patients without 3q21 abnormalities, however, were also found to express *GATA2* at elevated levels. Similarly to *MEL1* expression, this *GATA2* expression could be due to expansion of a subpopulation of cells which normally express *GATA2*, or the expression could be due to an oncogenic mechanism. As *GATA2* is known to be expressed in normal haematopoietic progenitor cells, the expansion of a normally expressing subpopulation due to leukemic transformation is a likely explanation. This is supported by the fact that every patient who expressed *MEL1* also expressed *GATA2*, although not all *GATA2* expressing patients expressed *MEL1*. This suggests that a subset of normal *GATA2* expressing haematopoietic cells also express *MEL1*. Expression of *EVI1* was also seen in all patients with t(3;3)(q21;q26) and inv(3)(q21q26), but also in some patients without a 3q26

rearrangement. This expression did not correlate with AML subtype, *GATA2* expression or *MEL1* expression.

None of the patients without a 1p36, 3q21 or 3q26 rearrangement showed the clinical phenotype of normal or elevated platelet counts and dysmegakaryopoiesis with which t(1;3)(p36;q21), t(3;3)(q21;q26) and inv(3)(q21q26) patients present. Therefore, neither *MEL1* expression, *GATA2* expression or *EVI1* expression could account for this phenotype. However, if the expansion hypothesis does explain *GATA2* expression and *MEL1* expression in non-translocation patients, it may be that the expression of either or both of these genes, mediated by translocation in a cell type in which they are not normally expressed, can cause or contribute to this phenotype.

Quantitative analysis showed that there may be a trend in which patients with 3q21 translocations express *GATA2* more strongly than patients who have *GATA2* upregulation without 3q21 translocation. If this trend was irrefutable, there would be an argument for a threshold-level effect of *GATA2* transcription being causative of the clinical phenotype associated with the 3q21 translocations. As this trend was not completely consistent, however, this argument awaits the *GATA2* expression analysis of more patients with and without 3q21 breakpoints.

CHAPTER 4

The

ins (12;1)(p13;p36p21)

Rearrangement

CHAPTER 4

ins(12;1)(p13;p36p21)

4.1 Rearrangements of 1p36, 12p13 and 1p21 in AML & MDS

4.1.1 Background

This chapter details an investigation of a complex rearrangement, an inverted insertion of the 1p21 - 1p36 region into 12p13. Translocations involving 12p13 are among the most frequent translocations in AML and MDS (Rowley 1999, Mitelman 2000), and often result in fusion transcripts involving the *Translocation-ETS Leukaemia* (*TEL*) gene (reviewed in Bohlander 2000). *TEL,* also known as *ETS Variant 6* (*ETV6*) gene, is a transcription factor involved in the regulation of haematopoietic development (Wang *et al* 1998), but was first identified as the fusion partner of the *Platelet Derived Growth Factor Receptor ß* (*PDGFRß*) gene as a result of the t(5;12)(q33;p13) in chronic myelomonocytic leukaemia (Golub *et al* 1994). Many of the translocations involving *TEL* have been characterised, and a list of the *TEL* fusion partner genes that have been identified is presented in Table 4.1

The TEL protein contains two important domains: a C-terminal E26 transformation-specific (ETS) domain (a DNA-binding domain; Sharrocks 2001), and an N-terminal helix-loop-helix (HLH) domain (or the pointed (PNT) domain), a domain which

Partner Gene	Translocation	Reference	Disease(s)
ARNT	t(1;12)(q21;p13)	Salomon-Nguyen *et al* 2000	AML
ARG	t(1;12)(q25;p13)	Cazzaniga *et al* 1999	AML
MDS1/EVI1	t(3;12)(q26;p13)	Peeters *et al* 1997a	AML, MDS
FGFR3	t(4;12)(p16;p13)	Yagasaki *et al* 2001	peripheral T cell lymphoma
CHIC2/BTL	t(4;12)(q12;p13)	Cools *et al* 1999	AML
ACS2	t(5;12)(q31;p13)	Yagasaki *et al* 1999	CML
PDGFRß	t(5;12)(q33;p13)	Golub *et al* 1994	AML, MDS
STL	t(6;12)(q23;p13)	Suto *et al* 1997	ALL
HLXB9	t(7;12)(q36;p13)	Beverloo *et al* 2001	AML
JAK2	t(9;12)(p24;p13)	Peeters *et al* 1997b	ALL, CML
PAX5	t(9;12)(q11;p13)	Cazzanigga *et al* 2001	ALL
ABL	t(9;12)(q34;p13)	Papadopoulos *et al* 1995	ALL, CML
CDX2	t(12;13)(p13;q12)	Chase *et al* 1999	AML, ALL
TRKC	t(12;15)(p13;q25)	Eguchi *et al* 1999	AML
AML1	t(12;21)(p13;q22)	Golub *et al* 1995, Romana *et al* 1995	ALL, AML
MN1	t(12;22)(p13;q11)	Buijs *et al* 1995, 2000	myeloid leukaemia

Table 4.1. The fusion partners of the *TEL* oncogene. The *TEL* oncogene forms fusion genes with many partner genes. This list was compiled by literature search.

facilitates protein homodimerisation and confers transcriptional repression activity (Fenrick *et al* 1999). The fusion proteins of TEL are especially interesting as they have been shown to contribute to the progression of leukaemia in at least three distinct ways (Buijs *et al* 2000, reviewed in Cools *et al* 2002). These are summarised below.

The first leukemogenic mechanism of TEL fusion proteins is represented by the most frequently occurring *TEL* translocation, the t(12;21)(p13;q22), which occurs in approximately 20 % of ALL patients (Look 1997) and is also known to occur in AML (Rubnitz *et al* 1999). This translocation forms the *TEL-AML1* fusion gene (Golub *et al* 1995, Romana *et al* 1995). *AML1* is a transcription factor which is responsible for the activation of numerous genes involved in haematopoiesis (reviewed in Speck 2001). *TEL* is known to repress the activity of at least some of the promoters that are activated by *AML1* (Fears *et al* 1997). The *TEL-AML1* fusion protein is therefore thought to act principally by blocking transcription of myeloid genes which are normally activated by *AML1* (Hiebert *et al* 1996, Song *et al* 1999).

Secondly, the *MN1-TEL* (Buijs *et al* 1995) and *BTL-TEL* (Cools *et al* 1999) fusion genes are unique among *TEL* fusions in that they contain only the 3' end of *TEL*, whereas all other known *TEL* fusions contain only the 5' end of *TEL*. The MN1-TEL and BTL-TEL fusion proteins therefore incorporate the DNA-binding portion of TEL but not the PNT protein interaction domain, inhibiting the recruitment of co-repressing molecules that is performed by the normal TEL protein (Chakrabati and Nucifora 1999, Lopez *et al* 1999). The addition of the transactivation domain from MN1 effectively converts TEL from a repressor into an activator of transcription with

transforming ability (Buijs *et al* 2000). The transforming potential of the *BTL-TEL* fusion gene is yet to be investigated and there is some doubt that the fusion gene is always expressed in patients presenting with the t(4;12)(q12;p13) (Cools *et al* 2002).

The third mechanism of action by a *TEL* fusion gene is exemplified by the *TEL-MDS1-EVI1* fusion gene, in which *TEL* contributes only its first one or two exons to the fusion protein (Hoglund *et al* 1996, Peeters *et al* 1997a). Therefore, no functional domain from *TEL* is present in the fusion gene. The function of this fusion protein is expected to be the same as that of the normal MDS1-EVI1 protein, and the leukemogenic effect of the translocation to be a result of the constitutive transcription of the *MDS1-EVI1* gene from the *TEL* promoter in haematopoietic cells (Peeters *et al* 1997a). This mechanism is shared by the *TEL-CDX2* fusion gene formed by the t(12;13)(p13;q12) (Chase *et al* 1999). *CDX2* is not expressed in normal bone marrow (Chase *et al* 1999). *MDS1-EVI1* is expressed at very low levels in normal bone marrow (Soderholm *et al* 1997).

Although *TEL* is commonly affected by 12p13 rearrangements, one study has suggested that approximately half of the translocations involving 12p13 in haematological malignancy do not involve a *TEL* fusion gene, and that there are at least two other breakpoint cluster regions in 12p13 (Sato *et al* 1997).

Rearrangements of both 1p36 and 12p13 individually are frequent in haematological malignancies, but there is only one report, published during the work presented in this thesis, of a rearrangement between 1p36 and 12p13, a single case of t(1;12)(p36;p13) in a 66 year old woman with MDS (RAEBT) (Odero *et al* 2001). Molecular

characterisation *via* FISH analysis of this rearrangement showed that the 12p13 breakpoint was within intron 2 of the *TEL* gene, but the precise position of the 1p36 breakpoint was not investigated (Odero *et al* 2001).

The Mitelman Database of Chromosome Aberrations in Cancer (http://cgap.nci.nih.gov/Chromosomes/Mitelman) lists no recurrent translocations involving 1p21 in leukaemia. Similarly, no 1p21 rearrangements in leukaemia are reported in the literature.

4.1.2 Patient Information

Patient 2 was a 24 year old male who presented in September 1993 with diahorrea, vomiting and lethargy. Analysis of a bone marrow aspirate led to the diagnosis of AML M1, with 72 % blasts. The blasts had a pleomorphic appearance, and normal haematopoiesis was markedly decreased. The patient was treated with chemotherapeutic agents beginning upon diagnosis. This treatment was successful for a time, but the patient died in May 1996.

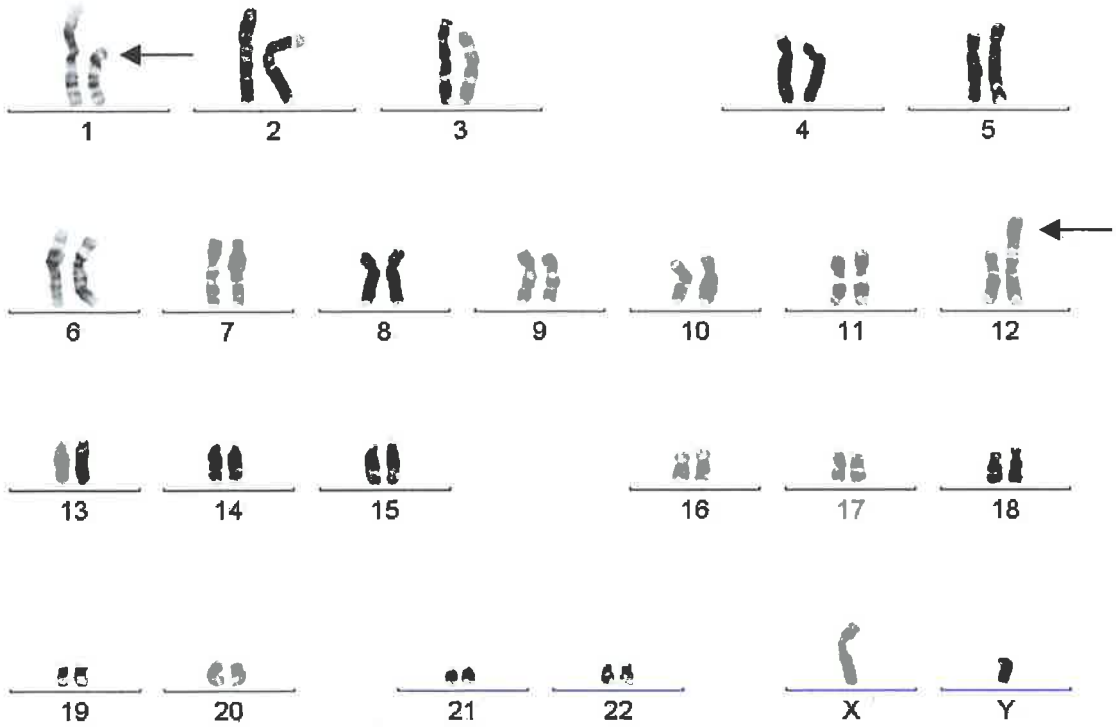
Cytogenetic analysis of Patient 2 bone marrow material was performed on presentation samples in September 1993. Twenty-eight metaphases were examined, and of these, 24 were abnormal. The abnormal karyotype was interpreted as 46, XY, ins(12;1)(p13;p36p21). At the outset of the present study, samples were sent to be karyotyped again. This analysis agreed with the earlier assessment, and this abnormal karyotype is shown in Figure 4.1.

Figure 4.1. Patient 2 Karyotype.

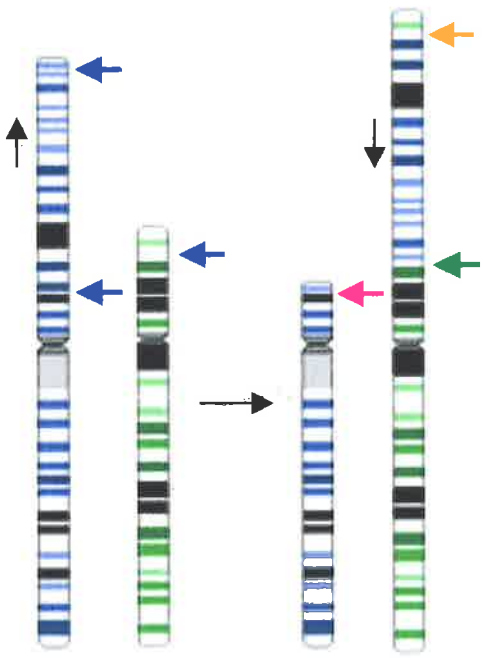
A) The aberrant karyotype of the leukemic blasts of Patient 2, showing 46, XY, inv ins (1;12)(p21-p36;p13). Karyotype analysis was performed by Sarah Moore (Institute of Medical and Veterinary Science, Adelaide, Australia).

B) Ideogram representing the ins inv (1;12)(p21-p36;p13) karyotype. The normal chromosomes are on the left of the horizontal black arrow. Derivative chromosomes are on the right of the horizontal black arrow. Chromosome 1 material is highlighted with blue sub-bands. Chromosome 12 material is highlighted with green sub-bands. The vertical black arrows indicate the orientation of the 1p21-1p36 section before and after rearrangement. The blue arrows indicate the 1p36, 1p21 and 12p13 breakpoints on the normal chromosomes. The pink arrow indicates the telomeric 1p36 - centromeric 1p21 junction. The green arrow indicates the centromeric 1p36 - centromeric 12p13 junction. The orange arrow indicates the telomeric 1p21 - telomeric 12p13 junction.

A)



B)



Prior to the commencement of the present study, preliminary FISH analysis was performed on Patient 2 material. These experiments positioned the 1p36 breakpoint telomeric of a PAC clone sequence located at 1p36.2 (Varga *et al* 2001).

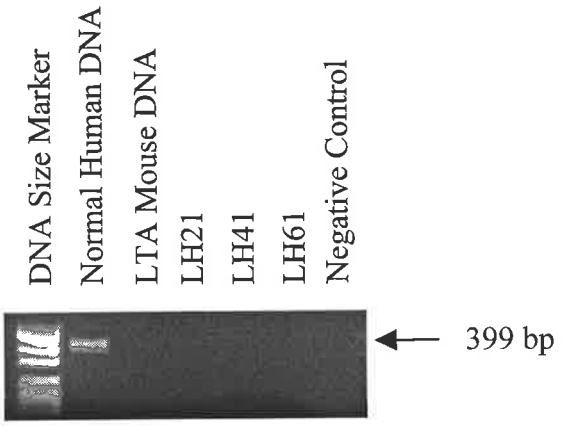
4.2 Somatic Cell Hybrid Analysis

Somatic cell hybrid lines were constructed (see Section 2.2.5.4) by fusion of the LTA murine fibroblast cell line with bone marrow cells collected from Patient 2 at presentation and frozen until the present study. Three colonies were obtained following the hybridisation, and these were expanded and screened by PCR for the presence of translocation-derived chromosomes.

4.2.1 Chromosome 1 Analysis

The three hybrid cell lines were initially screened with PCR markers D1S468 (1p36.3, telomeric marker) and D1S2661 (1p32, centromeric marker). All three hybrid cell lines were negative for both markers and it was concluded that they contained no chromosome 1 material (see Figure 4.2). However, with the publication of the Mochizuki *et al* (2000) study detailing the involvement of *MEL1* in the recurrent translocation t(1;3)(p36;q21) in AML and MDS, it became apparent that the initial screening did not include the *MEL1* locus, as the telomeric marker, D1S468, is centromeric to *MEL1* (D1S468 is within an intron of the *TP73* gene; see Section 3.2.1). The three hybrid cell lines were therefore rescreened with a more telomeric marker, 425F3Z, which is known to be telomeric to *MEL1* (see Section 3.2.1). One of the hybrid cell lines, LH61, was positive for this marker (see Figure 4.2). This

A) D1S2661 Screen.



B) D1S468 Screen.

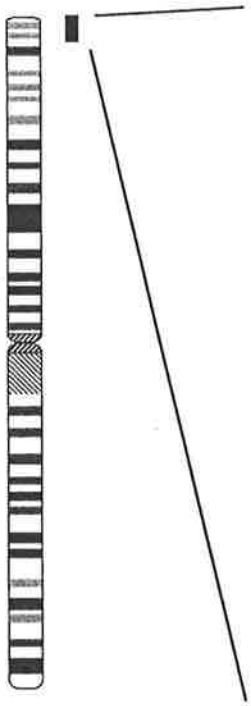
C) 425F3Z Screen.

Figure 4.2. Chromosome 1 Marker Screen of LH Hybrid Cell Lines.

Results of testing for presence or absence of chromosome 1 markers in the panel of somatic cell hybrids generated from Patient 2 bone marrow material. D1S2661 is located at 1p32 and so is centromeric of the proposed breakpoint. 4253Z is at the telomeric end of 1p36. D1S468 is between the other two markers but is at 1p36.3 and was initially used as the telomeric marker. Only one of the hybrids, LH61, retains any chromosome 1 material. LH61 is positive for 425F3Z but negative for both D1S468 and D1S2661, indicating that it contains the telomeric end of chromosome 1, and that the breakpoint at 1p36 lies between 425F3Z and D1S468.

indicated that LH61 retained the telomere of chromosome 1, which is present on the der 1 chromosome if the cytogenetic interpretation of the karyotype is correct (see Figure 4.1).

This result determined that the breakpoint lay between 425F3Z and D1S468, two markers that are separated by less than 1 Mb. It was therefore possible that the 1p36 breakpoint on chromosome 1 in Patient 2 was in a similar position to those in Patient 1 and the t(1;3)(p36;q21) patients reported in Mochizuki *et al* (2000) (see Section 3.2.1). A series of markers based on the Celera genomic scaffold GA_x2HTBKR38EG, which contains the t(1;3)(p36;q21) breakpoint, were used to analyse the cell line LH61 in order to precisely locate the 1p36 breakpoint of Patient 2 (see Figure 4.3).



Marker	Position on Scaffold (kb)	Presence in LH61
425F3Z	9	+
38EG40	40	+
38EG76	76	+
38EG178	178	+
38EG183	183	+
38EG188	188	-
333E3	198	-
38EG357	357	-

Figure 4.3. Markers used in Chromosome 1 Breakpoint Analysis of LH61. The table shows the distance of each marker from the 5' end of the genomic scaffold GA_x2HTBKR38EG and the result obtained on analysis of LH61for each marker. + indicates that the LH61 retained the marker. - indicates the marker was not present in LH61.

Successive marker analyses narrowed the breakpoint to between the markers 38EG183 (positive, telomeric marker) and 38EG188 (negative, centromeric marker), which are approximately 5 kb apart. This region is approximately 22 kb further telomeric (and further 5' of *MEL1*) than the t(1;3)(p36;q21) breakpoint in Patient 1 (see Section 3.2.1).

The cytogenetic interpretation of the karyotype involves the 1p21 - 1p36 region of chromosome 1 inserting into 12p13 (see Figure 4.1). If this is correct, the der 1 chromosome will in effect be a deletion chromosome, with the 1p21 - 1p36 region deleted. As this is the chromosome which is apparently retained by LH61 (as only telomeric 1p36 markers are present in the hybrid), primers for *AMPD2*, a 1p12 marker were designed to test this hypothesis. PCR marker analysis of the LH61 hybrid, shown in Figure 4.4, revealed that it did retain the marker *AMPD2* as expected, supporting this interpretation of the karyotype. Further PCR walking analysis to narrow the 1p21 breakpoint was not attempted, as the 1p36 breakpoint was expected to be the more biologically significant breakpoint because 1p21 is rarely involved in rearrangements in leukaemia (see Section 4.1.1). Also, the 1p21 breakpoint was expected to be identified by an inverse PCR strategy (described in Sections 4.4 and 4.5).

4.3 Identification of 1p36-12p13 Breakpoint Sequence

The karyotype given in Section 4.1.2 shows the centromeric side of the 1p36 breakpoint recombining with the centromeric side of the 12p13 breakpoint (see Figure 4.1B). The 1p36 marker closest to the der1 (1p21-1p36) breakpoint on the

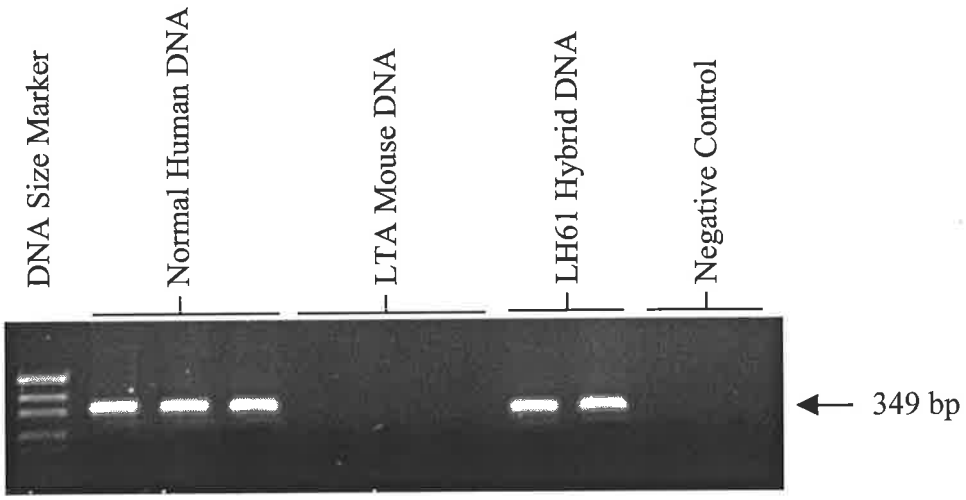


Figure 4.4. *AMPD2* Marker Analysis of LH61 Hybrids.

Results of testing the LH61 hybrid cell line for retention of the *AMPD2* marker. LH61 clearly does retain the marker, which is located at 1p12. In combination with the results presented in Figure 4.3, it shows LH61 to contain centromeric and telomeric sequences, but to have lost sequences in between. This result supports the interpretation of the Patient 2 karyotype as inv ins (12;1)(p13;p21-p36), as LH61 contains a deletion chromosome.

centromeric side (that is, negative in the hybrid containing the der1 chromosome), 38EG188, was used as a starting point to design an inverse PCR for use on Patient 2 DNA, to identify the 1p36-12p13 breakpoint (see Section 2.2.3.7). Restriction enzymes for use in the inverse PCR were chosen on the basis of availability and proximity of recognition sequences to 38EG188 on the centromeric side of this marker. The absence of recognition sequences on the telomeric side of 38EG188 was also taken into account. The three enzymes used in this instance were *Eco*NI, *Sac*II and *Xba*I. The design of this inverse PCR can be seen in Figure 4.5.

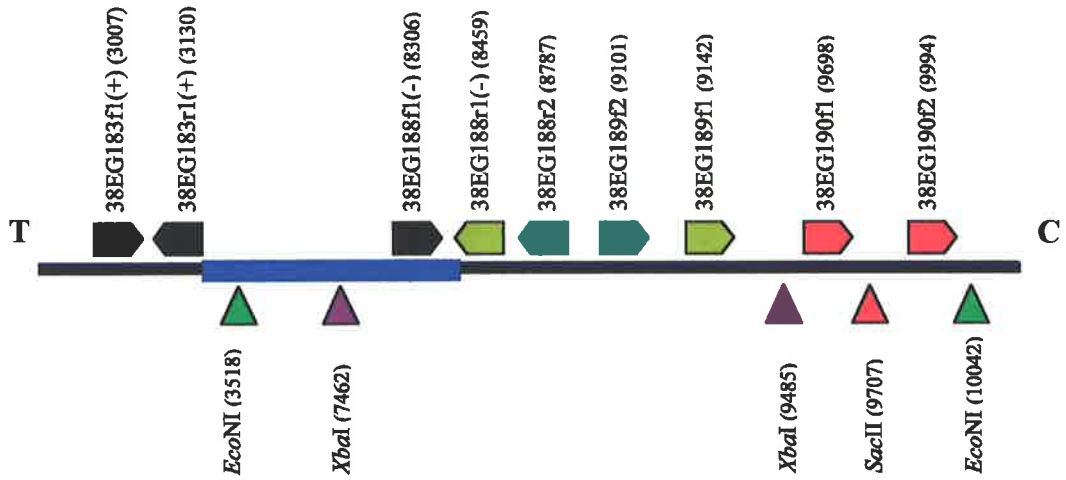


Figure 4.5. Design of Inverse PCR to Identify Partner Sequence of Centromeric 1p36 Breakpoint. The black line represents the genomic sequence of the 1p36 region. The letters T and C represent the telomeric and centromeric end of the region respectively. The blue box represents the breakpoint region as identified through somatic cell hybrid PCR walking. + and - indicate the presence or absence of markers tested against LH61. Primers used in either PCR walking, inverse PCR or inverse PCR sequencing are represented by pentagons above the line. Primers are colour-coded as follows: black for PCR walking primers, dark green for primary inverse PCR primers, light green for secondary inverse PCR primers and red for sequencing primers. Restriction sites utilised in the inverse PCR are represented by triangles below the line. Numbers in parentheses following a restriction enzyme site or primer name indicate the position of the object. Numbering is in accordance with RP11-333E3 (accession AL356984.13). The figure is not to scale.

Inverse PCR was performed as described in Section 2.2.3.7. Primers used were 38EG189f2 and 38EG188r2 for the primary reaction, and 38EG189f1 and 38EG188r1 for the nested secondary reaction. The result of this inverse PCR are shown in Figure 4.6.

Bands were generated from templates corresponding to each of the three enzymes. Predicted sizes for bands amplified from the normal chromosome were 6525 bp for *Eco*NI digested template and 2024 bp for *Xba*I digested template (see Figure 4.6). No normal product was expected from the *Sac*II digested template as the closest *Sac*II restriction site was outside of the amplifiable range. There were more than the expected two products (one from the normal chromosome and one from the translocated chromosome) generated from *Xba*I-digested template. These products were considered likely to be non-specific PCR artifacts and were not analysed further. Two products were amplified from the *Sac*II digested template. As no normal product was considered amplifiable, only one product, corresponding to the translocated chromosome, was expected. As either band could be derived from the translocated chromosome, both bands were therefore isolated by gel purification for sequencing. The single *Eco*NI product is smaller than the expected normal product, but this may be because the normal product was too large to amplify under the PCR conditions used. The band produced is therefore potentially amplified from the translocated chromosome and was isolated by gel purification for sequencing. Sequencing of all three bands was performed using the primer 38EG190f1. This primer is 9 bp from the *Sac*II cut site, and 344 bp from the *Eco*NI cut site (see Figure 4.5).

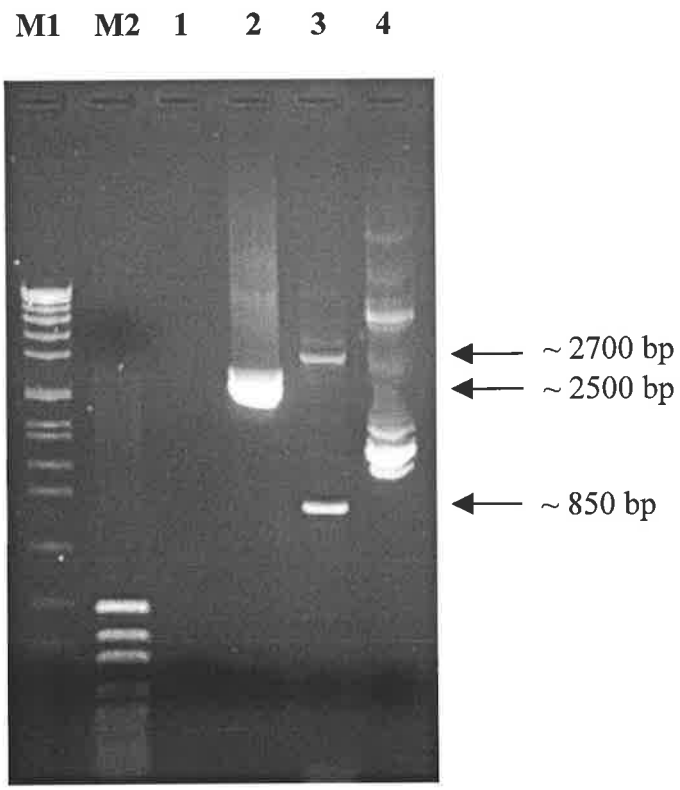


Figure 4.6. Inverse PCR to Identify Partner Sequence of Centromeric 1p36 Breakpoint.

Patient genomic DNA was prepared by digestion with the indicated restriction enzyme then circularised. The secondary PCR from templates prepared using each of the three restriction enzymes is shown. Primary PCR was performed using primers 38EG189f1 and 38EG188r2. Secondary PCR was performed using 38EG189f2 and 38EG188r1, and 2 µl of a 1:50 dilution of the primary PCR as template. The loading order is specified below.

M1. SppI/*Eco*RI DNA size marker.
M2. pUC19/*Hpa*II DNA size marker.
1. H_2O control.
2. *Eco*NI secondary inverse PCR.
3. *Sac*II secondary inverse PCR.
4. *Xba*I secondary inverse PCR.

Gel purification of the larger band in the *Sac*II lane did not give sufficient yield to allow sequencing of sufficient quality. However, the sequence data obtained from the smaller *Sac*II band indicated that the *Sac*II site which was expected to be cut was actually intact (data not shown). Subsequent investigation of the literature revealed that certain *Sac*II sites, apparently identical to other *Sac*II sites, are refractory to cutting (Oller *et al* 1991). This phenomenon is thought to be related to the fact that *Sac*II enzyme requires simultaneous interaction with two copies of the recognition sequence before DNA cleavage takes place, meaning that, in addition to the presence of the recognition sequence, cutting may also require the proximity of a second binding site (Kruger *et al* 1988, Conrad and Topal 1989). Analysis of the *Sac*II inverse PCR products was therefore taken no further.

Sequence analysis of the *Eco*NI inverse PCR product using the primer 38EG190f1 produced sequence data which corresponded to normal chromosome 1 sequence for approximately 125 bp, at which point the sequence contains a run of seventeen consecutive adenosines, following which the quality of the sequencing data rapidly diminished. This was possibly due to depletion of adenosine nucleotides in the Big Dye Terminator sequencing reaction mix. To overcome this problem, a new primer, 38EG190f2, was designed closer to the *Eco*NI cut site (47 bp; see Figure 4.5) to avoid the requirement of sequencing over the A-rich sequence. Sequencing performed with this primer produced good quality sequence data for approximately 500 bp. BLASTN analysis of this sequence showed that the first 32 bp corresponded to normal chromosome 1 sequence (from RP11-333E3, accession AL356984.13), including the recognition sequence for *Eco*NI, and the sequence after the cut site corresponded to a clone designated LL12NC01-95H4 (accession U81834.1), which is derived from

chromosome 12p13. An alignment of the sequence of the *Eco*NI inverse PCR product with the sequence of the two clones as determined by BLASTN analysis is presented in Figure 4.7.

Because this sequence does not correspond to the actual breakpoint of the patient, but rather to the ligation at a common restriction site, created during the construction of the inverse PCR template, it was important to amplify across the breakpoint directly from the patient material to confirm the inverse PCR result. Primers were therefore designed that would amplify from RP11-333E3 across the breakpoint to LL12NC01-95H4 from patient template DNA. The design of these primers is illustrated in Figure 4.8. Long template PCR was performed using these primers and the result of the PCR is shown in Figure 4.9. The PCR amplified a product from Patient 2 DNA and no product from normal human DNA. This result confirmed that there was a recombination of chromosome region 1p36 and 12p13 in Patient 2. The PCR product generated from Patient 2 material using primers specific for RP11-333E3 and LL12NC01-95H4 was sequenced to determine the precise sequence at the breakpoint. This data is presented in Figure 4.10. Comparison of the sequence of this product to the NCBI Human Genomic Sequence Assembly by BLASTN analysis confirmed that the chromosome 12 derived sequence runs centromere to telomere and the chromosome 1 derived sequence runs telomere to centromere (that is, 12C - 12T - 1T - 1C) along the derivative chromosome. The orientation of the rearranged chromosome is therefore consistent with expectations in the event of an inverted insertion.

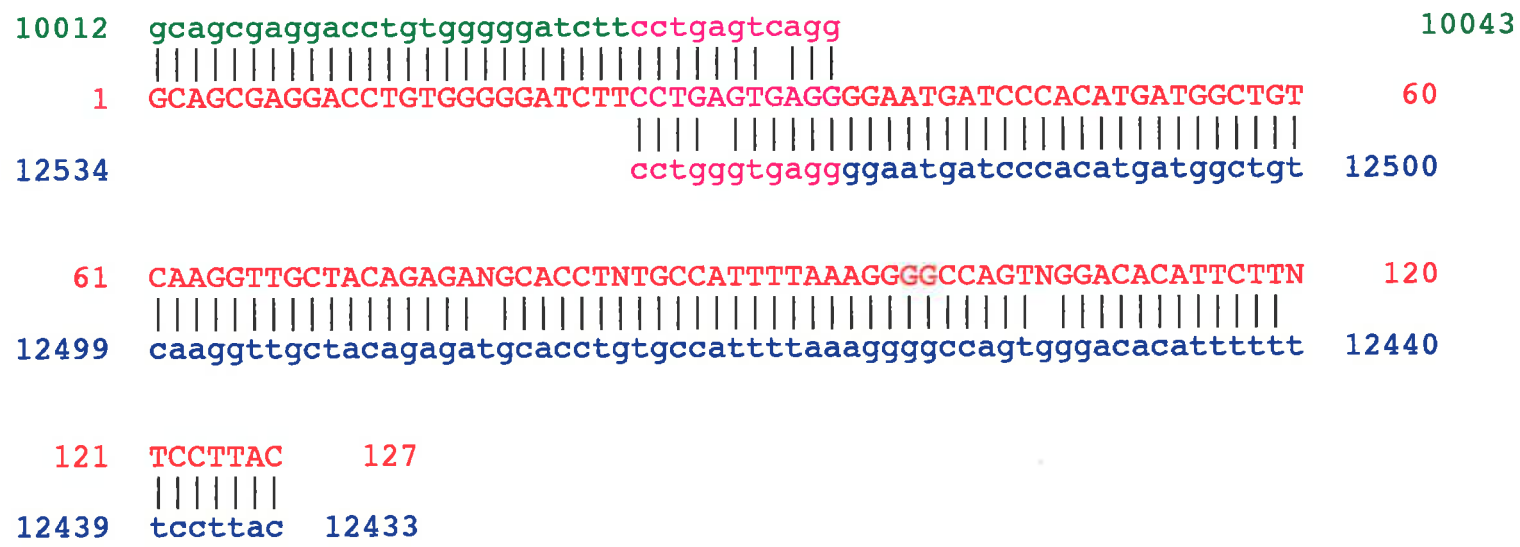


Figure 4.7. Sequence alignment of *Eco*NI inverse PCR product.

The sequence of the *Eco*NI inverse PCR product, as sequenced with the 38EG190f2 primer, is represented in red. It is aligned with sequences from 1p36 (RP11-333E3 (AL356984.13), represented in green) and 12p13 (LL12NCO1-95H4 (U81834.1), represented in blue). The presence of the *Eco*NI restriction site (CCTNNNNNAGG, where N represents any nucleotide) is highlighted in pink.

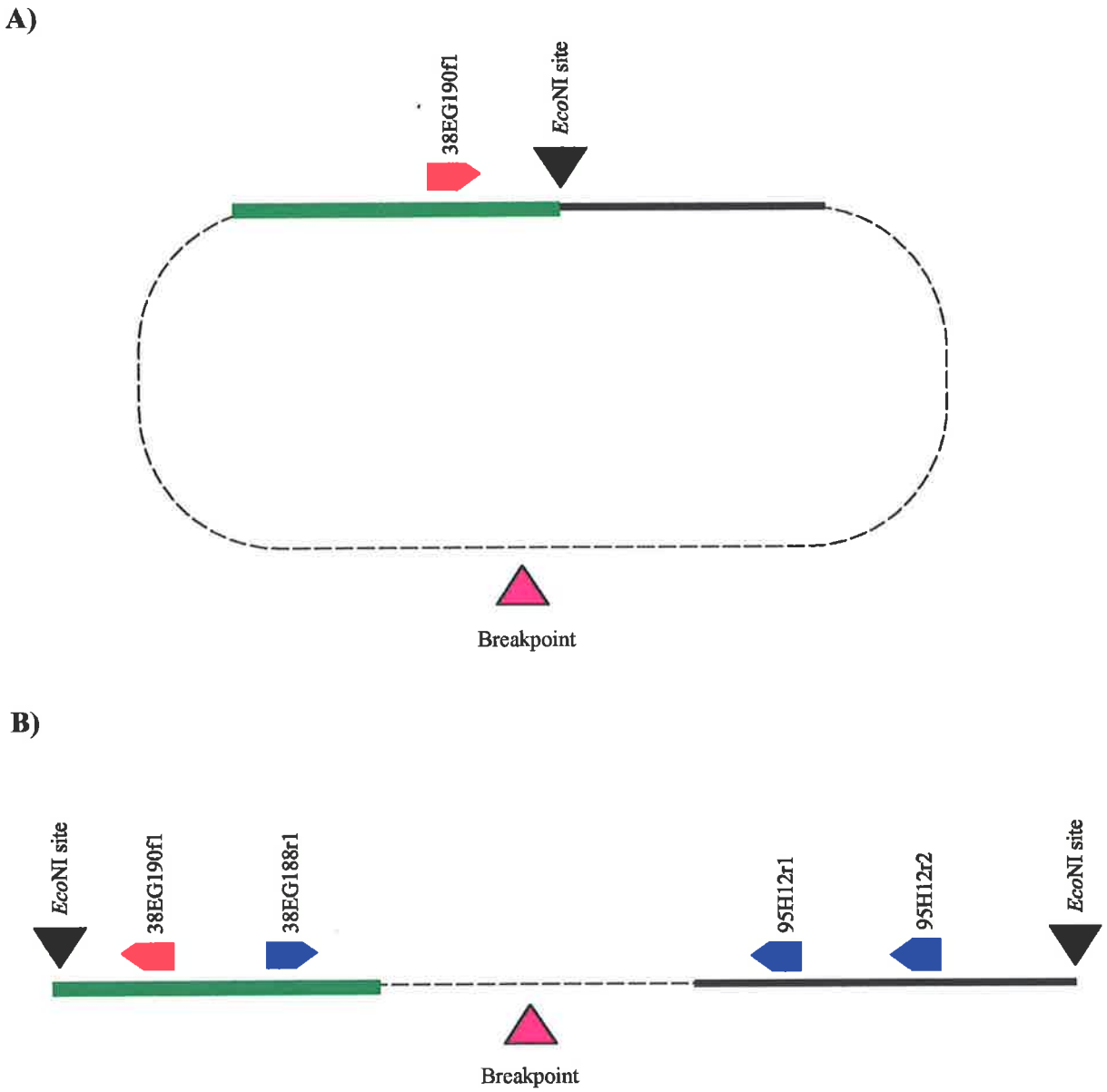


Figure 4.8. Design of Primers for Amplification of 1p36 - 12p13 Breakpoint.

A) Schematic representation of the *Eco*NI inverse PCR product. Known 1p36 sequence is represented in green. Known 12p13 sequence is shown as an unbroken black line. Unknown sequence is shown as a broken black line. The position of the sequencing primer 38EG190f1 is shown by a red pentagon. The *Eco*NI site which enabled circularisation of the inverse PCR product is shown by a black triangle. The approximate position of the breakpoint is represented by a pink triangle.

B) Illustration of the position of the same features as they exist on Patient 2 DNA. All features are represented as in panel A. The primers used for amplification of the breakpoint sequence are represented as blue pentagons.

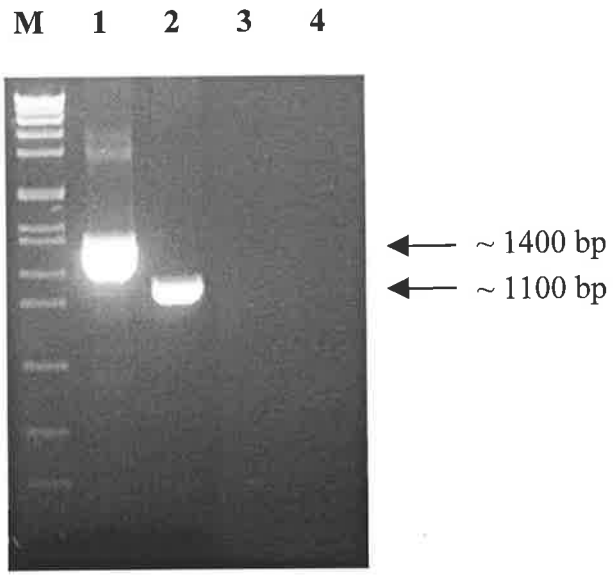
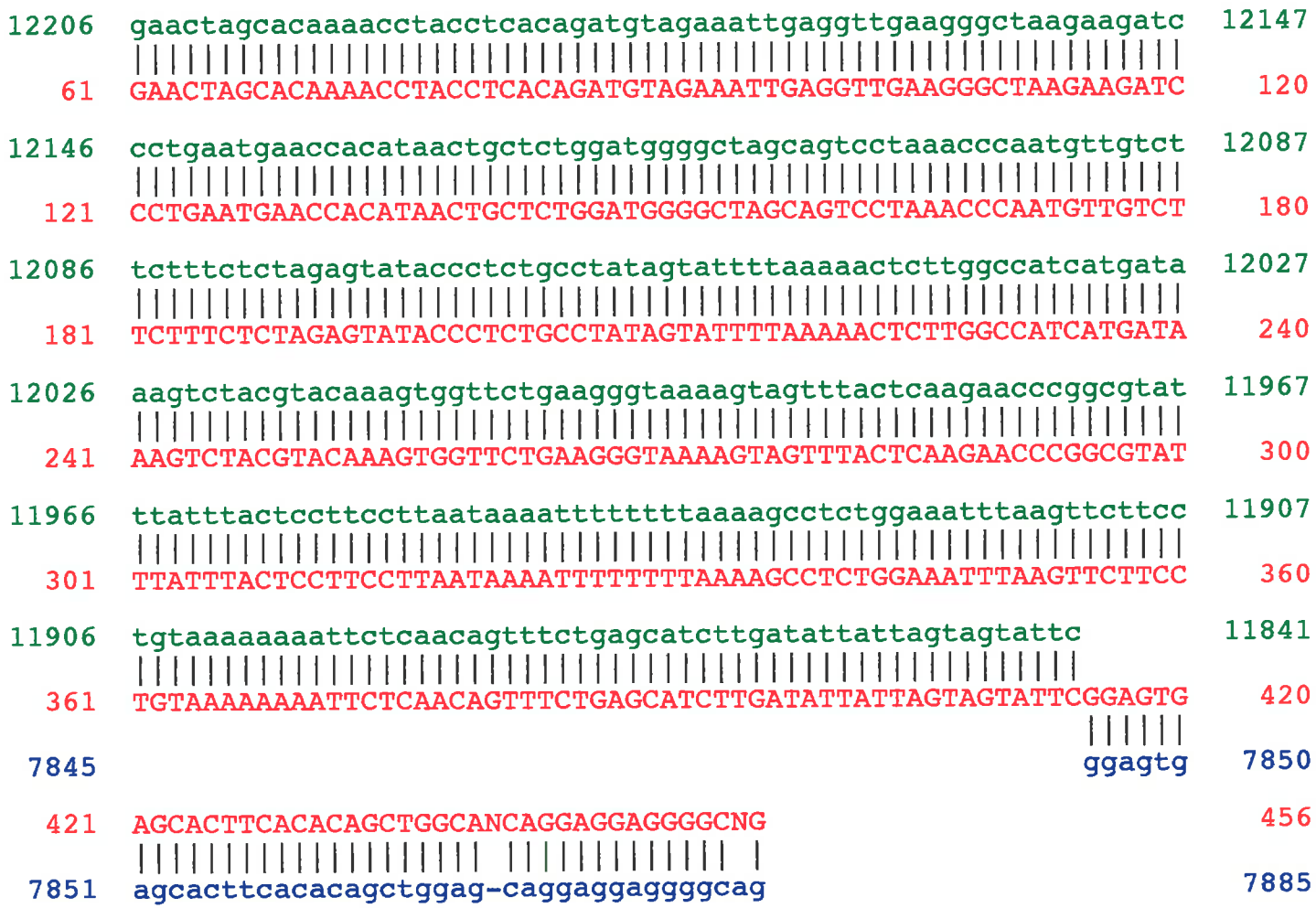


Figure 4.9. Amplification of Patient 2 1p36-12p13 Breakpoint Sequence.

Result of long template PCR using a 1p36 primer (38EG188r1) and a 12p13 primer (95H12r2 or 95H12r1) to amplify across the 1p36-12p13 translocation breakpoint in Patient 2. Normal bone marrow DNA was used as a negative control. The loading order is specified below.

M. SppI/*Eco*RI DNA size marker.
1. 95H12r2-38EG188r1 PCR from Patient 2 bone marrow DNA
2. 95H12r1-38EG188r1 PCR from Patient 2 bone marrow DNA
3. 95H12r2-38EG188r1 PCR from normal bone marrow DNA
4. 95H12r1-38EG188r1 PCR from normal bone marrow DNA

Figure 4.10. Sequence alignment of Patient 2 1p36-12p13 Breakpoint Sequence. The sequence of the PCR product from the primers 38EG188r1 and 95H12r2 from Patient 2 material, as sequenced with the 95H12r1 primer, is represented in red. It is aligned with sequences from 12p13 (LL12NCO1-95H4 (U81834.1), represented in green) and 1p36 (RP11-333E3 (AL356984.13), represented in blue).


12206 gaactagcacaaaacctacctcacagatgtagaaattgaggttgaagggctaagaagatc 12147
61 GAACTAGCACAAAACCTACCTCACAGATGTAGAAATTGAGGTTGAAGGGCTAAGAAGATC 120
12146 cctgaatgaaccacataactgctctggatggggctagcagtcctaaacccaatgttgtct 12087
121 CCTGAATGAACCACATAACTGCTCTGGATGGGGCTAGCAGTCCTAAACCCAATGTTGTCT 180
12086 tctttctctagagtataccctctgcctatagtattttaaaaactcttggccatcatgata 12027
181 TCTTTCTCTAGAGTATACCCTCTGCCTATAGTATTTTAAAAACTCTTGGCCATCATGATA 240
12026 aagtctacgtacaaagtggttctgaagggtaaaagtagtttactcaagaacccggcgtat 11967
241 AAGTCTACGTACAAAGTGGTTCTGAAGGGTAAAAGTAGTTTACTCAAGAACCCGGCGTAT 300
11966 ttatttactccttccttaataaaatttttttttaaaagcctctggaaatttaagttcttcc 11907
301 TTATTTACTCCTTCCTTAATAAAATTTTTTTTAAAAGCCTCTGGAAATTTAAGTTCTTCC 360
11906 tgtaaaaaaaattctcaacagtttctgagcatcttgatattattagtagtattc 11841
361 TGTAAAAAAAATTCTCAACAGTTTCTGAGCATCTTGATATTATTAGTAGTATTCGGAGTG 420
7845 ggagtg 7850
421 AGCACTTCACACAGCTGGCANCAGGAGGAGGGGCNG 456
7851 agcacttcacacagctggag-caggaggaggggcag 7885

BLASTN analysis of the 12p13 breakpoint showed that the breakpoint lies within intron 1 of the *TEL* gene. This gene is involved in numerous translocations in leukaemia, and in many cases forms fusion genes as a result of translocation. In this instance, the 5' end of the locus is presumably fused to 1p21 sequence, as it is the 3' end of the *TEL* locus which is joined to 1p36 sequence (see Figure 4.1 and Figure 4.11). As stated in Section 4.2.1, the 1p36 breakpoint is in a similar position to that of the breakpoint in the t(1;3)(p36;q21) translocation in that it is 5' of the *MEL1* gene. However, following the rearrangement, *TEL* and *MEL1* are oriented in opposite directions, as illustrated by the gene map presented in Figure 4.11. Both genes normally run from telomere to centromere, and as the 1p21 - 1p36 section of chromosome 1 is inverted prior to insertion into chromosome 12, the two genes are not correctly oriented to allow the formation of *TEL-MEL1* a fusion gene. Furthermore, according to the NCBI Human Genomic Sequence Assembly and further analysis of EST sequences using BLASTN analysis, no known gene in the 1p36 breakpoint region is oriented correctly to form a fusion transcript with the 3' end of the *TEL* gene. Therefore, any *TEL* fusion arising from this rearrangement is more likely to involve a 1p21 gene. This outcome would also be consistent with the observation that most *TEL* fusion genes involve the 5' end of the gene. For this reason, the identity of the 1p21 partner sequence was investigated.

4.4 Investigation of 12p13-1p21 Breakpoint Sequence

Knowledge of the sequence of the 1p36-12p13 breakpoint was utilised in the design of an inverse PCR to amplify from the telomeric side of the 12p13 breakpoint to the presumed 1p21 partner sequence. The design of this inverse PCR assumes that there is

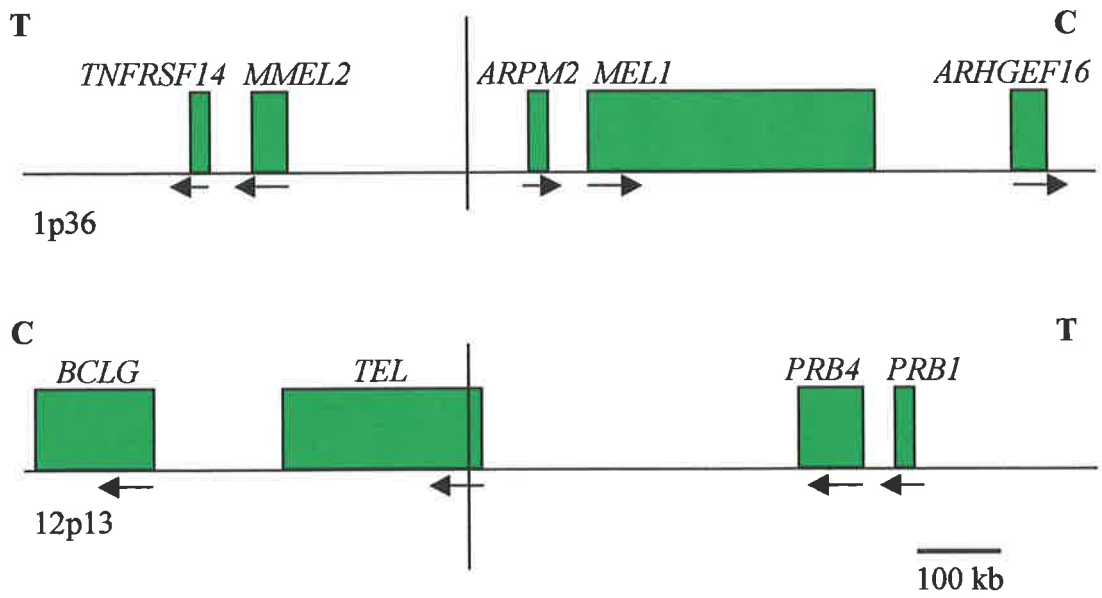


Figure 4.11. Gene Map of 1p36 - 12p13 Breakpoint in Patient 2.

The horizontal lines represent normal chromosome 1 (top) and normal chromosome 12 (bottom), and the vertical lines passing through each represents the breakpoints. Note that, because of the complexity of the inv ins (1;12)(p21-p36;p13), only the centromeric portions of these chromosomes recombine with each other. The portion of chromosome 1 telomeric of this breakpoint recombines with the centromeric portion of the 1p21 breakpoint, and the portion of chromosome 12 telomeric of this breakpoint recombines with the telomeric portion of the 1p21 breakpoint (see Figure 4.1). "T" and "C" indicate the telomeric and centromeric side of the breakpoint respectively. The green boxes indicate the position of the genes near the breakpoints. All genes are labelled above the appropriate boxes, and the arrows below the boxes indicate the direction of transcription. The approximate scale is indicated by the 100 kb scale bar at lower right.

no loss of material from 12p13 as a result of the translocation. Primers were positioned telomeric to the 12p13 breakpoint identified in Section 4.3. Restriction enzymes for use in this inverse PCR were chosen on the basis of availability and proximity of recognition sequences to the known breakpoint on the telomeric side, and the absence of recognition sequences in the region between the primers. The three enzymes chosen were *Ava*II, *Bst*NI and *Eco*RI. The design of the inverse PCR can be seen in Figure 4.12.

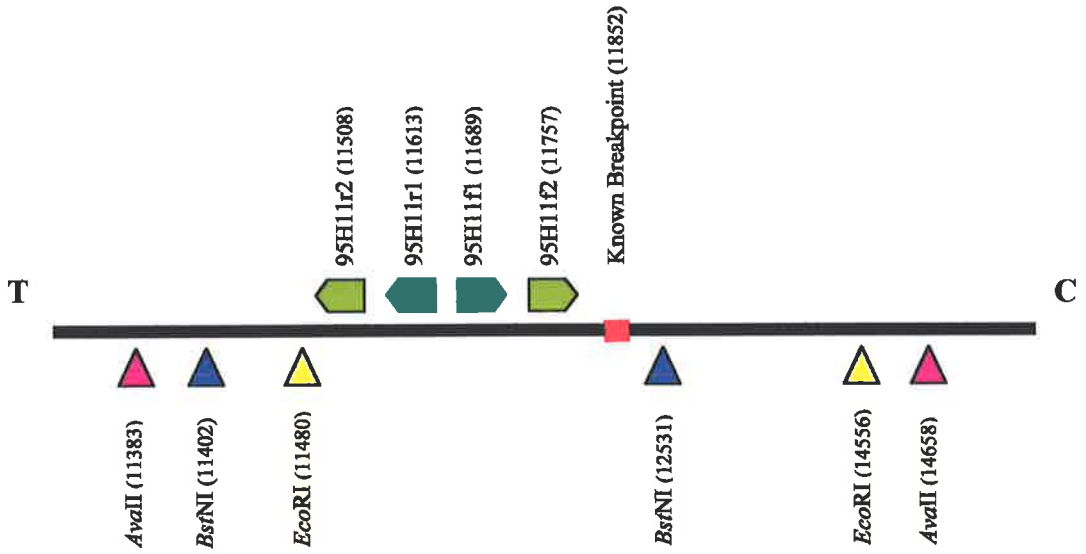


Figure 4.12. Design of Inverse PCR to Identify Partner Sequence of Telomeric 12p13 Breakpoint. The black line represents the genomic sequence of the 12p13 region. The letters T and C represent the telomeric and centromeric end of the region respectively. The red box represents the breakpoint as identified *via* the inverse PCR performed in Section 4.3. Primers used in the inverse PCR are represented by pentagons above the line. Primers are colour-coded as follows: dark green for primary inverse PCR primers and light green for secondary inverse PCR primers. Restriction sites utilised in the inverse PCR are represented by triangles below the line. Numbers in parentheses following a restriction enzyme site or primer name indicate the position of the object. Numbering is in accordance with LL12NC01-95H4 (accession U81834.1). The figure is not to scale.

Inverse PCR was performed as described in Section 2.2.3.7. Primers used were 95H11f1 and 95H11r1 in the primary reaction, and 95H11f2 and 95H11r2 in the nested secondary reaction. The result of this inverse PCR is presented in Figure 4.13.

The product amplified from the *Ava*II template appeared to be of the predicted size for the product amplified from the normal chromosome 12 (3026 bp; see Figure 4.12). Sequencing confirmed that this was the case (data not shown). The product amplified from the *Eco*RI template was also of the size expected for a product amplified from the normal chromosome 12 (2827 bp; see Figure 4.12). Sequencing confirmed this conclusion also (data not shown). No product was amplified from the *Bst*NI template. Given that a normal product of 880 bp was expected to be amplified from the normal chromosome 12 (see Figure 4.12), it is possible that the *Bst*NI template was incorrectly prepared.

As no products could be amplified by this inverse PCR that did not correspond to normal chromosome 12 products, the inverse PCR was not pursued further. It is possible that the reason no abnormal products were amplified is that the region adjacent to the breakpoint in Patient 2 on 12p13 was deleted during translocation, and the binding sites for the inverse PCR primers therefore exist only on the normal chromosome 12.

4.5 Investigation of 1p36-1p21 Breakpoint Sequence

Another means of identifying the 1p21 breakpoint is by the design of an inverse PCR to amplify across the recombination point of the telomeric portion of 1p36 to 1p21

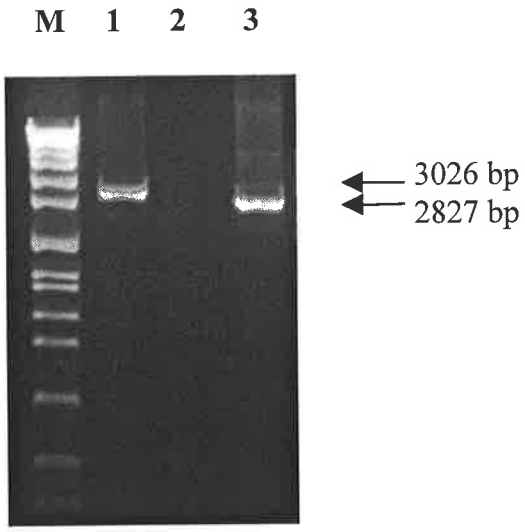


Figure 4.13. Inverse PCR to Identify Partner Sequence of Telomeric 12p13 Breakpoint.

Templates were prepared from Patient 2 DNA by restriction with the indicated restriction enzyme followed by circularisation. The secondary PCR from templates prepared using each of the three restriction enzymes is shown. Primary PCR was performed using primers 95H11f1 and 95H11r1. Secondary PCR was performed using 95H11f2 and 95H11r2, and 2 µl of a 1:50 dilution of the primary PCR as template. The loading order is specified below.

M1. SppI/*Eco*RI DNA size marker.
1. *Ava*II secondary inverse PCR.
2. *Bst*NI secondary inverse PCR.
3. *Eco*RI secondary inverse PCR.

(see Figure 4.1). The marker closest to the 1p36 breakpoint on the telomeric side, 38EG183, was used as a starting point for the inverse PCR design. Restriction enzymes were chosen on the basis of availability and proximity of recognition sequences to 38EG183 on the telomeric side. The absence of recognition sequences on the centromeric side was also taken into account. The three enzymes chosen were *Alw*NI, *Ear*I and *Nci*I. The design of this inverse PCR can be seen in Figure 4.14.

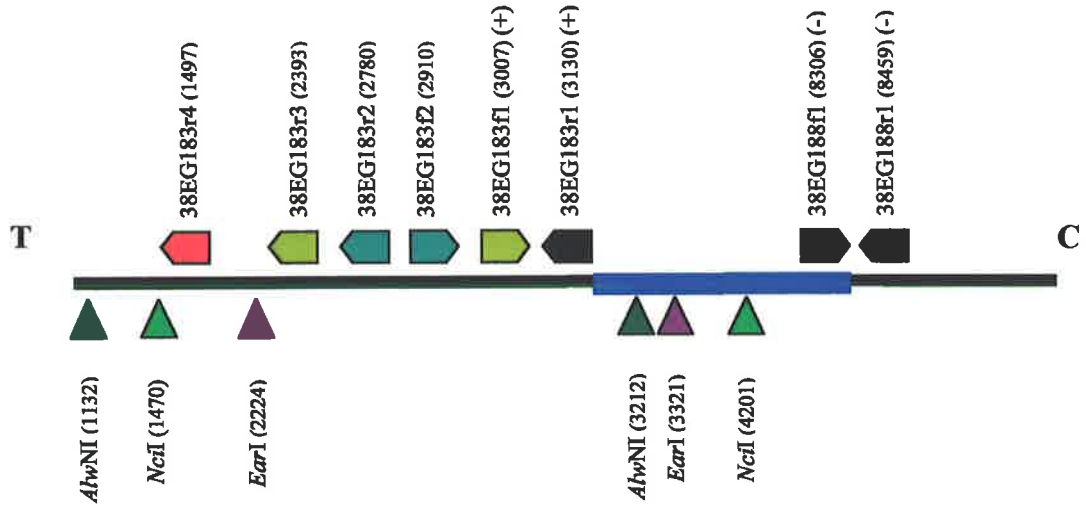


Figure 4.14. Design of Inverse PCR to Identify Partner Sequence of Telomeric 1p36 Breakpoint. The black line represents the genomic sequence of the 1p36 region. The letters T and C represent the telomeric and centromeric end of the region respectively. The blue box represents the breakpoint region as identified through somatic cell hybrid PCR walking. + and - indicate the presence or absence of markers tested against LH61. Primers used in either PCR walking, inverse PCR or inverse PCR sequencing are represented by pentagons above the line. Primers are colour-coded as follows: black for PCR walking primers, dark green for primary inverse PCR primers, light green for secondary inverse PCR primers and red for sequencing primers. Restriction sites utilised in the inverse PCR are represented by triangles below the line. Numbers in parentheses following a restriction enzyme site or primer name indicate the position of the object. Numbering is in accordance with RP11-333E3 (accession AL356984.13). The figure is not to scale.

Inverse PCR was performed as described in Section 2.2.3.7. Primers used were 38EG183f2 and 38EG183r2 in the primary reaction, and 38EG183f1 and 38EG183r3 in the nested secondary reaction. The result of this inverse PCR is presented in Figure 4.15.

For products generated from normal, non-rearranged chromosomes, the expected sizes are as follows. A 1513 bp product should be amplified from *Alw*NI digested template, a 530 bp product should be amplified from *Ear*I digested template, and a 2164 bp product should be amplified from *Nci*I digested template (see Figure 4.14). The band present in the *Alw*NI lane is of the expected size, and therefore likely to have been amplified from the normal chromosome 1. The lower band in the *Nci*I lane is also of the expected size. However, there is a second, larger band in the *Nci*I lane which may be amplified from a rearranged chromosome. Unfortunately, despite several attempts, this product remained unsequenced. As these PCR reactions produced more than one product, gel purification of the products was required prior to sequencing. This necessitated exposure to UV light, and it is thought that this exposure damaged the products and may have made them unsuitable for sequencing. This was only a problem for larger sized bands, however, and the band generated from the *Ear*I template, which was also not of the expected size and therefore also possibly amplified from a rearranged chromosome, was sequenced.

BLASTN analysis of the sequence obtained from the *Ear*I inverse PCR product showed that the restriction enzyme had cut at the recognition site correctly, although another *Ear*I recognition site elsewhere in the product remained uncut. The BLASTN analysis showed that the cut *Ear*I site had recombined with a sequence as illustrated

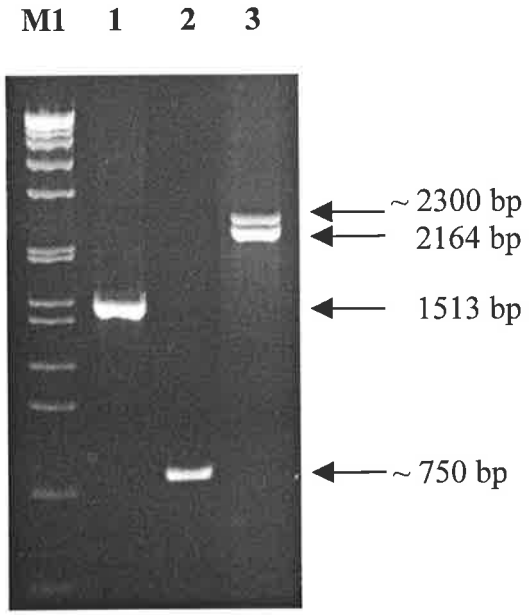


Figure 4.15. Inverse PCR to Identify Partner Sequence of Telomeric 1p36 Breakpoint.

Templates were prepared by restriction with the indicated restriction enzyme followed by circularisation. The secondary PCR from templates prepared using each of the three restriction enzymes is shown. Primary PCR was performed using primers 38EG183f2 and 38EG183r2. Secondary PCR was performed using 38EG183f1 and 38EG183r3, and 2 µl of a 1:50 dilution of the primary PCR as template. The loading order is specified below.

M1. SppI/*Eco*RI DNA size marker.
1. *Alw*NI secondary inverse PCR.
2. *Ear*I secondary inverse PCR.
3. *Nci*I secondary inverse PCR.

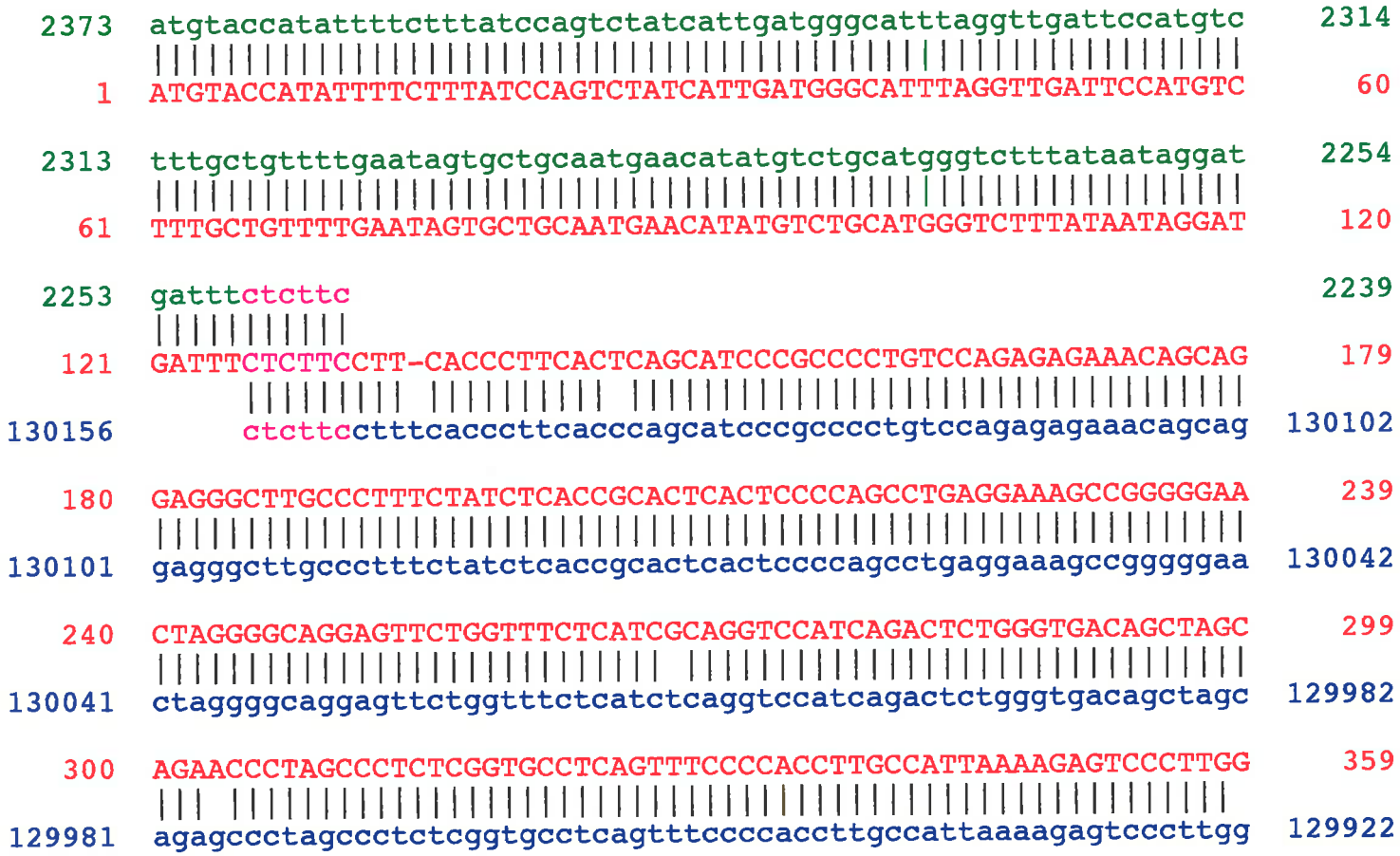
by the sequence alignment presented in Figure 4.16. The sequence identified adjacent to this restriction site, when analysed by the BLASTN algorithm against Build 29 of the NCBI Human Genome Assembly (May 2002), was found to be a repetitive sequence.

The chromosomes to which the sequence was originally ascribed were chromosomes 1, 2, 6, 7 and 13. However, a later BLASTN search, against Build 30 of the NCBI Human Genome Assembly (August 2002), resulted in matches to only two locations within the genome. These were 1q21 (clone RP11-763B22 (accession AL592492.10) on contig NT_034402; this is the sequence illustrated in Figure 4.16) and 2p11 (clone RP11-165D20 (accession AC027612.6) on contig NT_032994). BLAST analysis of 2 sequences (Tatusova and Madden, 1999) showed that these two clones have a large region of highly significant homology. Using numbering corresponding to RP11-763B22 (AL592492.10), the overlap stretches from bp 125793 - 154263, a region of almost 30 kb, and is of approximately 99 % identity. Whether both of these BLASTN matches to the *Ear*I inverse PCR product sequence are legitimate, indicating that the two regions are truly repetitive, or whether the overlap is an error associated with the assembly of the genome (as three of the five original matches are now presumed to be), remains to be seen. However, neither of these clones provides the expected match to 1p21, as predicted by the karyotype of the leukemic cells. There are three possible explanations for this. Firstly, it is possible that the *Ear*I PCR product is an artifact of the inverse PCR and not a true indication of a recombination that has occurred in the patient. Secondly, the chromosomal rearrangement in Patient 2 may be more complex than suggested by the karyotype analysis, and either 2p11 or, more likely, 1q21 are involved in the rearrangement. Finally, the 1p21 sequence involved in the

rearrangement may not be present in the database as it may not yet have been fully sequenced. The latter possibility is considered the more likely, for two reasons. Firstly, the *Ear*I inverse PCR product sequence, while a good match for both the 2p11 and 1p21 sequences, was not a perfect match for either, with three mismatches and a single bp deletion for the 1q21 match, and four mismatches for the 2p11 match, over the approximately 200 bp of partner chromosome sequence in each case (see Figure 4.16). This is an indication that neither of these two database sequences are the correct match for the inverse PCR sequence. Secondly, examination of the composition of the sequence data available in Build 30 of the NCBI Human Genome Assembly at 1p21, where the partner sequence was expected to be found, shows several regions in which contigs are represented in yellow, indicating that these contigs are constructed from unfinished sequences, as opposed to contigs constructed from finished sequence, which are shown in blue. While contigs constructed from unfinished sequences are present to an extent in most regions of the genome map, the fact that the sequence data at 1p21 is not yet finished to a high degree of confidence means that the true partner sequence of the telomeric 1p36 breakpoint may not yet be available.

Irrespective of the fact that the telomeric 1p36 breakpoint partner sequence had not been positively identified, an attempt was made to design primers that would amplify from RP11-333E3 across the breakpoint to the partner sequence from patient template DNA. This was done by designing the primers for the unknown partner sequence using the portion of the sequenced *Ear*I inverse PCR product beyond the *Ear*I site, that corresponded to the unknown partner chromosome. The Repeatmasker II algorithm (http://repeatmasker.genome.washington.edu/) was used to identify repetitive regions. One primer was designed within the repetitive region and one was

Figure 4.16. Sequence alignment of *Ear*I inverse PCR product. The sequence of the *Ear*I inverse PCR product, as sequenced with the 38EG183r3 primer, is represented in red. It is aligned with sequences from 1p36 (RP11-333E3 (AL356984.13), represented in green) and 1q21 (RP11-763B22 (AL592492.10), represented in blue). The presence of the *Ear*I recognition sequence is highlighted in pink. The *Ear*I restriction site is one base 3' of the recognition sequence on the 5' - 3' strand, and introduces a 3 base overhang on the 3' - 5' strand.

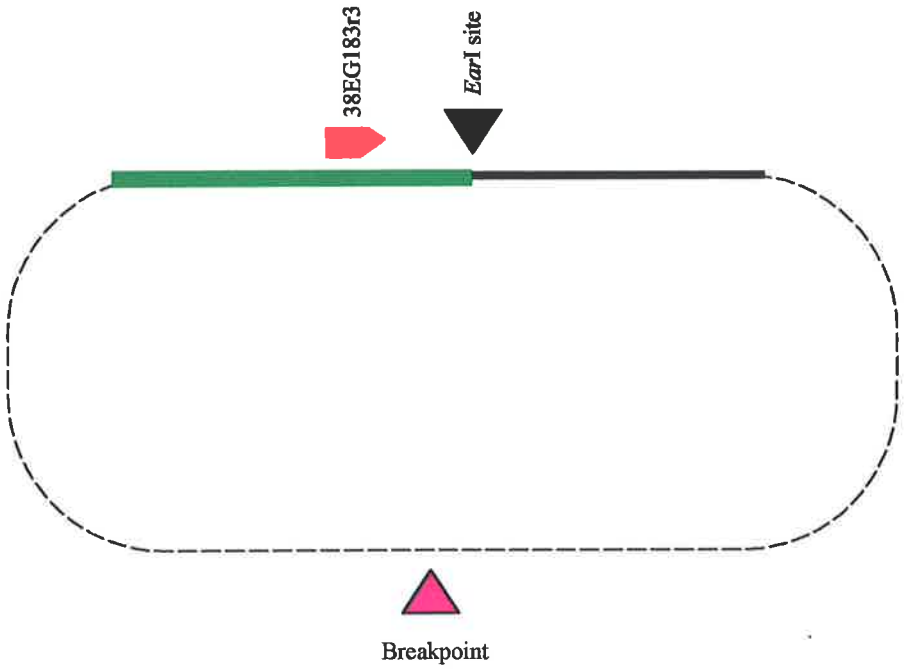

2373 atgtaccatattttctttatccagtctatcattgatgggcatttaggttgattccatgtc 2314
1 ATGTACCATATTTTCTTTATCCAGTCTATCATTGATGGGCATTTAGGTTGATTCCATGTC 60
2313 tttgctgttttgaatagtgctgcaatgaacatatgtctgcatgggtctttataataggat 2254
61 TTTGCTGTTTTGAATAGTGCTGCAATGAACATATGTCTGCATGGGTCTTTATAATAGGAT 120
2253 gatttctcttc 2239
121 GATTTCTCTTCCTT-CACCCTTCACTCAGCATCCCGCCCCTGTCCAGAGAGAAACAGCAG 179
130156 ctcttcctttcacccttcacccagcatcccgcccctgtccagagagaaacagcag 130102
180 GAGGGCTTGCCCTTTCTATCTCACCGCACTCACTCCCCAGCCTGAGGAAAGCCGGGGGAA 239
130101 gagggcttgcccttctatctcaccgcactcactccccagcctgaggaaagccgggggaa 130042
240 CTAGGGGCAGGAGTTCTGGTTTCTCATCGCAGGTCCATCAGACTCTGGGTGACAGCTAGC 299
130041 ctaggggcaggagttctggtttctcatctcaggtccatcagactctgggtgacagctagc 129982
300 AGAACCCTAGCCCTCTCGGTGCCTCAGTTTCCCCACCTTGCCATTAAAAGAGTCCCTTGG 359
129981 agagccctagccctctcggtgcctcagtttccccaccttgccattaaaagagtcccttgg 129922

designed in an apparently non-repetitive region, although BLASTN analysis of both primer sequences still identified several different regions of the genome, suggesting that both primers were still non-specific. The design of these primers (38EG183f1 and 763B130r1 and 38EG183f3 and 763B130r2) is illustrated in Figure 4.17. Both 763B130r1 and 763B130r2 are exact matches to the sequence from RP11-763B22, while the 763B130r1 primer has a single base pair mismatch at the 3' end compared to the 2p11 sequence RP11-165D20. The 763B22r2 primer matches both the RP11-763B22 and RP11-165D20 sequences exactly. Long Template PCR was performed using these primers, in various combinations and under many varied PCR conditions, on each of patient 2 DNA, normal human DNA, LH61 hybrid DNA and LTA DNA. No specific product could be amplified from any template using any combination of these primers, although many non-specific products were amplified (data not shown). The specificity of the PCR could not be improved by altering the conditions under which the reaction was conducted, despite many attempts. This is likely to be a reflection of the non-specific nature of the primers, due to the repetitive nature of the partner chromosome sequence. It is difficult to draw anything conclusive from this result.

4.6 *TEL* Northern Analysis

To further investigate the possibility that the Patient 2 rearrangement resulted in the formation of a *TEL* fusion gene, an RNA probe was made to assess this possibility by Northern analysis. As only a very small region of the *TEL* transcript was 5' of the breakpoint, the probe was designed to span the exon 1 - exon 2 junction, therefore spanning the breakpoint, with approximately the same amount of sequence on either

A)



B)

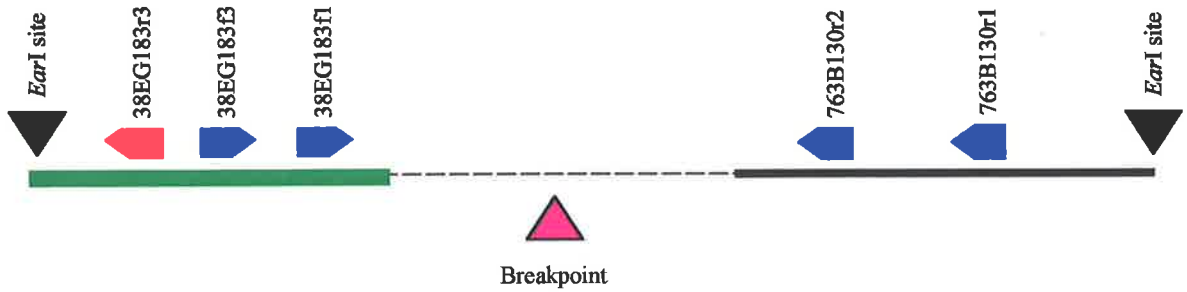


Figure 4.17. Design of Primers for Amplification of 1p36 - 1p21 Breakpoint.

A) Schematic representation of the EarI inverse PCR product. Known 1p36 sequence is represented in green. Known 1p21 sequence is shown as an unbroken black line. Unknown sequence is shown as a broken black line. The position of the sequencing primer 38EG183r3 is shown by a red pentagon. The *Ear*I site which enabled circularisation of the inverse PCR product is shown by a black triangle. The approximate position of the breakpoint is represented by a pink triangle.

B) Illustration of the position of the same features as they exist on Patient 2 DNA. All features are represented as in panel A. The primers used for amplification of the breakpoint sequence are represented as blue pentagons.

side. This probe should therefore detect any *TEL* fusion, whether it involves the 5' portion or the 3' portion of *TEL*. The probe was generated using primers rTELf4 and rTELr3 (see Figure 4.18). Patient 2 RNA, as well as RNA from two normal bone marrow samples, was poly(A)$^+$ selected, electrophoresed and transferred to a nylon membrane as described in Section 2.2.4. The *TEL* RNA probe was hybridised to the membrane and signal detection was conducted as described in Section 2.2.4. The result of this hybridisation is presented in Figure 4.19.

Figure 4.18. Position of *TEL* Primers within *TEL* coding region. The intron/exon structure of the *TEL* gene. A blue box represents an exon. A gap between boxes indicates an intron. To show the direction of transcription of the gene, the 5' and 3' orientation is indicated. Exon size is represented in approximate proportion, but the figure is not to scale. The black pentagons represent the position and orientation of RT-PCR primers with the name of the primer immediately above the pentagon. The arrow below the gene indicates the position of the breakpoint relative to the exons of *TEL*.

The major band has run slightly differently in each lane. This may be attributable to either unequal loading or buffering problems in the electrophoresis of the samples. This could not be confirmed by comparison with the EtBr stained gel as the RNA was poly(A)$^+$ selected and therefore there were no 28S or 18S RNA bands visible on the gel. Nonetheless, it is clear that there is one major band at approximately 2 kb in each lane, corresponding to the expected size of the full length *TEL* transcript. No other transcripts were detected using this probe. As no aberrantly sized transcripts were able to be detected in Patient 2 RNA by this experiment, it is considered unlikely that a *TEL* fusion transcript results from this rearrangement.

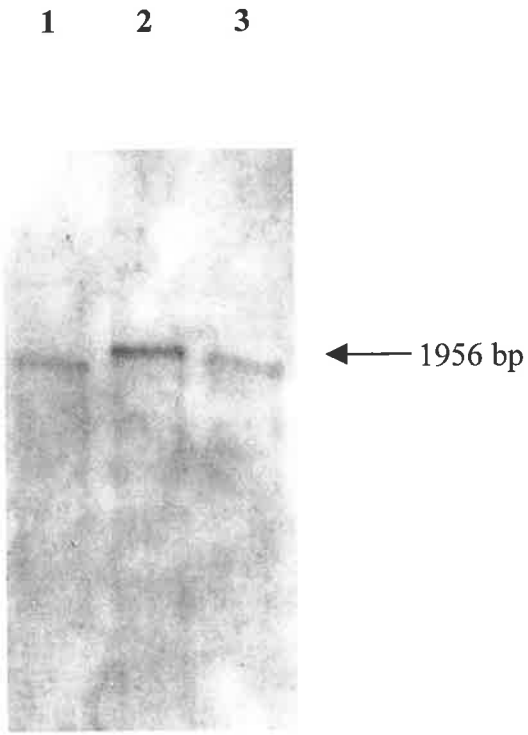


Figure 4.19. *TEL* Northern Analysis.

Result of Northern Analysis of Patient 2 RNA and two normal bone marrow samples, using a probe designed to flank the known *TEL* breakpoint in Patient 2. The lanes have run slightly unevenly. This is expected to be due to either unequal loading or buffering problems. However, it is apparent that each lane contains only one band, which is of the expected size for the normal *TEL* transcript. This result demonstrates that Patient 2 is unlikely to express a *TEL* fusion transcript, although it does not exclude the possibility that a fusion transcript of very similar size to the normal *TEL* transcript is expressed. The loading order is specified below.

1. Patient 2 bone marrow RNA
2. Normal bone marrow RNA donor 1
3. Normal bone marrow RNA donor 2

As a *TEL* fusion does not seem to be formed by this rearrangement, other mechanisms of leukemogenesis must be considered. As shown in Chapter 3, Patient 2 expresses *MEL1* at elevated levels compared to normal bone marrow (see Figure 3.15). While *MEL1* expression does occur in patients without 1p36 rearrangements, expression is significantly linked to the presence of the t(1;3)(p36;q21) (see Section 3.4.1). Given that the 1p36 breakpoint to the *MEL1* gene in Patient 2 is in a similar position to those in t(1;3)(p36;q21) patients, the possibility that *MEL1* is upregulated as a result of the rearrangement in Patient 2 must be considered. A recent study shows that 12p13 translocations with similar breakpoints to that of Patient 2 may result in upregulation of genes on the partner chromosome as a result of position effects of sequences from 12p13 (Cools *et al* 2002). The orientation of the *TEL* and *MEL1* genes in the Patient 2 rearrangement is in agreement with this position effect mechanism, and it is therefore likely that the upregulation of *MEL1* in Patient 2 is caused by position effects of sequences from 12p13. This mechanism is analogous to the position effect of the *RPN1* enhancer seen in t(1;3)(p36;q21), as well as in t(3;3)(q21;q26) and inv(3)(q21q26) (see Sections 3.1.1 and 3.5.2.2).

4.7 Discussion

4.7.1 Identification of the 1p36 - 12p13 Breakpoint Sequence

Somatic cell hybrid chromosome walking identified the 1p36 breakpoint in the Patient 2 chromosome rearrangement, ins(12;1)(p13;p36p21), as being similar to those in patients bearing the t(1;3)(p36;q21), in that it was 5' of the *MEL1* locus. Expression

analysis described in Section 3.4.1 revealed that Patient 2 did express *MEL1*, and this was considered likely to be an outcome of the rearrangement.

However, inverse PCR analysis of the centromeric side of this breakpoint led to the discovery that the partner chromosome was 12p13, as expected from karyotype analysis, and that the breakpoint on 12p13 lay within intron 1 of the *TEL* gene, which is involved in numerous fusion genes in leukaemia. This result contrasts with the intron 2 breakpoint reported in the t(1;12)(p36;p13) (Odero *et al* 2001), but suggests that a similar mechanism of leukemogenesis may result from both rearrangements. The definition of the *TEL* intron 1 breakpoint in Patient 2 raised the question of how *TEL* contributed to the leukemogenic nature of the rearrangement. This was likely to be through the formation of a fusion gene, given the position of the breakpoint within the *TEL* gene. However, no gene near the 1p36 breakpoint was correctly oriented to be involved in a fusion transcript with *TEL*. Also, the majority of *TEL* fusion genes involve the 5' end of *TEL*, which, given the orientation of *TEL* in this rearrangement, would have to involve the 3' end of a 1p21 gene. For this reason, characterisation of the 1p21 breakpoint was pursued.

4.7.2 The 1p36-1p21 and 12p13-1p21 Breakpoints

The other breakpoints generated by the complex rearrangement in Patient 2, ins(12;1)(p13;p36p21), are presumed to be a 1p36-1p21 breakpoint and a 12p13-1p21 breakpoint. These breakpoints proved difficult to characterise and as such their relevance to the leukemic phenotype is difficult to ascertain.

The inverse PCR designed to identify the partner chromosome for the telomeric portion of 12p13 was unsuccessful, as only bands corresponding to normal chromosome 12 were amplified. This could be due to deletion of a small part of 12p13 in association with the rearrangement. As the hybrid cell line used to investigate this rearrangement, LH61, contained the der 1 chromosome and therefore no chromosome 12 material, it was impossible to investigate this possibility by somatic cell hybrid marker analysis. As the extent of any deletion, if present, could not be known, redesign of the inverse PCR to allow for the missing sequence was not possible.

Sequence obtained from the inverse PCR designed to identify the partner chromosome band of the telomeric side of the 1p36 breakpoint was not an identical match for any sequence in the Genbank database. The best matches were to either 1q21, the long arm of chromosome 1 rather than the short arm as expected, or 2p11 based on the sequence obtained from the inverse PCR. Both 1q21 and 2p11 are known to be involved in chromosomal rearrangements in leukaemia, as briefly discussed below. It is possible that the rearrangement is more complex than suggested following the karyotype analysis. The possibility of a pericentric inversion of chromosome 1 prior to the translocation taking place would explain a 1q21 breakpoint. This is purely speculative, however. Importantly, the LH61 hybrid retained the marker *AMPD2* at 1p12 (see Section 4.2.1). This suggests that 1p12 is retained on the der 1 chromosome, such that loss (or insertion into chromosome 12) of all of 1p36-1q21 is not a possibility. This is also confirmed by the retention of a centromere on the der 1 chromosome (see Figure 4.1).

The principal chromosomal rearrangement in leukaemia involving 1q21 is the t(1;11)(q21;q23) in AML, which results in the fusion of *MLL* with *AF1Q*, a 1q21 gene (Tse *et al* 1995). The *AF1Q* locus is approximately 2 Mb from where the Patient 2 breakpoint appeared to be according to the inverse PCR result (Entrez MapView (http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/maps.cgi)), and so would be unlikely to be affected by any such rearrangement that may have taken place in Patient 2. Interestingly, there have also been two reports of an inv(1)(p36;q21) in AML M2 (Mitelman Database of Chromosome Aberrations in Cancer (http://cgap.nci.nih.gov/Chromosomes/Mitelman)). The precise breakpoints and molecular outcome in these cases were not investigated. Chromosome band 2p11 is involved in a recurrent t(2;12)(p11;p12), which is interesting given that chromosome 12 breakpoints for translocations involving *TEL* rearrangement are often reported as being at 12p12 rather than 12p13 due to the position of *TEL* at the boundary of 12p12 and 12p13 (Baens *et al* 1999). The precise breakpoints and molecular outcome in this rearrangement are also unknown.

Given that neither of these matches was identical, and also that both are inconsistent with the karyotype analysis, neither 1q21 nor 2p11 are likely to be the true partner chromosome band. The failure to obtain an identical match by BLASTN searching of any Genbank database with the *Ear*I inverse PCR product suggests that the true partner sequence involved in the rearrangement is yet to be entered into the databases. The match illustrated in Figure 4.16 is the best match found in the database, and it contains 3 mismatches and a single bp insertion in the portion of the inverse PCR product corresponding to the partner sequence (which derives from 1q21 in this case). The quality of the sequence data obtained was of a high standard, and this is borne out

by the fact that there are no mismatches in the portion of the sequence that corresponds to the 1p36 sequence. It seems most likely, therefore, that there is 1p21 region, similar to the repeated region present at 1q21 and 2p11, that is yet to be sequenced and entered into the database, and that this region corresponds to the 1p21 breakpoint in Patient 2.

4.7.3 Expression of *MEL1* in Patient 2

MEL1, the gene shown to be upregulated by the t(1;3)(p36;q21), was also shown to be expressed in Patient 2 (see Figure 3.15). *MEL1* is also expressed in some patients who do not have 1p36 rearrangements. However, statistical analysis showed that translocation-mediated expression was a genuine outcome of the t(1;3)(p36;q21) (see Chapter 3). As the 1p36 breakpoint in Patient 2 is in the same region as those in the t(1;3)(p36;q21), it seems likely that the rearrangement is responsible for the upregulation of *MEL1*. However, this could not be achieved *via* fusion of *TEL* to *MEL1*, in a mechanism similar to that of the *TEL-MDS1-EVI1* fusion transcript, as the *MEL1* and *TEL* genes were incorrectly oriented for such a fusion to occur.

A precedent for a model of upregulation of genes juxtaposed to 12p13 sequences, but without the formation of a *TEL* fusion gene, is provided by the recent findings of Cools *et al* (2002). In a case of t(4;12)(q11-q12;p13), which had in other cases been shown to result in an in-frame *CHIC2-TEL* fusion gene (Cools *et al* 1999), these investigators found no *TEL* fusion gene was formed. Similarly, the t(5;12)(q31;p13), which had in other patients been shown to result in an out-of-frame *TEL-ACS2* fusion gene (Yagasaki *et al* 1999), Cools *et al* (2002) again found no *TEL* fusion gene was

formed. After investigation of the expression of other genes at 4q11-4q12 and 5q31 in these cases, Cools *et al* (2002) found that patients with the t(4;12)(q11-q12;p13) translocation ectopically expressed the *GSH2* gene located at 4q11-4q12. Similarly, patients with the t(5;12)(q31;p13) translocation ectopically expressed *IL-3*, located at 5q31. These results argue for a mechanism of dysregulation of oncogenes by juxtaposition of regulatory sequence in 12p13 translocations with breakpoints within the *TEL* locus, irrespective of the formation of a fusion gene. This mechanism is likely to apply to both *MEL1* and *EVI1* upregulation in 12p13 rearrangements with 1p36 and 3q26 respectively. The *TEL-EVI1* fusion is formed in some cases, and it may be that some patients with t(1;12)(p36;p13) will express a *TEL-MEL1* fusion transcript. But whether or not a fusion is formed, the outcome of the translocation is likely to be transcriptional upregulation of *EVI1* or *MEL1*, respectively.

4.7.4 The *TEL-MDS2* Fusion Gene in t(1;12)(p36;p13)

During the writing of this thesis, the detailed molecular characterisation of the case reported in Section 4.1.1 was published (Odero *et al* 2002). The findings of this investigation contrast with the findings presented in this thesis, in that the t(1;12)(p36;p13) reported by Odero *et al* (2002) had a 1p36.1 breakpoint and the breakpoint in Patient 2 was mapped to 1p36.3. This demonstrates that at least two distinct rearrangements involving 1p36 and 12p13 occur in haematological malignancy. Odero *et al* (2002) found that the 1p36 breakpoint in the t(1;12)(p36;p13) in their patient was at 1p36.1, and disrupted a novel gene which they named *MDS2*. *MDS2* is not related to *MDS1*, discussed in Chapter 3, as determined by BLASTN analysis. *MDS2* formed a fusion transcript with the first two exons of *TEL* as an

outcome of the t(1;12)(p36;p13), but the resultant open reading frame encoded 58 amino acids, only four of which were derived from *MDS2*, and the predicted fusion protein does not contain any functional domains. This fusion is therefore unlikely to produce a protein that could contribute to the progression of the malignancy. Importantly, Odero *et al* (2002) also investigated the expression of other genes near the breakpoint and found that the *RPL11* gene on 1p36 showed significant transcriptional upregulation in the patient carrying the t(1;12)(p36;p13). This finding is further evidence of the position effect that 12p13 translocations can have on juxtaposed genes. As this breakpoint was at 1p36.1, and *MEL1* is at 1p36.3, the upregulation of *MEL1* expression is not a plausible outcome of this t(1;12)(p36;p13). In combination with the present study, the findings of Odero *et al* (2002) indicate a heterogeneity in 1p36 translocation breakpoints generally, as well as in rearrangements with 12p13 specifically. Investigation of more t(1;12)(p36;p13) patients will reveal whether all such translocations involve *MDS2*. Alternately, most t(1;12)(p36;p13) 1p36 breakpoints may cluster to the same region as those of Patient 2 and the t(1;3)(p36;q21) patients and hence result in upregulation of *MEL1*, and the translocation investigated by Odero *et al* (2002) may be a rare anomaly.

4.8 Conclusions

The 1p36 breakpoint in the rearrangement in Patient 2, ins(12;1)(p13;p36p21), was mapped and shown to be located in the same breakpoint region as the 1p36 breakpoints, from patients with the t(1;3)(p36;q21) examined in this study (see Section 3.2.1) and in Mochizuki *et al* (2000).

The 12p13 sequence which recombined with the telomeric end of 1p36 was identified by inverse PCR. The 12p13 breakpoint was shown to be within intron 1 of the *TEL* gene. The arrangement of the genes on the der 12 chromosome did not allow for the possibility that a *TEL* fusion could be formed with any known genes on 1p36. The possibility of a fusion between the 5' end of *TEL* and the 3' part of a gene from 1p21 (or the chromosome band found to have recombined with the telomeric 12p13 breakpoint) still exists. Sequence data for the 1p21 breakpoint region was obtained but did not match any 1p21 sequence data present within the Genbank databases. This is likely to be due to the incomplete nature of the publicly available human genomic sequence.

The presence of a *TEL* fusion transcript was investigated by Northern analysis. No aberrant band was detected, and so the existence of a fusion transcript as a result of this rearrangement was unlikely in this case. It is possible that a fusion transcript may be of similar length to the normal transcript, and so be masked on the Northern by the normal *TEL* band.

MEL1 was found to be expressed in Patient 2 (see Section 3.4.1 and Figure 3.15). Given the position of the 1p36 breakpoint, and the findings of Cools *et al* 2002, who found that translocations involving 12p13 both within and without the *TEL* gene, could cause upregulation of juxtaposed genes without the formation of a *TEL* fusion gene, the conclusion reached by this study is that *MEL1* expression is likely to be the primary molecular outcome of this chromosomal rearrangement.

CHAPTER 5

The

(1;3)(p36;p21)

Translocation

CHAPTER 5

t(1;3)(p36;p21)

5.1 t(1;3)(p36;p21) as a Recurrent Translocation

5.1.1 Background

The t(1;3)(p36;p21) has only recently been recognised as a recurrent translocation (Sato *et al* 2002). There have been eight cases reported: three in ALL patients, one in a CML patient, two in MDS patients and two in AML M3 patients (Sato *et al* 2002). One of the AML M3 patients had a variant form of the translocation, a t(1;2;3)(p36;q21;p21). Five of the eight patients had received chemotherapy prior to detection of the translocation, prompting suggestions from the authors that the translocation is therapy-related. All eight patients had chromosome abnormalities other than the t(1;3)(p36;p21), and in two cases the t(1;3)(p36;p21) occurred only in post-chemotherapy relapse analyses. Other than the potential correlation with chemotherapy, there was no apparent clinical feature that marked patients with the t(1;3)(p36;p21). Survival times of the patients ranged from 25 days to greater than 16 years. This may support the hypothesis that the t(1;3)(p36;p21) is a secondary aberration. Alternately, it may be that, while the rearrangements appear identical cytogenetically, they are different at the molecular level, and therefore contribute differently to leukemogenesis.

5.1.2 Patient Information

Patient 3 was a 65 year old male who presented in November 1988 with megaloblastic anaemia, and was diagnosed with a lymphoproliferative disorder of the bone marrow. Later studies showed that, although the leukemic blasts appeared morphologically lymphoid in origin, they stained with Sudan Black, which indicated that they were of myeloid origin. The bone marrow aspirate showed a reduction in normal marrow elements, including reduced megakaryocyte numbers, abnormal erythropoiesis and abnormal lymphocyte population. The final diagnosis was of acute leukaemia of indeterminate type, but which was more consistent with the myeloid lineage. The patient died in April 1990, 17 months after presentation.

Karyotype analysis was performed three times during the patient's treatment, but the reports are inconsistent and inconclusive in their assessment. Fifteen cells were typed in each case, and between ten and thirteen were found to be abnormal. In all abnormal cells, rearrangements of the short arm of chromosome 1 and chromosome 3 were seen, but the translocation was variously reported as balanced and unbalanced, and two reports were unable to determine the origin of the recombinant material on the der 1 and der 3 chromosomes.

Frozen samples were sent for karyotyping at the outset of this study. The karyotype obtained is shown in Figure 5.1. Metaphase spread preparation was difficult as the sample had been stored for many years and as a result the banding pattern is difficult to interpret. However, there are clearly abnormalities of chromosome arms 1p and 3p, and the karyotype was reported as 46, XY, t(1;3)(p36;p23),

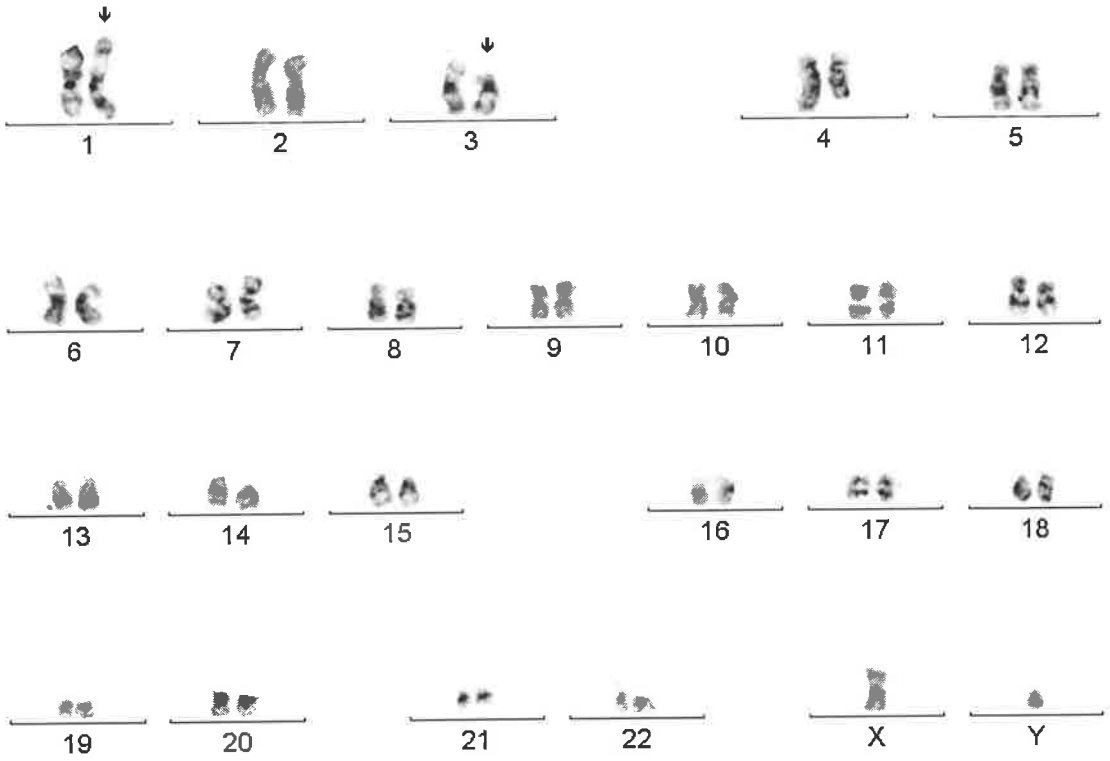


Figure 5.1. Patient 3 Karyotype.

The aberrant karyotype of the leukemic blasts of Patient 3, showing 46, XY, t(1;3)(p36;p23), der(1)ins(1)(p36p12p32)t(1;3). Aberrant chromosomes are denoted with arrows. Karyotype analysis was performed by Sarah Moore (Institute of Medical and Veterinary Science, Adelaide, Australia).

der(1)ins(1)(p36p12p32)t(1;3) (Sarah Moore, Institute of Medical and Veterinary Science, Adelaide, Australia). The possibility that the rearrangement involved 3p21 rather than 3p23, and was thus a variant of the recurrent t(1;3)(p36;p21) reported in Sato *et al* (2002), could not be ruled out due to the difficulty of interpretation.

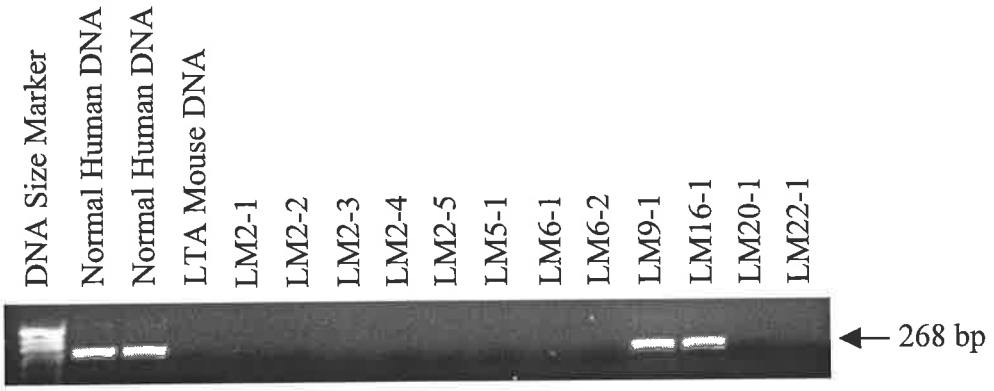
5.2 Somatic Cell Hybrid Analysis

Hybrid cell lines were constructed, using the murine fibroblast line LTA and patient bone marrow material collected at presentation (see Section 2.2.5.4). Twelve colonies were obtained, and these were subcloned and cultured for analysis.

5.2.1 Chromosome 1 Analysis

The twelve hybrid cell lines were screened with 1p36 markers 425F3Z (telomeric marker) and D1S2661 (centromeric marker) to identify hybrid cell lines which retain derivative chromosomes. Ten of the hybrid lines were negative for both markers, indicating that they had discarded both derivative chromosomes and the normal chromosome 1, and so were not informative for breakpoint analysis. Two hybrid lines, LM9 and LM16, were positive for both markers (see Figure 5.2). This pattern suggested that both hybrid lines had retained the normal chromosome 1, but it was also possible that they retained both derivative chromosomes, or the normal chromosome 1 and either (or both) derivative chromosomes. Irrespective of the exact combination of chromosomes, neither hybrid cell line was informative for investigating the chromosome 1 breakpoint. To test whether any of the hybrid cell

A) 425F3Z Screen



B) D1S2661 Screen

Figure 5.2. Chromosome 1 Marker Screen of LM Hybrid Cell Lines.

Results of testing for presence or absence of chromosome 1 markers in the panel of somatic cell hybrids generated from Patient 3 bone marrow material. Panel A shows the 425F3Z screen, and Panel B shows the D1S2661 screen. The loading order is shown above Panel A and is the same in both panels. 425F3Z is at the telomeric end of 1p36. D1S2661 is on chromosome band 1p32 and so centromeric of the presumed breakpoint. The DNA size marker used was pUC19/*Hpa*II. A sample of DNA from a normal human individual was used as a positive control. A sample of DNA from the murine LTA cell line was used as a negative control. Hybrids are named LM, followed by numbers corresponding to the plates on which they were grown. The expected size of the PCR product is indicated on the right. Only two of the hybrids, LM9 and LM16 retain any chromosome 1 material, and both appear to retain the entire normal chromosome 1. They are therefore not informative for chromosome 1 PCR walking.

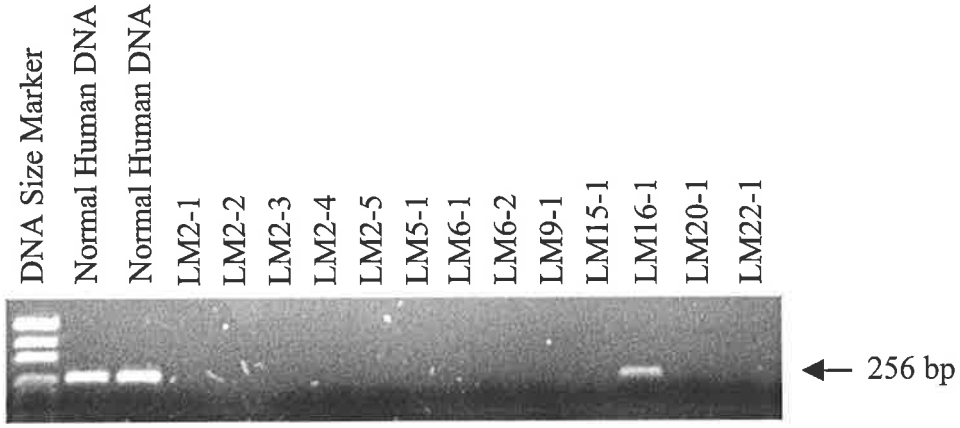
lines were informative for the chromosome 3 breakpoint, a screen using chromosome 3 markers was subsequently conducted.

5.2.2 Chromosome 3 Analysis

To further investigate the combination of chromosomes retained by the hybrid cell lines LM9 and LM16, a screen was also performed with chromosome 3p markers *FHIT* (3p14, centromeric marker) and *CHL1* (3p26, telomeric marker). The two hybrid cell lines, LM9 and LM16, which were found to contain chromosome 1 material, were also found to contain chromosome 3 material, as shown in Figure 5.3.

No hybrid cell line other than LM9 or LM16 showed retention of either chromosome 3 marker, indicating that they had discarded all chromosome 3 material. LM16 retained both of these chromosome 3 markers, indicating that LM16 retains a normal chromosome 3 (or both derivative chromosomes, or a combination of the three chromosomes that includes a normal chromosome 3), and was therefore not analysed further. LM9 retained *FHIT* but had lost *CHL1*, indicating that it retained the der 3 chromosome (and no normal chromosome 3), in addition to a normal chromosome 1. LM9 was therefore an informative hybrid cell line for further analysis, although this analysis would of necessity focus on the chromosome 3 breakpoint rather than the chromosome 1 breakpoint. The positive marker *FHIT* is located at 3p14, and the negative marker *CHL1* is located at 3p26, so a series of markers was designed along the length of the chromosome arm to narrow the region known to contain the breakpoint (see Figure 5.4).

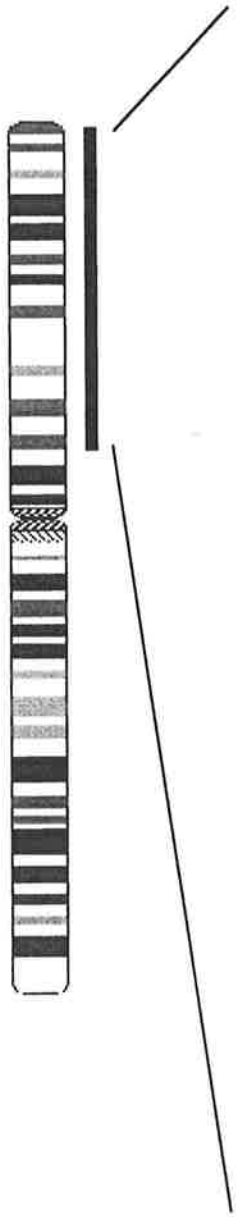
A) *CHL* Screen



B) *FHIT* Screen

Figure 5.3. Chromosome 3 Marker Screen of LM Hybrid Cell Lines.

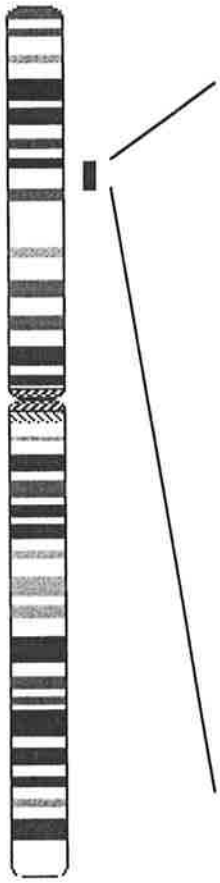
Results of testing for presence or absence of chromosome 1 markers in the panel of somatic cell hybrids generated from Patient 3 bone marrow material. Panel A shows the *CHL* screen, and Panel B shows the *FHIT* screen. The loading order is shown above Panel A and is the same in both panels. Note that LM15-1, which had not grown sufficiently at the time of the chromosome 1 screening process, has been included in the panel. LM15-1 was screened independently for chromosome 1 markers and found to contain no chromosome 1 material. *CHL* is at 3p26 and so telomeric of the presumed breakpoint. *FHIT* is at 3p14 and so centromeric of the presumed breakpoint. Only two of the hybrids, LM9 and LM16 retain any chromosome 3 material. LM16-1 appears to retain the entire normal chromosome 3, and is therefore not informative for chromosome 3 PCR walking. LM9-1 contains *FHIT* but not *CHL*, and therefore is expected to contain the der 3 translocated chromosome.



Marker	Position on Chromosome (Mb)	Presence in LM9
CHL1	2.6	-
PCAF	23.4	-
RARB	29	-
RBMS3	32.5	-
D3S12	33.4	-
TGFBR2	33.7	-
FBXL2	34.6	-
5' *LOC90670*	36.2	-
5' *LOC131656*	39.1	-
LRRFIP2	39.7	-
KIAA0342	40	-
GOLGA4	40.2	-
LOC152440	40.3	-
ITGA9	40.4	+
ENTPD3	43.2	+
FHIT	62.8	+

Figure 5.4. Markers used in Chromosome 3 Breakpoint Analysis of LM9. The table shows the distance, in Mb, of each marker from the p arm telomere on chromosome 1, and the result obtained on analysis of LM9 for each marker. + indicates that the LM9 retained the marker. - indicates the marker was not present in LM9.

This series of PCR markers narrowed the breakpoint to between the markers *LOC152440* and *ITGA9*. BLASTN analysis with these two pairs of primers revealed that they were both present on a single genomic database entry (Genbank accession AP000493), and that they were located approximately 125 kb apart on chromosome band 3p21. As the breakpoint is in 3p21, the translocation which occurred in Patient 3 is hereafter referred to as t(1;3)(p36;p21). Markers were then designed to this sequence, between *LOC152440* and *ITGA9*, to precisely define the breakpoint (see Figure 5.5).



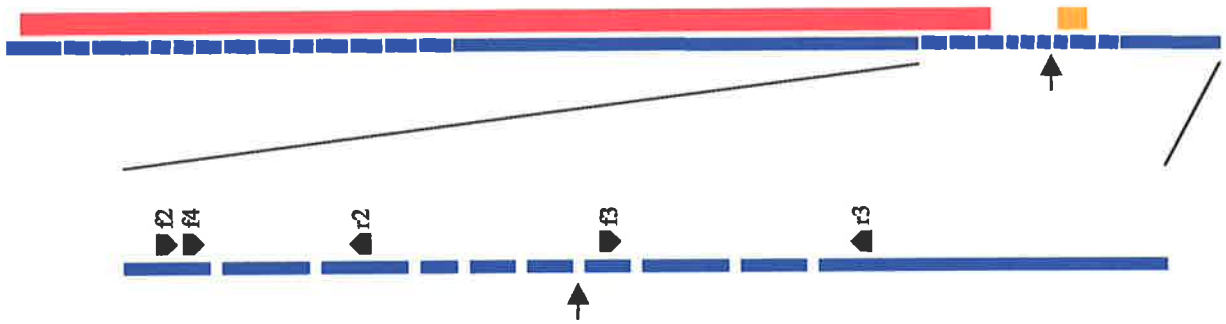
Marker	Position on AP000493 (kb)	Presence in LM9
LOC152440	0.5	-
493AP15	15	-
493AP16	16	+
493AP20	20	+
493AP23	23	+
493AP31	31	+
493AP61	61	+
ITGA9	125	+

Figure 5.5. Markers used in Fine Chromosome 3 Breakpoint Analysis of LM9. The table shows the location, in kb, of each marker on the NCBI clone AP000493, and the result obtained on analysis of LM9 for each marker. + indicates that the LM9 retained the marker. - indicates the marker was not present in LM9.

This process narrowed the breakpoint region to a distance of approximately 1 kb, between the markers 493AP15 (negative, telomeric marker) and 493AP16 (positive, centromeric marker). These markers both map to the same intron within the gene *GOLGA4* on 3p21 (see Figure 5.6). This means that the breakpoint is within the gene and therefore that a fusion gene is a possible outcome of this translocation.

The Genbank entry provided for *GOLGA4 via* the RefSeq database (accession NM_002078) was presumed to be the full-length *GOLGA4* transcript. This transcript is 7694 bp in length, in agreement with the approximately 7.7 kb transcript seen by Northern analysis (Erlich *et al* 1996), and was determined to have 24 exons by BLASTN analysis against genomic sequence databases. However, the Entrez MapViewer shows only the 11 exons at the 3' end of the transcript. This inconsistency correlates with the difference between finished sequence and draft sequence on the Entrez Map. These 11 exons match the finished genomic sequence which contains the breakpoint detailed above (accession AP000493). The remainder of the transcript does not match finished sequence and the region in which these exons should be shown is covered by unfinished sequence. The numbering of exons that will be used for purposes of discussing the *GOLGA4* gene structure will be that of the 24 exons determined by BLASTN analysis. By this numbering, the breakpoint occurs within intron 20, between exons 20 and 21 (see Figure 5.6). This corresponds to between bp 6757 and 6758 of the NM_002078 transcript sequence.

A) *GOLGA4*



B) *MEL1*

Figure 5.6. Exonic Structure of *GOLGA4* and *MEL1*.

A) The intron/exon structure of each of the gene *GOLGA4* is shown in blue. A blue box represents an exon. A gap between boxes indicates an intron. The top line is the full length gene. The second line is a close up view of exons 15 to 24 of the gene, as indicated. Exon size is represented roughly proportionally, but the figure is not to scale. The red box indicates the portion of the gene which codes for the coiled-coil domain of the *GOLGA4* protein. The orange box indicates the portion of the gene which encodes the GRIP domain. The black pentagons represent the position and orientation of RT-PCR primers with the name of the primer immediately above the pentagon. The arrows indicate the position of the breakpoint in Patient 3 relative to the exons of *GOLGA4*.

B) The positions of the *MEL1* primers used for the PCR to detect a potential fusion transcript between *GOLGA4* and *MEL1* are indicated with respect to the structure of the *MEL1* gene. A green box represents an exon of *MEL1*. Exon size is represented roughly proportionally, but the figure is not to scale. Note that the primers are at either end of the gene, to provide the maximum possibility of detecting any fusion products.

5.3 Northern Analysis

Given that the breakpoint on chromosome 3 in Patient 3 is within the *GOLGA4* gene, a fusion gene involving *GOLGA4* is a possible outcome of the translocation. Northern analysis using RNA probes for the *GOLGA4* mRNA sequence was conducted with the aim of detecting an aberrant transcript which was specific to Patient 3 and which would therefore represent a fusion transcript. As the location of the breakpoint is known precisely with respect to the coding region of the gene (that is, the breakpoint region is contained entirely within one intron), it was possible to design probes on either side of the breakpoint. In the event of a fusion gene resulting from a chromosome translocation, often only one of the two potential fusion transcripts is expressed. Designing one Northern probe for either side of the breakpoint ensures that a fusion gene will be detected if it is transcribed, and also allows determination of the portion of the gene which is present in the fusion transcript.

The position of the primers used to generate the probes GOLGA4#2 (5' probe, amplified using primers rGOLGA4f2 and rGOLGA4r2) and GOLGA4#3 (3' probe, amplified using primers rGOLGA4f3 and rGOLGA4r3) within the coding region of the *GOLGA4* gene is illustrated in Figure 5.6. Patient 3 bone marrow RNA, as well as RNA from two bone marrow samples from normal donors (in which *GOLGA4* had previously been shown to be expressed by RT-PCR analysis (data not shown)), was electrophoresed and transferred to a nylon membrane as described in Section 2.2.4. RNA probes were hybridised to the membrane and signal detection was conducted as described in Section 2.2.4.

Hybridisation with the GOLGA4#2 probe, 5' of the breakpoint in Patient 3, was unsuccessful in all samples after several attempts. The result of the hybridisation with GOLGA4#3 probe, 3' of the breakpoint in Patient 3, is presented in Figure 5.7.

The major band has run slightly differently in each lane. This may be attributable to buffering problems in the electrophoresis of the samples or unequal loading of the samples. However, it is clear that there is one major band at approximately 7.7 kb in each lane, which corresponds to the expected size of the full-length *GOLGA4* transcript (7694 bp, based on NM_002078). There is also a band at approximately 8.5 kb present only in the lane containing Patient 3 RNA, which possibly represents a fusion transcript.

5.4 Investigation of a Fusion Between *GOLGA4* and *MEL1*

As there appears to be a fusion gene created by fusion of the 3p21 gene *GOLGA4* with a 1p36 gene, it is reasonable to examine the possibility that the 1p36 gene in question is *MEL1*. *GOLGA4* runs 5' to 3' from telomere to centromere, as does *MEL1*. Therefore the t(1;3)(p36;p21) could potentially result in in-frame fusion of these two genes. *MEL1* is known to be upregulated in the t(1;3)(p36;q21) and in the inv ins (1;12)(p21-p36;p13) (see Figure 3.15). *MEL1* has also been shown to be expressed in Patient 3 (see Figure 3.15), and this was considered likely to be a result of the t(1;3)(p36;p21) carried by Patient 3. As discussed in Section 3.5.2.2, however, there may be a non-translocation mechanism for transcriptional activation of *MEL1* in AML, so the expression of *MEL1* in Patient 3 may be due to this second mechanism and therefore unrelated to the t(1;3)(p36;p21). *MEL1* is related to *MDS1-EVI1* (see

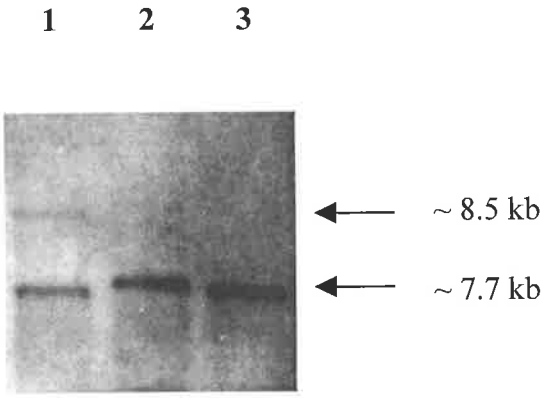


Figure 5.7. *GOLGA4* Northern Analysis

Result of Northern Analysis of Patient 3 bone marrow RNA and two normal bone marrow samples, using a probe designed to flank the known *GOLGA4* breakpoint in Patient 3. The lanes have run slightly unevenly. This is expected to be due to either unequal loading or buffering problems. A dominant band can be seen at approximately 7.7 kb in every lane. This corresponds to the expected normal transcript of *GOLGA4*. A second band, at approximately 8.5 kb, in the Patient 3 lane, indicates a likely fusion transcript. The loading order is provided below.

1. Patient 3 bone marrow RNA
2. Normal bone marrow RNA donor 1
3. Normal bone marrow RNA donor 2

Section 3.1.1), which is involved in non-fusion translocated-mediated transcriptional upregulation as well as fusion transcripts. While the involvement of *MEL1* in a fusion gene is yet to be demonstrated in any translocation, the possibility, given the occurrence of *MDS1-EVI1* fusion genes, was considered significant enough to investigate. This possibility was therefore examined by RT-PCR.

Long Template PCR was performed with primers MEL1f3 and GOLGA4r3, which would detect a fusion between 5' *MEL1* and 3' *GOLGA4*, or with primers GOLGA4f2 and MEL1r2, which would detect a fusion between 5' *GOLGA4* and 3' *MEL1*. The two *GOLGA4* primers were chosen with respect to the position of the breakpoint, such that the reverse primer was 3' of the breakpoint and the forward primer was 5' of the breakpoint. Any breakpoint within *MEL1* was unknown, so primers were positioned as close as was feasible to the 5' and 3' end of the full length *MEL1* transcript. The positions of these primers within the coding regions of *GOLGA4* and *MEL1* are illustrated in Figure 5.6. Full length Patient 3 bone marrow cDNA, prepared using the protocol for first strand cDNA synthesis for 3' RACE (Section 2.2.3.6), was used in this PCR. Control primer combinations GOLGA4f2 with GOLGA4r3 and MEL1f3 with MEL1r2 were used to ensure that the PCR conditions were suitable for all primers. The result of this PCR is shown in Figure 5.8.

A normal *MEL1* product of 4278 bp was successfully amplified as a control. A shorter product corresponding to the *MEL1* splice variant characterised in Section 3.4.1.1 was also amplified but is not distinguishable in Figure 5.8 as the resolution of the gel is insufficient to separate the two *MEL1* bands. The second band was visualised after the gel had been run substantially further (data not shown). A normal *GOLGA4* product

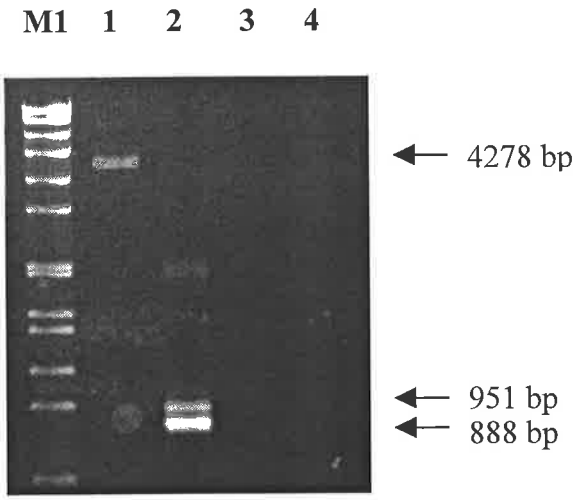


Figure 5.8. *GOLGA4-MEL1* Fusion RT-PCR.

RT-PCR performed with various combinations of primers for *GOLGA4* and *MEL1* to attempt to amplify a fusion product of these two genes. Two normal *GOLGA4* products (951 bp and 888 bp) and normal *MEL1* product (4278 bp) are the only products apparent on the gel. A second band is present in the normal *MEL1* lane as a result of alternate splicing, described in Section 3.4.1.1. The band cannot be distinguished from the full length band at this resolution, however. The shorter *GOLGA4* band is a result of alternate splicing and is discussed further in Section 5.4.1. All PCRs used full length Patient 3 bone marrow cDNA, prepared using the protocol for first strand PCR for 3' RACE, as template. PCRs were performed under Long Template conditions. The loading order is specified below.

M1. SppI/*Eco*RI DNA size marker.
1. MEL1f3 - MEL1r2
2. GOLGA4f2 - GOLGA4r3
3. MEL1f3 - GOLGA4r3
4. GOLGA4f2 - MEL1r2

of 951 bp was also amplified as a control. All three of these products were of the expected sizes. A second, slightly shorter product was amplified in the *GOLGA4* control PCR. This product was determined to be the result of splice variation and is discussed in more detail in Section 5.4.1. No fusion product was amplified using either combination of *MEL1* and *GOLGA4* primers. The maximum size of any such fusion product would be less than 5 kb, which is well within the amplifiable range under these conditions. Therefore, it was concluded that a *GOLGA4-MEL1* fusion transcript is not expressed in Patient 3 as a result of this rearrangement.

5.4.1 *GOLGA4* Splice Variant

A second band was amplified as a result of the *GOLGA4* control PCR utilised in Section 5.4. This band was purified and sequenced, and revealed to be a splice variant which was represented by sequences in the Genbank database and which confirms data previously reported (Kjer-Nielsen *et al* 1999). The full-length transcript is represented in the Genbank database by accession NM_002078 and the variant by accession U31906. The fact that this variant was not detected in the *GOLGA4* Northern analysis in Section 5.3 is not surprising, as the size difference is negligible with respect to the length of the full-length transcript and segregation of the two transcripts would not be expected. The shorter variant band was found to be created by the splicing out of the 63 bp exon 23, accounting for the size difference of the two bands in Figure 5.8. Exon 23 contains the stop codon for the full-length transcript, and the shorter splice variant would therefore translate part of exon 24 before reaching a different stop codon. Each variant protein therefore has a different COOH terminus. The sequences of these are illustrated in Figure 5.9. Both COOH-terminal sequences

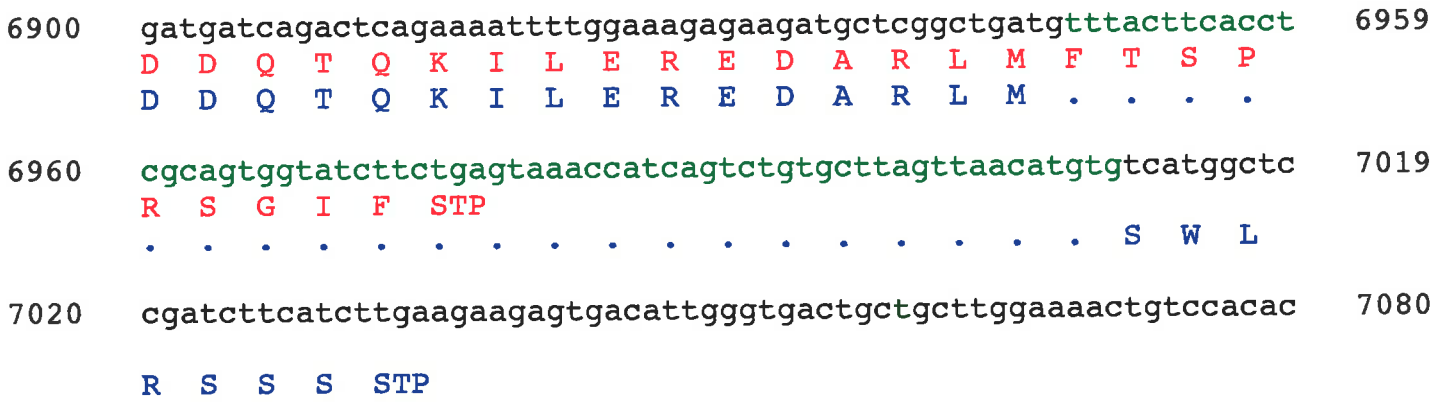


Figure 5.9. *GOLGA4* Splice Variation.

Sequence differences between the two variant splice forms of *GOLGA4*. The splice variation is caused by the variable inclusion of exon 23, the sequence of which is represented in green. Sequences from exons 22 and 24 are represented in black and are present in both variants. The translation shown in red is the protein sequence encoded by the full-length transcript. The stop codon for this translation is contained within exon 23 as shown. The shorter variant therefore has a different stop codon, in exon 24, and the protein sequence encoded by this variant is represented in blue. Numbering of the transcript sequence is according to Genbank accession NM_002078.

were analysed by PRODOM (http://prodes.toulouse.inra.fr/prodom/doc/prodom.html) and PSORT II (http://psort.nibb.ac.jp/), but no sequences of biological significance were found using these programs. This is consistent with the finding that both variant proteins localise to the Golgi complex correctly (Kjer-Nielsen *et al* 1999). The biological relevance, if any, of this splice variant and its expression is therefore unknown.

5.5 3' RACE Analysis

For reasons discussed in Section 5.7.2, it is more likely that the biologically relevant fusion transcript would involve the 5' end of *GOLGA4*, rather than the 3' end, which was shown to be involved in a fusion transcript *via* Northern analysis in Section 5.3. As Northern analysis with RNA probe GOLGA4#2, which is 5' of the breakpoint, was unsuccessful, an alternate method to investigate the presence of a fusion transcript involving the 5' end of *GOLGA4* was to use 3' RACE analysis (see Section 2.2.3.6). This is a method of obtaining the 3' end sequence of transcripts for which only 5' sequence is known. This is ideal for characterisation of fusion transcripts where only one partner is known. Ideally, the identification of a fusion transcript would occur by observing a band in the Patient 3 lane that does not exist in the control lanes.

In this instance, the precise position of the breakpoint was known. Primers were therefore positioned 5' of this breakpoint. Primers used for 3' RACE analysis were GOLGA4f2 in the primary reaction and GOLGA4f4 in the secondary reaction (see Figure 5.6). Full length cDNAs, from Patient 3 bone marrow, two normal bone marrow samples and two normal peripheral blood samples, were prepared using the

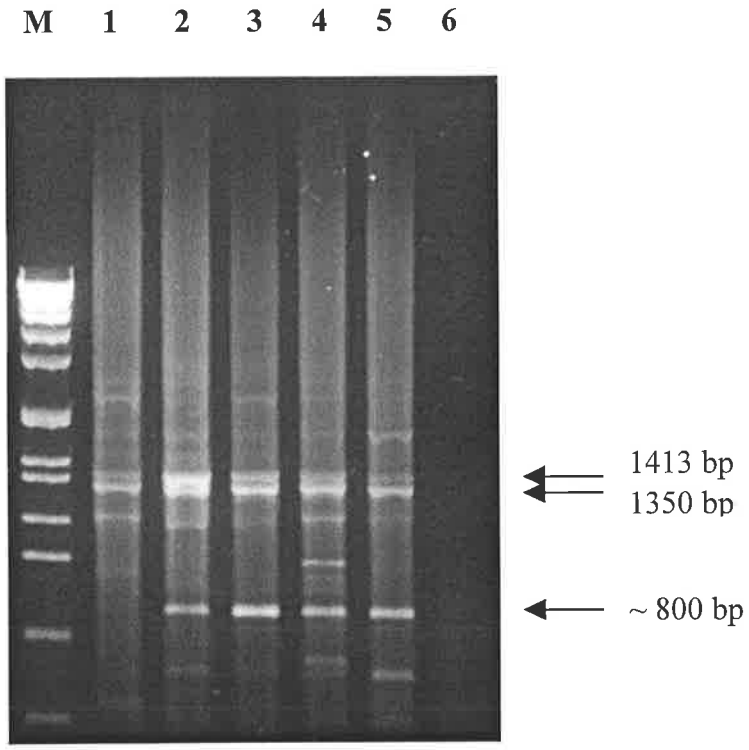
protocol for first strand PCR for 3' RACE described in Section 2.2.3.4. The reaction was performed as described in Section 2.2.3.6 and the results are displayed in Figure 5.10.

One of the major bands amplified and present in every lane corresponds with the expected size of 1413 bp and is arrowed in Figure 5.10. A second band of 1350 bp is also present in every lane and corresponds to the *GOLGA4* splice variant identified in Section 5.4.1. Every band present in the Patient 3 lane is also present in at least one of the control lanes. This indicates that the presence of a fusion transcript containing the 5' end of *GOLGA4* is unlikely. As no aberrant bands could be detected under these conditions, further analysis of the amplified products was not pursued.

Interestingly, the major difference between the bands present in the Patient 3 lane and the control lanes was the absence of a band in the Patient 3 lane. The band, which is of approximately 800 bp in length and is strongly amplified in each of the controls, was not amplified from Patient 3 cDNA in several repeats of the experiment. A control PCR performed using only the AUAP primer (that is, with no gene-specific primer) also resulted in the amplification of this band (data not shown). The band is therefore the result of non-specific priming, and hence unrelated to *GOLGA4*. The band was therefore not characterised further.

M. SppI/*Eco*RI DNA size marker.
1. Patient 3 cDNA.
2. Normal bone marrow donor 1.
3. Normal bone marrow donor 2.
4. Normal PBMNC donor 1.
5. Normal PBMNC donor 2.
6. H_2O negative control.

A)



B)

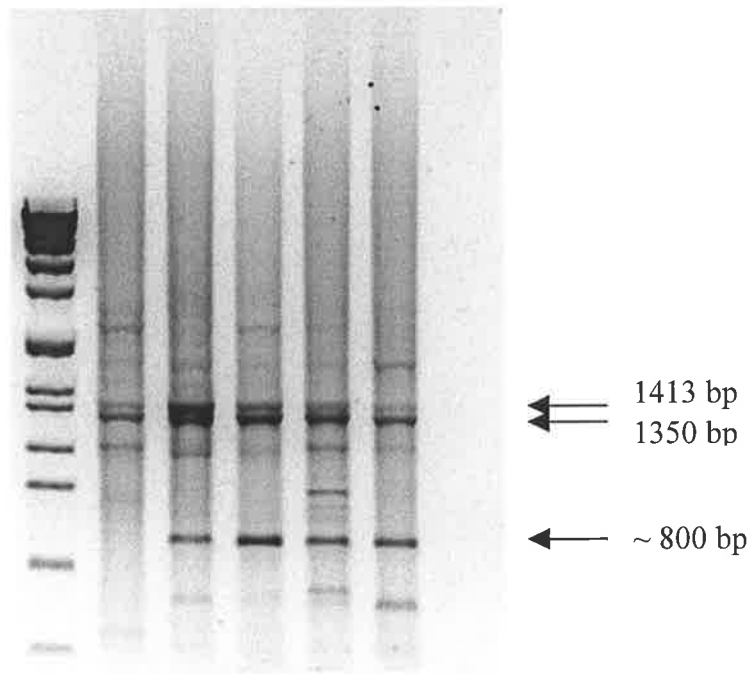


Figure 5.10. 3' RACE Analysis of *GOLGA4* Transcripts in Patient 3.

3' RACE was performed as described in Section 2.2.3.6, using primers rGOLGA4f2 in the primary reaction and rGOLGA4f4 in the secondary reaction. The indicated bands are discussed in the text. Panel A is the normal EtBr stained agarose gel. Panel B is the inverted image to allow easier visualisation of bands. The loading order is presented on the opposite page.

5.6 Inverse PCR

As RNA based methods did not yield informative results, an inverse PCR was designed to identify the partner sequence of the 3p21 breakpoint in Patient 3. The positive marker closest to the 3p21 breakpoint on the telomeric side, 493AP16, was used as a starting point. Restriction enzymes were chosen on the basis of availability and proximity of recognition sequences to 493AP16 on the telomeric side. The absence of recognition sequences on the centromeric side was also taken into account. The three enzymes chosen were *Ava*II, *BstU*I and *Xba*I. The design of this inverse PCR can be seen in Figure 5.11.

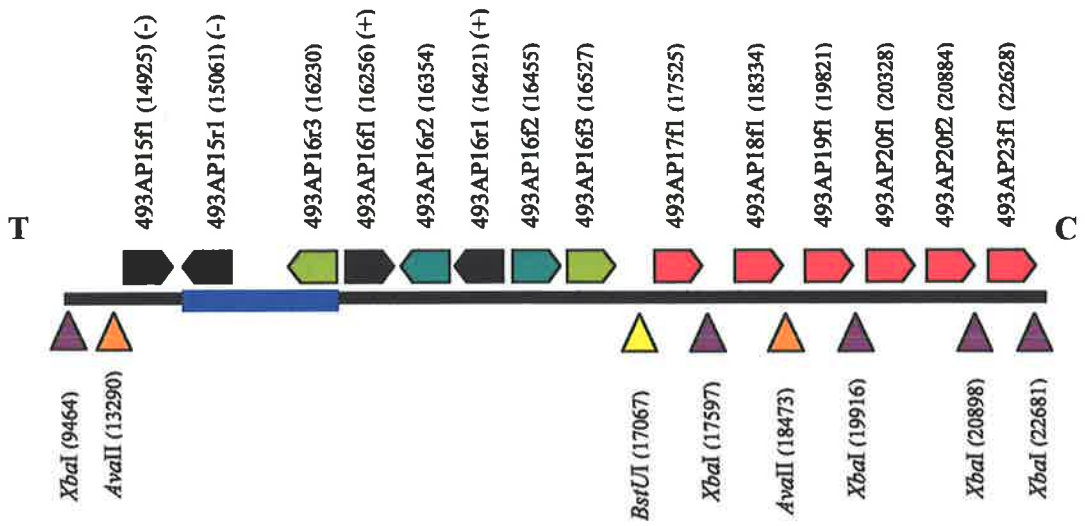


Figure 5.11. Design of Inverse PCR to Identify Partner Sequence of the 3p21 Breakpoint. The black line represents the genomic sequence of the 3p21 region. The letters T and C represent the telomeric and centromeric end of the region respectively. The blue box represents the breakpoint region as identified through somatic cell hybrid PCR walking. + and - indicate the presence or absence of markers tested against LM9. Primers used in either PCR walking, inverse PCR or inverse PCR sequencing are represented by pentagons above the line. Primers are colour-coded as follows: black for PCR walking primers, dark green for primary inverse PCR primers, light green for secondary inverse PCR primers and red for sequencing primers. Restriction sites utilised in the inverse PCR are represented by triangles below the line. Numbers in parentheses following a restriction enzyme site or primer name indicate the position of the object. Numbering is in accordance with accession AP000493.1. The figure is not to scale.

Inverse PCR was performed on Patient 3 DNA as described in Section 2.2.3.7. Primers used were 493AP16f2 and 493AP16r2 in the primary reaction, and 493AP16f3 and 493AP16r3 in the nested secondary reaction. The result of this inverse PCR is presented in Figure 5.12.

The expected sizes of products generated from normal, non-rearranged chromosomes were predicted to be 4886 bp from *Ava*II digested template and 7836 bp from the *Xba*I digested template. A 7836 bp product is at the outer limit of the amplification range of the Long Template PCR system used and may not be successfully amplified for this reason. There are no *BstU*I sites on the telomeric side of the breakpoint region which are within amplifiable range, so no product generated from a normal chromosome was expected to be amplified from *BstU*I template.

Two bands were amplified from the *Ava*II digested template. One of these, at approximately 5 kb, was of the size expected for a product amplified from the normal chromosome. This band was isolated, and sequencing confirmed that this was the expected product from the normal chromosome 3. The second band, at 900 bp, is shorter than any product amplified from a correctly digested template should be, as any product should be a minimum of 1946 bp, the distance from the 493AP16f3 primer to the *Ava*II restriction site (see Figure 5.11 and Section 5.7.5). Attempts to sequence this band with the 493AP18f1 primer were unsuccessful. It was concluded that this product was the result of an artifact of either the PCR or the template restriction and ligation process. It was therefore not analysed further.

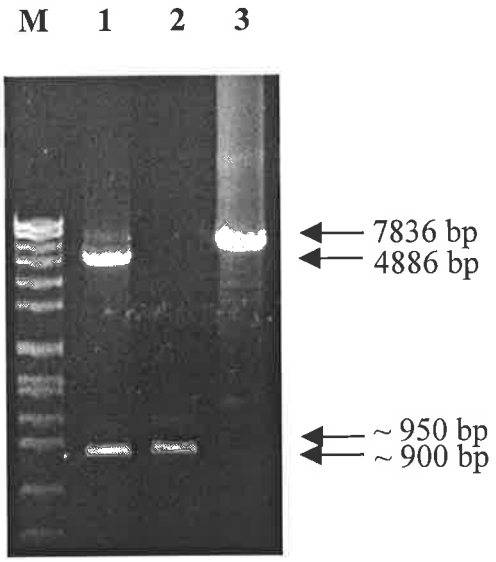


Figure 5.12. Inverse PCR to Identify Partner Sequence of 3p21 Breakpoint.

Templates were prepared from Patient 3 DNA by restriction with the indicated restriction enzyme followed by circularisation. The secondary PCR from templates prepared using each of the three restriction enzymes is shown. Primary PCR was performed using primers 493AP16f2 and 493AP16r2. Secondary PCR was performed using 493AP16f3 and 493AP16r3, and 2 µl of a 1:50 dilution of the primary PCR as template. The indicated bands are discussed in the text. The loading order is specified below.

M1. SppI/*Eco*RI DNA size marker.
1. *Ava*II secondary inverse PCR.
2. *Bst*UI secondary inverse PCR.
3. *Xba*I secondary inverse PCR.

A single band of 950 bp was amplified from *BstU*I digested product. This band was purified and sequenced with the primer 493AP16f3. The sequencing data indicated that the restriction site was intact, and therefore the entire set of sequencing data corresponded to the normal chromosome 3.

A single band of approximately 7 kb was amplified from the *Xba*I digested template. This is shorter than the size expected of the product generated from the normal chromosome. The 7 kb product was isolated and sequenced with the primer 493AP17f2, and the *Xba*I site at position 17597 was shown to be intact. Because *Xba*I restriction can be affected by the methylation status of the recognition site, it was considered possible that while this site was refractory to restriction, other sites may not be. There were three other *Xba*I sites within the length of normal chromosome 3 that could theoretically be within the 7 kb inverse PCR product. Primers were designed such that each of these sites could be sequenced within this product (see Figure 5.11). Sequencing with the primer 493AP19f1 showed that the restriction site at position 19916 was also intact. Sequencing with the primers 493AP20f2 and 493AP23f1 failed, probably indicating that the DNA break that led to circularisation of the template took place between the end of the sequence obtained with 493AP19f1 and the position of the 493AP20f2 primer. To attempt to confirm this, sequencing was conducted with the 493AP20f1 primer, which falls within this region. Sequencing with this primer was successful, and the sequence corresponded to normal chromosome 3 sequence up until the quality of the data diminished, 56 bp from the binding site of 493AP20f2. The DNA break which led to circularisation must therefore have taken place within this 56 bp region. As the break was clearly illegitimate (in that it was not due to restriction at a recognised site), any

determination made from this product as to the nature of the partner sequence of the translocation would be highly unreliable, and analysis of this product was not pursued further.

5.7 Discussion

5.7.1 Identification of the 3p21 Breakpoint within *GOLGA4*

Somatic cell hybrid chromosome walking analysis identified the 3p21 breakpoint in the t(1;3)(p36;p21) of Patient 3 within intron 20 of the *GOLGA4* gene, also known as *p230* or *golgin-245*. *GOLGA4* is a member of the golgin family of genes, which encode proteins which are localised to the Golgi complex, and which play a poorly understood role in facilitating protein transport, presumably by involvement with vesicular docking or fusion (Erlich *et al* 1996, Kjer-Nielsen *et al* 1999). The full-length 2231 amino acid GOLGA4 protein contains two major domains. A very large coiled-coil domain, which may be more correctly defined as several domains, spans the first approximately 2100 amino acids, which are encoded by bp 286 - 6585 on NM_002078. This was confirmed by analysis using COILS 2.1, available at http://www.ch.embnet.org/software/COILS_form.html. The protein also contains a GRIP domain as reported in the annotation for Genbank accession NP_002069. GRIP domains have been shown to be sufficient for targeting to the Golgi complex (Munro and Nichols 1999). The GRIP domain is encoded by NM_002078 bp 6796 - 6930. The breakpoint within intron 20, at bp 6757 of NM_002078, separates these two functional domains, such that the 5' potential fusion transcript would contain the

coiled coil domain, and the 3' potential fusion transcript would contain the GRIP domain (see Figure 5.6).

GOLGA4 has not been previously implicated in a fusion gene or in leukaemia. The only previous report of the involvement of any golgin gene in cancer is that of a fusion gene between *GOLGA5* and *RET* in a papillary thyroid carcinoma (PTC) from a patient exposed at a young age to radioiodine released from the Chernobyl reactor (Klugbauer *et al* 1998). *RET* rearrangements are common in post-Chernobyl PTCs, but rare in spontaneous PTCs (Rabes *et al* 2000). The *RET* proto-oncogene is located on chromosome 10q11.2 and is a member of the cadherin superfamily. It encodes a receptor tyrosine kinase, and is involved in the transduction of signals required for cell growth and differentiation. *GOLGA5* is located at chromosome 14q32. It has not been implicated in any other chromosome rearrangements or in any cancer other than its involvement in the PTC *RET* fusion. In this rearrangement, both reciprocal fusion products were expressed (Klugbauer *et al* 1998). In *RET* fusions generally, however, it is the 3' portion of the *RET* transcript, which contains the tyrosine kinase domain, that is contained within the biologically active fusion transcript (Rabes *et al* 2000). Presumably, therefore, the *GOLGA5-RET* fusion gene is the biologically active outcome of the rearrangement.

5.7.2 Confirmation of the Presence of a *GOLGA4* Fusion Gene in Patient 3

Two RNA probes intended for use in Northern analysis were designed, one on either side of the breakpoint, as illustrated in Figure 5.6. This strategy was chosen in order to detect aberrant transcripts involving either end of *GOLGA4*, and was possible because

definition of the breakpoint region within a single intron was achieved prior to Northern analysis being performed.

The 5' probe, GOLGA4#2, failed to produce any signal despite repeated hybridisation attempts. This is possibly due to a fault with the manufacture of the probe, although numerous tests were performed to demonstrate that the probe was labelled and of the expected size. The PCR product from which the RNA probe was generated was also fully sequenced to confirm the identity of the product prior to transcription. The lack of signal could not be due to a lack of expression of *GOLGA4* in the samples used to construct the membrane, as expression of normal *GOLGA4* was detected by RT-PCR on the same samples, and also because expression was confirmed using the same membrane and the GOLGA4#3 probe. This result also rules out a fault with membrane construction, although it is possible that the stripping of the membrane following hybridisation with the GOLGA4#3 probe damaged the membrane in such a way as to prevent subsequent successful hybridisation.

Results obtained with the GOLGA4#3 probe, 3' of the breakpoint, show an aberrant transcript of approximately 8.5 kb. Given the position of the breakpoint, 921 bp of this transcript should correspond to *GOLGA4* sequence, meaning that another 7.5 kb is contributed to the transcript by the 5' fusion partner. The aberrant band is considerably weaker than the major *GOLGA4* band, indicating that the fusion transcript is not expressed as strongly. This is not overly surprising, as *GOLGA4* appeared to be highly expressed by RT-PCR analysis (data not shown), and as the detected fusion transcript would be under the regulatory control of the promoter sequence of the fusion partner gene. Given the arrangement of the domains of

GOLGA4, the fusion gene detected on this Northern would retain the GRIP domain but not the coiled-coil domain. This fusion transcript may not be the leukemogenic outcome of the translocation, as there are several reasons to suspect that the reciprocal transcript may be more significant. A leukemogenic model for fusion transcripts encoding coiled-coil domains which enable oligomerisation and hence recruitment of transcriptional coactivators has been demonstrated for fusions of both RARα and AML1 (Minucci *et al* 2000). Also, based on the intensity of the bands on the Northern, expression of the fusion transcript could be expected to be higher from the *GOLGA4* promoter region, in the situation where expression level may be relevant to the leukemogenic function of the fusion transcript. Furthermore, although the *RET-GOLGA5* fusion resulted in transcription of both reciprocal fusion transcripts, the majority of *RET* fusions involve the 3' portion of *RET*, and so presumably it is the transcript containing this (and therefore the 5' portion of *GOLGA5*) which is biologically active. This further supports the notion that the reciprocal product, involving the 5' portion of *GOLGA4*, may be expressed in Patient 3.

5.7.3 Investigation of a Fusion Between *GOLGA4* and *MEL1*

The *MEL1* gene, which is dysregulated in leukaemia by the t(1;3)(p36;q21) and which is known to be expressed in Patient 3 (see Figure 3.15), may potentially be involved in a fusion with *GOLGA4*. The 1p36 breakpoint is unknown, but given the proximity of *MEL1* to the breakpoints in both the t(1;3)(p36;q21) in several patients (Mochizuki *et al* 2000; this report) and the inv ins (12;1)(p13;p21-p36) in Patient 2, it seemed feasible that *MEL1* might be involved in the t(1;3)(p36;p21) in Patient 3.

Long Template PCR was performed with combinations of *MEL1* primers and *GOLGA4* primers which would amplify normal transcripts and either of the two possible reciprocal fusion transcripts if present. However, no fusion product was amplified by this RT-PCR. A normal *GOLGA4* product was amplified as a control, as was a normal *MEL1* product. The amplification of these products indicate that the primers will amplify in the conditions used for the PCR, and that they will amplify from this Patient 3 material. Therefore, it was concluded that a *GOLGA4-MEL1* fusion transcript is not expressed in Patient 3 as a result of this translocation. This does not exclude the possibility that *MEL1* upregulation may be a result of the t(1;3)(p36;p21) in Patient 3, however. *MEL1* involvement in translocations reported to date has been limited to upregulation without fusion (see Chapters 3 and 4). Also, translocations resulting in the formation of a fusion gene, which is apparently non-functional, in combination with upregulation of an unrelated gene have been demonstrated (Cools *et al* 2002, Odero *et al* 2002). The possibility that the *GOLGA4* fusion is non-functional, and that the biologically relevant outcome of the translocation is upregulation of *MEL1* transcription, must be considered. However, *MEL1* expression in Patient 3 may not be an outcome of the translocation, but rather be caused by a different mechanism (see Section 3.5.2.2).

5.7.4 3' RACE Analysis of *GOLGA4* Transcripts

RACE is a technique designed to amplify the ends of cDNA sequences of which only partial sequence is known. This is ideal for amplifying fusion transcripts where only one fusion partner is known. This technique generally gives rise to a complex pattern of amplification for two major reasons. Firstly, the reverse primer used in both the

primary and secondary reactions is one that of necessity amplifies from the 3' end of all transcripts, and the specificity of the technique is therefore quite low. For this reason, ten rounds of single-primer amplification are performed prior to the addition of the AUAP primer, to enrich the reaction for gene-specific sequences (see Section 2.2.3.6). However, this modification of the technique does not fully remove the lack of specificity from the reaction. Secondly, even transcripts which are amplified specifically may be of several lengths because of splice variation and incomplete processing of transcripts. The use of normal controls allows identification of transcripts which differ in size from normal transcripts, and these can be isolated and sequenced to identify the partner gene.

When 3' RACE was performed using *GOLGA4* primers, however, all products present in the Patient 3 lane were apparently present in at least one of the control lanes. This result indicates that there may not be a fusion transcript involving the 5' end of *GOLGA4*, but does not exclude the possibility that such a fusion transcript exists. There may be a fusion transcript that is of a similar size to one (or more) of the bands resulting from the normal *GOLGA4*, such that the band is difficult to distinguish on the gel. An examination of the intensity of the full-length normal band and the splice variant band suggest that this is unlikely, however, as in all cases, including Patient 3, the ratio of the intensity of the lower band to the intensity of the upper band is similar. A fusion transcript which co-migrated with one of these bands would be expected to increase the relative intensity of that band, as any fusion transcript is likely to co-migrate with only one of these bands. The alternative splice variation occurs 3' of the breakpoint in Patient 3 and so will not be present in a fusion transcript containing the 5' portion of *GOLGA4*. Another possibility is that the fusion transcript may be

significantly larger than the normal transcript, leaving the abundantly expressed normal transcript to be amplified preferentially due to its shorter length. If, as the 3' RACE results suggest, there is no fusion transcript involving the 5' end of *GOLGA4*, this does not conflict with the Northern analysis result, as the aberrant band seen in the Northern involved *GOLGA4* sequence 3' of the breakpoint.

The major difference between the bands amplified from the Patient 3 sample and the control samples was the absence of a band in Patient 3. The band was also amplified by a control PCR using only the AUAP primer (that is, with no gene-specific primer). The band is therefore non-specific, and unrelated to *GOLGA4*. Nonetheless, the absence of the band from the Patient 3 lane is difficult to explain. As it had been shown to be unrelated to *GOLGA4*, however, this band was not characterised further

5.7.5 Inverse PCR Analysis of 3p21 Breakpoint

The inverse PCR analyses performed on this breakpoint were unsuccessful for varied reasons. Analysis with each enzyme will be discussed separately.

Two specific products were amplified from the *Ava*II digested template. One of these was of the correct size (approximately 5 kb) to have been generated from the normal copy of chromosome 3, and this was confirmed by sequencing. The second band was of an inappropriate size to have amplified from correctly digested template as, at 900 bp, it was too short to have been restricted at the expected site. The known *Ava*II recognition site was 1946 bp from the nested inverse PCR primer 493AP16f3 binding site and so any product amplified from a correctly digested template would therefore

be at least this long. The failure of the sequencing reaction with the 493AP18f1 primer supported the expectation that the sequence corresponding to 493AP18f1 was not present in the product. This indicates that breakage and religation of the template DNA had taken place between 493AP16f3 and 493AP18f1. This may have occurred due to a polymorphic restriction site in this region, which was present in Patient 3 but not in the Genbank database entry. Alternately, the DNA breakage may have been due to shearing during manipulation, or possibility due to star activity of the restriction enzyme. As this product was not amplified from correctly constructed template, characterisation was not pursued further.

*BstU*I digested template allowed amplification of only a single band of approximately 950 bp. No band was expected to be amplified from the normal chromosome, and so any band produced was likely to be derived from the translocated chromosome. In this case, this band was not shorter than expected, as the *BstU*I site was only 540 bp from the nested inverse PCR primer 493AP16f3 binding site, and the amplified product was larger than 540 bp. Sequencing of the amplified product, however, revealed that the *BstU*I site was intact in the product, and that the full length of sequenced product (approximately 1 kb) was derived from the normal chromosome 3 sequence. No explanation could be found for the failure of this enzyme to restrict at this particular site. As for the nature of the DNA break that led to circularisation of the template, the explanation may be similar to that for the *Ava*II template that gave rise to the shorter amplification product.

The *Xba*I digested template produced a single band of approximately 7 kb. This is shorter than the size expected of the product generated from the normal chromosome,

which may not have been amplified as, at nearly 8 kb, it is at the outer limit of the Long Template PCR system under the conditions employed. The 7 kb product therefore was likely to have been amplified from template generated from the translocated chromosome. However, sequencing revealed that the *Xba*I site expected to be cut was intact. *Xba*I restriction is known to be blocked by DNA methylation, so this is a possible reason that this restriction may have failed. More *Xba*I sites were present further along the sequence, and primers were designed to allow for the sequencing of these restriction sites also (see Figure 5.11). Sequence corresponding to the expected normal chromosome was returned from sequencing reactions with each of 493AP18f1, 493AP19f1 and 493AP20f1. However, no sequence data was returned from sequencing reactions with 493AP20f2 or 493AP23f1, presumably due to the absence of the primer binding sites within the product. Sequence obtained with 493AP20f1 went to within 56 bp of the binding site for 493AP20f2 before the sequence quality diminished. Therefore, as the 493AP20f2 binding sequence is not present within this product, there must have been a DNA break within this 56 bp region. The cause of the break is unknown. Again, it may be due to DNA shearing during manipulation, or to restriction enzyme star activity.

The inverse PCR methodology therefore failed to identify the partner sequence of the 3p21 breakpoint. Clearly template preparation was not ideal, and a repeat of the experiment, with the possibility of the use of different enzymes, is worthwhile. Unfortunately, time constraints prevented performing the experiment again.

5.8 Conclusions

The 3p21 breakpoint in the t(1;3)(p36;p21) in Patient 3 was mapped by somatic cell hybrid PCR walking and determined to be within intron 20 of the *GOLGA4* gene. Northern analysis of *GOLGA4* using a probe 3' of the breakpoint demonstrated the existence of a fusion transcript of ~ 8.5 kb. Strategies for the identification of the partner chromosome sequence and partner gene were unsuccessful. Future strategies may include: (1) inverse PCR analysis with different enzymes; (2) 5' RACE analysis; and (3) construction of a genomic library from Patient 3 material, to be screened with sequences adjacent to the known 3p21 breakpoint and for identified clones to then be sequenced.

CHAPTER 6

Conclusions

CHAPTER 6

Conclusions

6.1 Heterogeneity of 1p36 Translocation Breakpoints

This study involved the mapping of three distinct chromosome translocations, all involving 1p36. Two of these translocations, ins(12;1)(p13;p36p21) and t(1;3)(p36;q21), were fully characterised, with the 1p36 breakpoints in both cases found to lie in the same breakpoint cluster region, 5' of the *MEL1* gene at 1p36.3. The patients with these two translocations were found to express *MEL1*, and this was attributed to the chromosomal rearrangement in each case. *MEL1* expression was also seen in patients without 1p36 translocation, however, so there must be a second mechanism of *MEL1* expression in leukaemia. In the case of the third translocation investigated, t(1;3)(p36;p21), only the 3p21 breakpoint was mapped. However, the patient carrying the translocation was also found to express *MEL1*, and so it is possible that the 1p36 breakpoint in this patient also lies in a similar region and is the cause of increased *MEL1* expression. Therefore, from the results presented in this thesis, the suggestion would be that 1p36 translocations in haematological malignancy generally cluster to this region.

A recent publication shows a novel 1p36 breakpoint, however, at 1p36.1 (Odero *et al*, 2002). This breakpoint was found in an MDS patient with t(1;12)(p36;p13), a translocation which involves the recombination of the same chromosome bands as

that of Patient 2 (an ins(12;1)(p13;p36p21)). The Patient 2 rearrangement was originally thought to be a variant of the t(1;12)(p36;p13) and indeed the 12p13 breakpoints were found to be similar in both cases. The 12p13 breakpoint in the Odero *et al* (2002) patient with t(1;12)(p36;p13) is in intron 2 of the *TEL* gene, while the 12p13 breakpoint in the ins(12;1)(p13;p36p21) in Patient 2 was shown to be in intron 1 of *TEL*. However, the subsequent mapping of the 1p36 breakpoint in the t(1;12)(p36;p13) reported in Odero *et al* (2002) placed this breakpoint at 1p36.1, and showed that it resulted in the fusion of *TEL* to a novel gene, *MDS2*. This breakpoint is distinct from the *MEL1* breakpoint cluster region, and so it is clear that there are at least two distinct breakpoints on 1p36 in haematological translocations. Odero *et al* (2002) performed FISH analysis using probes located near the *MDS2* gene in ten patients with other 1p36 chromosomal rearrangements and found that none of these patients had *MDS2* rearrangements. They concluded that *MDS2* was not frequently rearranged, and therefore it may be that the *MEL1* breakpoint cluster region is more frequently involved in translocation than is *MDS2*. The investigation of more patients with distinct 1p36 rearrangements may confirm this.

This situation may be further clarified by the identification of the 1p36 breakpoint in the t(1;3)(p36;p21) in Patient 3. Strategies to achieve this may include the construction of a genomic library from Patient 3 material and screening this with sequences adjacent to the known 3p21 breakpoint. The inserts of the clones known to contain regions proximal to the breakpoint could then be sequenced. This strategy could overcome the artifactual nature of the difficulties encountered with the inverse PCR described in Section 5.6. In addition, 5' RACE analysis may be used to identify the partner gene in the fusion transcript visualised by Northern analysis.

6.2 Molecular Outcomes of 1p36 Rearrangements

Each of the three patients investigated in this thesis were shown to express *MEL1*. The 1p36 breakpoints in the two patients for whom 1p36 breakpoints were characterised were found to cluster in the area 40 - 70 kb 5' of *MEL1*, and so transcriptional upregulation of the gene by juxtaposition to different transcriptional enhancer sequences is a possible mechanism of leukemogenesis. The 1p36 breakpoint of Patient 3 may also be in the same region, and the t(1;3)(p36;p21) may also result in *MEL1* expression, as expression is detected in this patient. However, the issue is clouded by results demonstrating the expression of *MEL1* in patients without 1p36 rearrangements. This expression could be due to either of two mechanisms.

Firstly, as *MEL1* was shown to be expressed at low levels in normal bone marrow, it is possible that the *MEL1* expression observed in leukaemia patients is due to a clonal expansion of a haematopoietic progenitor cell which normally expresses *MEL1*. This hypothesis was supported by the pattern of *MEL1* expression examined by AML subtype, where expression was more common in leukaemias derived from immature haematopoietic progenitor cells. Expression of the gene under these circumstances is less likely to contribute to leukemogenesis as, in the context of the cell, it is not abnormal.

Secondly, there may be an alternate mechanism of activation of *MEL1* transcription in leukaemia. The most likely such mechanism is the aberrant expression of an upstream regulator of *MEL1*, the identity of which is currently unknown. Expression of *MEL1*

in these circumstances is likely to be leukemogenic as the expression occurs in a cell type in which it does not normally occur.

Regardless of which of these two mechanisms occurs in non-translocation patients, there is sufficiently strong correlation between 1p36 rearrangement and MEL1 expression to argue that *MEL1* expression in patients with the t(1;3)(p36;q21) translocation is caused by a mechanism distinct from that occurring in non-translocation patients (see Section 3.4.1). Also, Patient 2 and Patient 3, the only two patients investigated who carried 1p36 rearrangements other than the t(1;3)(p36;q21), showed a quantifiably higher rate of transcription of *MEL1* than did non-1p36 rearrangement patients (see Figure 3.15). An explanation for this effect is that the higher levels of transcription reflect the varying strengths of the transcriptional enhancers to which *MEL1* is juxtaposed in these translocations. Additionally, there are precedents for upregulation of juxtaposed genes by both 3q21 sequences (RPN1 enhancers cause upregulation of *EVI1* in 3q21q26 syndrome; reviewed in Nucifora 1997) and 12p13 sequences (Cools *et al* 2002; Odero *et al* 2002). In combination, this data suggests that *MEL1* transcription is a genuine result of chromosome rearrangement.

In the case of the t(1;3)(p36;q21), an additional molecular outcome of the translocation appears to be transcriptional upregulation of *GATA2*. *GATA2* expression was observed at high levels in the two t(1;3)(p36;q21) patients, as well as in three 3q21q26 patients. *GATA2* expression is also observed at low levels in normal bone marrow as a result of expression in a subset of normal haematopoietic progenitor cells, and so it was not surprising to see it expressed at high levels in a subset of non-

3q21 rearranged patients, presumably as a result of the clonal expansion of haematopoietic progenitor cells in a similar way to that proposed for *MEL1*. In fact, every non-translocation patient who expressed *MEL1* also expressed *GATA2*, supporting this expansion hypothesis for the expression of both genes in non-translocation patients. This suggests that a subset of *GATA2* expressing normal haematopoietic cells also express *MEL1*, and it is the expansion of this population of cells as a result of transformation that results in the leukaemias without 1p36 or 3q21 translocation which express both genes. Similarly to *MEL1*, statistical analysis showed that *GATA2* expression in 3q21 translocation patients occurred by a distinct mechanism than expression in non-translocation patients. The translocation is therefore a likely means of mediation of this transcription, and this is supported by the proposal of a molecular mechanism of dysregulation of *GATA2* transcription *via* separation of a haematopoietic regulatory element from the gene by all known 3q21 translocation breakpoints (reviewed in Wieser 2002). *GATA2* expression *via* 3q21 translocation is an outcome that is apparently shared by patients with either t(1;3)(p36;q21), t(3;3)(q21;q26) or inv(3)(q21q26). Patients with these translocations also share a distinctive clinical phenotype of elevated platelet counts and trilineage dysplasia. This commonality suggests that *GATA2* expression, expressed out of its normal cellular context, plays a role in establishing this phenotype.

6.3 Molecular Mechanisms of Transformation

Direct investigation of the potential mechanisms of transformation as a result of these 1p36 rearrangements was outside the scope of this thesis. However, certain aspects of

the molecular outcomes of the investigated translocations that might contribute to leukemogenesis can be speculated upon.

Two molecular outcomes are likely to result from the t(1;3)(p36;q21). The first of these is upregulation of *GATA2*, caused by separation of the *GATA2* promoter from an upstream haematopoietic regulatory unit. *GATA2* normally plays a role in haematopoiesis, and is expressed in early progenitor cells before being downregulated as the haematopoietic cells begin to differentiate. *GATA2* is thought to be involved in the maintenance of the haematopoietic stem cell population, and therefore inappropriate expression of the gene may result in the inappropriate promotion of immortality within the leukemic blast. Additionally, GATA2 shares partial functional redundancy with GATA1, which is involved in the maturation of erythroid and megakaryocytic lineages and platelet formation. If GATA2 is able to fulfil part of the function of GATA1 in cells in which it is aberrantly expressed, this may be the means by which it contributes to the clinical phenotype of AML patients with either t(1;3)(p36;q21), t(3;3)(q21;q26) or inv(3)(q21q26), which is normal or elevated platelet counts and dysmegakaryopoiesis. The commonality of the 3q21 breakpoint in these translocations, and the upregulation of *GATA2* observed in patients with either of the three translocations, suggest that there is such a contribution from *GATA2* to this phenotype.

The second outcome of the t(1;3)(p36;q21) was upregulation of *MEL1*. *MEL1* is closely related to *MDS1-EVI1*, and although they are likely to regulate the expression of different genes, it is interesting to speculate that, when overexpressed, *MEL1* and *EVI1* have similar downstream effects. EVI1 and MDS1-EVI1 are both variously

reported as repressors and activators of transcription, although the two isoforms are generally regarded as antagonistic (Morishita *et al* 1995, Bartholomew *et al* 1997, Wimmer *et al* 1998, Izutsu *et al* 2001). The transcriptional regulatory ability of MEL1 is yet to be investigated. Both *MEL1* and *MDS1-EVI1* are weakly expressed in normal bone marrow, which is suggestive of a role in haematopoiesis. If *MEL1* expression in non-translocation AML patients is explained by the expansion of a normally expressing haematopoietic progenitor, then it appears that *MEL1* and *EVI1* are not expressed in the same haematopoietic precursors, as the expression patterns of the two genes in the panel of non-translocation patients is different. There could be a suggestion, therefore, that the two genes fulfil similar roles in a different precursor cell type. Nothing more in terms of molecular mechanisms of leukemogenic activity of these genes can be known without knowing the identity of the target genes of MEL1 and EVI1 regulation.

While no target genes have yet been identified for MEL1, the first target gene for EVI1 was identified in a recent publication (Takahashi and Licht 2002). The gene identified is the *promyelocytic leukaemia zinc finger* (*PLZF*) gene, which is known to be involved in leukaemia *via* its fusion to *RARα* by the t(11;17)(q23;q21) (Chen *et al* 1993). The PLZF protein is a transcriptional repressor which is involved in the suppression of cell growth *via* cell cycle arrest (Yeyati *et al* 1999). Transcription of the gene was found to depend on the presence of the EVI1 protein. Therefore, it is possible *EVI1* overexpression may have an impact on cell cycle regulation, which could be a key to its oncogenic function. Intriguingly, PLZF has also recently been shown to promote megakaryocytic development *via* a direct interaction with the GATA1 protein (Labbaye *et al* 2002). This has major implications for the

establishment of the elevated platelet count and dysmegakaryopoietic phenotype that accompanies the t(3;3)(q21;q26) or inv(3)(q21q26) in AML patients. Also, this finding further implicates the GATA genes in the outcome of the translocations. Perhaps, as a result of either the t(3;3)(q21;q26) or inv(3)(q21q26), *EVI1* upregulation activates *PLZF* expression, and PLZF interacts with GATA2, which is also upregulated by the translocation and may fulfil part of the same haematopoietic function as GATA1. This interaction could be responsible for the abnormal megakaryopoiesis seen in patients with these translocations.

It is possible that *MEL1* plays a similar role in the molecular outcome of the t(1;3)(p36;q21) to the one *EVI1* plays in the 3q26 translocations. However, it is impossible to be confident of this without some study of the target genes of MEL1 regulation. If *PLZF* does play a role in the establishment of the clinical phenotype shared by patients with t(1;3)(p36;q21), t(3;3)(q21;q26) or inv(3)(q21q26), however, there must be a mechanism for upregulation of *PLZF* (or another gene with a similar function) which does not involve *EVI1*, and any such mechanism may involve *MEL1*.

It is fascinating that three members (*MEL1*, *EVI1* and *GATA2*) of two gene families which appear to interact and play roles in similar pathways should be upregulated by a set of three translocations which correlate with a distinct phenotype in AML patients. The further determination of the roles that these genes play in haematopoiesis, and the extent to which they interact both directly and indirectly, will be of major interest in understanding the molecular consequences of these three translocations in AML.

It is important to note that *MEL1* and *EVI1* are likely to have oncogenic functions which are quite separate from their contribution to the clinical phenotype exhibited by patients with t(1;3)(p36;q21), t(3;3)(q21;q26) or inv(3)(q21q26). This is the case as not all translocations that involve *EVI1* result in this clinical phenotype, but the aberrant expression of *EVI1* still contributes to leukemogenesis (reviewed in Nucifora 1997). A similar circumstance was demonstrated for *MEL1* in the present study, with the finding that the ins(12;1)(p13;p36p21) rearrangement in Patient 2, who had a decreased platelet count and no apparent dysmegakaryopoiesis, resulted in the upregulation of *MEL1*. Transgenic studies, similar to those that have been performed for *EVI1* (Kurokawa *et al* 1995, Sitailo *et al* 1999) are required to demonstrate an oncogenic effect for *MEL1* overexpression. Whether the difference between the oncogenic function of these genes and their contribution to the clinical phenotype are the result of wholly separate functions, such as the activation of distinct sets of genes, or a result of the fact that *GATA2* is not also aberrantly expressed, will require further investigation. It remains possible also that *EVI1* and *MEL1* do not contribute to the clinical phenotype at all, and that this phenotype is purely a result of *GATA2* upregulation.

Characterisation of the t(1;3)(p36;p21) in Patient 3 led to the demonstration that a fusion gene involving the 3' end of *GOLGA4* was expressed. This fusion transcript includes the portion of the *GOLGA4* transcript which encodes the GRIP domain, a localisation signal that targets the protein to the Golgi complex. Currently, it is difficult to envision a leukemogenic function for such a fusion. The reciprocal transcript would potentially be a more biologically relevant fusion because it would encode a large coiled-coil domain fused to the 3' end of the partner gene. Coiled-coil

domains have been demonstrated to be essential for the leukemogenic function of both the *PML-RARα* and *AML1-ETO* fusion genes (Minucci *et al* 2000). The mechanism involves fusion of coiled-coil domains (from *PML* and *ETO*, respectively) to a transcription factor (*RARα* and *AML1*). The coiled-coil domains allow the fusion proteins to oligomerise, leading to the formation of high molecular weight complexes. This leads to recruitment of the nuclear corepressor/histone deacetylase (N-CoR/HDAC) complex, which contributes to leukemogenesis by blocking differentiation of haematopoietic cells. A similar mechanism has recently been proposed for the leukemogenic action of the *NUP98* fusion genes (Hussey and Dobrovic, 2002). The non-homeobox gene fusion partners of *NUP98* all contain a coiled-coil domain, and this is thought to similarly promote oligomerisation of the fusion genes. Although NUP98 is a nucleoporin protein and not a transcription factor, it contains a region of FG repeats which are known to have transcriptional activity following recruitment of CBP/p300 (Kasper *et al* 1999). Therefore, if the reciprocal fusion product of the t(1;3)(p36;p21) involving the 5' end of *GOLGA4* is transcribed, this is a likely mechanism by which it may have leukemogenic activity. It will be necessary to first determine whether this reciprocal fusion is transcribed in patients with this translocation, and then to determine whether the fusion partner gene has transactivating potential to further this hypothesis.

6.4 Summary of Major Findings

1) The breakpoints of a t(1;3)(p36;q21) were mapped and found to be consistent with those previously reported (Mochizuki *et al* 2000).

2) *MEL1* was confirmed to be upregulated as an outcome of this translocation in two patients, also in agreement with previous findings (Mochizuki *et al* 2000).

3) *MEL1* was found to be expressed in a subset of AML and MDS patients without t(1;3)(p36;q21). *MEL1* was also found to be expressed at low levels in normal bone marrow, in contrast to previous findings and possibly suggesting a role in haematopoiesis.

4) *GATA2* expression was observed in the t(1;3)(p36;q21) in two patients, the t(3;3)(q21;q26) in one patient and the inv(3)(q21q26) in two patients, suggesting a potential role for expression of *GATA2* in promoting the elevated platelet count and dysmegakaryopoiesis observed in patients with these translocations.

5) The 1p36 and 12p13 breakpoints of a ins(12;1)(p13;p36p21) were mapped. The 1p36 breakpoint clustered with the t(1;3)(p36;q21) 1p36 breakpoints, and the 12p13 breakpoint was mapped to intron 1 of *TEL*.

6) *MEL1* was found to be expressed as a result of the ins(12;1)(p13;p36p21) . No *TEL* fusion was observed arising from this translocation. This supports a position effect upregulation mechanism for 12p13 translocations that do not result in *TEL* fusion (Cools *et al* 2002).

7) The 3p21 breakpoint of a t(1;3)(p36;p21) was mapped to intron 20 of the *GOLGA4* gene. *GOLGA4* had not been previously implicated in leukaemia.

8) An aberrant band was detected in a t(1;3)(p36;p21) patient RNA sample using a *GOLGA4* probe in Northern analysis, strongly suggesting the formation of a fusion transcript.

REFERENCES

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J.D. (1994). *Molecular Biology of the Cell.* Garland Science Publishing. New York.

Alimena, G., De Cuia, M. R., Mecucci, C., Arcese, W., Mauro, F., Screnci, M., Mancini, M., Cedrone, M., Nanni, M., Montefusco, E., *et al.* (1990). Cytogenetic follow-up after allogeneic bone-marrow transplantation for Ph1-positive chronic myelogenous leukemia. *Bone Marrow Transplant,* **5(2):** 119-127.

Andersen, M. K., Christiansen, D. H., Jensen, B. A., Ernst, P., Hauge, G. and Pedersen-Bjergaard, J. (2001). Therapy-related acute lymphoblastic leukaemia with MLL rearrangements following DNA topoisomerase II inhibitors, an increasing problem: report on two new cases and review of the literature since 1992. *Br J Haematol,* **114(3):** 539-543.

Andersen, M. K., Larson, R. A., Mauritzson, N., Schnittger, S., Jhanwar, S. C. and Pedersen-Bjergaard, J. (2002). Balanced chromosome abnormalities inv(16) and t(15;17) in therapy-related myelodysplastic syndromes and acute leukemia: report from an international workshop. *Genes Chromosomes Cancer,* **33(4):** 395-400.

Aplan, P. D., Chervinsky, D. S., Stanulla, M. and Burhans, W. C. (1996). Site-specific DNA cleavage within the MLL breakpoint cluster region induced by topoisomerase II inhibitors. *Blood,* **87(7):** 2649-2658.

Armstrong, S. A., Staunton, J. E., Silverman, L. B., Pieters, R., Den Boer, M. L., Minden, M. D., Sallan, S. E., Lander, E. S., Golub, T. R. and Korsmeyer, S. J. (2002). MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nat Genet,* **30(1):** 41-47.

Ayton, P. M. and Cleary, M. L. (2001). Molecular mechanisms of leukemogenesis mediated by MLL fusion proteins. *Oncogene,* **20(40):** 5695-5707.

Baens, M., Wlodarska, I., Corveleyn, A., Hoornaert, I., Hagemeijer, A. and Marynen, P. (1999). A physical, transcript, and deletion map of chromosome region 12p12.3 flanked by ETV6 and CDKN1B: hypermethylation of the LRP6 CpG island in two leukemia patients with hemizygous del(12p). *Genomics,* **56(1):** 40-50.

Barr, F. G. (1998). Translocations, cancer and the puzzle of specificity. *Nat Genet,* **19(2):** 121-124.

Bartholomew, C., Kilbey, A., Clark, A. M. and Walker, M. (1997). The Evi-1 proto-oncogene encodes a transcriptional repressor activity associated with transformation. *Oncogene,* **14(5):** 569-577.

Bartram, C. R., De Klein, A., Hagemeijer, A., Van Agthoven, T., Geurts Van Kessel, A., Bootsma, D., Grosveld, G., Ferguson-Smith, M. A., Davies, T., Stone, M., *et al.* (1983). Translocation of c-ab1 oncogene correlates with the presence of a Philadelphia chromosome in chronic myelocytic leukaemia. *Nature,* **306(5940):** 277-280.

Bassing, C. H., Swat, W. and Alt, F. W. (2002). The mechanism and regulation of chromosomal V(D)J recombination. *Cell,* **109(Suppl):** S45-55.

Bennett, J. M., Catovsky, D., Daniel, M. T., Flandrin, G., Galton, D. A., Gralnick, H. R. and Sultan, C. (1976). Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group. *Br J Haematol,* **33(4):** 451-458.

Bennett, J. M., Catovsky, D., Daniel, M. T., Flandrin, G., Galton, D. A., Gralnick, H. R. and Sultan, C. (1985a). Proposed revised criteria for the classification of acute myeloid leukemia. A report of the French-American-British Cooperative Group. *Ann Intern Med,* **103(4):** 620-625.

Bennett, J. M., Catovsky, D., Daniel, M. T., Flandrin, G., Galton, D. A., Gralnick, H. R. and Sultan, C. (1985b). Criteria for the diagnosis of acute leukemia of megakaryocyte lineage (M7). A report of the French-American-British Cooperative Group. *Ann Intern Med,* **103(3):** 460-462.

Bennett, J. M., Catovsky, D., Daniel, M. T., Flandrin, G., Galton, D. A., Gralnick, H. R. and Sultan, C. (1991). Proposal for the recognition of minimally differentiated acute myeloid leukaemia (AML-MO). *Br J Haematol,* **78(3):** 325-329.

Bennett, J. M. (2000). World Health Organization classification of the acute leukemias and myelodysplastic syndrome. *Int J Hematol,* **72(2):** 131-133.

Bessho, F., Mizutani, S., Hayashi, Y., Moriwaki, K., Yokota, S. and Inaba, T. (1989). Chronic myelomonocytic leukemia with chromosomal changes involving 1p36 and hepatocellular carcinoma in a case of Fanconi's anemia. *Eur J Haematol,* **42(5):** 492-495.

Beverloo, H. B., Panagopoulos, I., Isaksson, M., Van Wering, E., Van Drunen, E., De Klein, A., Johansson, B. and Slater, R. (2001). Fusion of the homeobox gene HLXB9 and the ETV6 gene in infant acute myeloid leukemias with the t(7;12)(q36;p13). *Cancer Res,* **61(14):** 5374-5377.

Bitter, M. A., Neilly, M. E., Le Beau, M. M., Pearson, M. G. and Rowley, J. D. (1985). Rearrangements of chromosome 3 involving bands 3q21 and 3q26 are associated with normal or elevated platelet counts in acute nonlymphocytic leukemia. *Blood,* **66(6):** 1362-1370.

Bloomfield, C. D., Garson, O. M., Volin, L., Knuutila, S. and De La Chapelle, A. (1985). t(1;3)(p36;q21) in acute nonlymphocytic leukemia: a new cytogenetic-clinicopathologic association. *Blood,* **66(6):** 1409-1413.

Bloomfield, C. D., Archer, K. J., Mrozek, K., Lillington, D. M., Kaneko, Y., Head, D. R., Dal Cin, P. and Raimondi, S. C. (2002). 11q23 balanced chromosome aberrations in treatment-related myelodysplastic syndromes and acute leukemia: report from an international workshop. *Genes Chromosomes Cancer,* **33(4):** 362-378.

Boettner, B. and Van Aelst, L. (2002). The role of Rho GTPases in disease development. *Gene,* **286(2):** 155-174.

Bohlander, S. K. (2000). Fusion genes in leukemia: an emerging network. *Cytogenet Cell Genet,* **91(1-4):** 52-56.

Booden, M. A., Siderovski, D. P. and Der, C. J. (2002). Leukemia-associated Rho guanine nucleotide exchange factor promotes G alpha q-coupled activation of RhoA. *Mol Cell Biol,* **22(12):** 4053-4061.

Boultwood, J. (2001). Ataxia telangiectasia gene mutations in leukaemia and lymphoma. *J Clin Pathol,* **54(7):** 512-516.

Boxer, L. M. and Dang, C. V. (2001). Translocations involving c-myc and c-myc function. *Oncogene,* **20(40):** 5595-5610.

Brewer, A. C., Guille, M. J., Fear, D. J., Partington, G. A. and Patient, R. K. (1995). Nuclear translocation of a maternal CCAAT factor at the start of gastrulation activates Xenopus GATA-2 transcription. *Embo J,* **14(4):** 757-766.

Brodeur, G. M., Green, A. A., Hayes, F. A., Williams, K. J., Williams, D. L. and Tsiatis, A. A. (1981). Cytogenetic features of human neuroblastomas and cell lines. *Cancer Res,* **41(11 Pt 1):** 4678-4686.

Buijs, A., Sherr, S., Van Baal, S., Van Bezouw, S., Van Der Plas, D., Geurts Van Kessel, A., Riegman, P., Lekanne Deprez, R., Zwarthoff, E., Hagemeijer, A., *et al.* (1995). Translocation (12;22) (p13;q11) in myeloproliferative disorders results in fusion of the ETS-like TEL gene on 12p13 to the MN1 gene on 22q11. *Oncogene,* **10(8):** 1511-1519.

Buijs, A., Van Rompaey, L., Molijn, A. C., Davis, J. N., Vertegaal, A. C., Potter, M. D., Adams, C., Van Baal, S., Zwarthoff, E. C., Roussel, M. F. and Grosveld, G. C. (2000). The MN1-TEL fusion protein, encoded by the translocation (12;22)(p13;q11) in myeloid leukemia, is a transcription factor with transforming activity. *Mol Cell Biol,* **20(24):** 9281-9293.

Cambrin, G. R., Mecucci, C., Van Orshoven, A., Tricot, G. and Van Den Berghe, H. (1986). Translocation t(1;3)(p36;q21) in malignant myeloid stem cell disorders. *Cancer Genet Cytogenet,* **22(1):** 75-81.

Cantor, A. B. and Orkin, S. H. (2001). Hematopoietic development: a balancing act. *Curr Opin Genet Dev,* **11(5):** 513-519.

Cazzaniga, G., Tosi, S., Aloisi, A., Giudici, G., Daniotti, M., Pioltelli, P., Kearney, L. and Biondi, A. (1999). The tyrosine kinase abl-related gene ARG is fused to ETV6 in an AML-M4Eo patient with a t(1;12)(q25;p13): molecular cloning of both reciprocal transcripts. *Blood,* **94(12):** 4370-4373.

Cazzaniga, G., Daniotti, M., Tosi, S., Giudici, G., Aloisi, A., Pogliani, E., Kearney, L. and Biondi, A. (2001). The paired box domain gene PAX5 is fused to ETV6/TEL in an acute lymphoblastic leukemia case. *Cancer Res,* **61(12):** 4666-4670.

Chakrabarti, S. R. and Nucifora, G. (1999). The leukemia-associated gene TEL encodes a transcription repressor which associates with SMRT and mSin3A. *Biochem Biophys Res Commun,* **264(3):** 871-877.

Chase, A., Reiter, A., Burci, L., Cazzaniga, G., Biondi, A., Pickard, J., Roberts, I. A., Goldman, J. M. and Cross, N. C. (1999). Fusion of ETV6 to the caudal-related homeobox gene CDX2 in acute myeloid leukemia with the t(12;13)(p13;q12). *Blood,* **93(3):** 1025-1031.

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(3):** 1161-1167.

Chen, D. and Zhang, G. (2001). Enforced expression of the GATA-3 transcription factor affects cell fate decisions in hematopoiesis. *Exp Hematol,* **29(8):** 971-980.

Cheng, T., Shen, H., Giokas, D., Gere, J., Tenen, D. G. and Scadden, D. T. (1996). Temporal mapping of gene expression levels during the differentiation of individual primary hematopoietic cells. *Proc Natl Acad Sci U S A,* **93(23):** 13158-13163.

Conrad, M. and Topal, M. D. (1989). DNA and spermidine provide a switch mechanism to regulate the activity of restriction enzyme Nae I. *Proc Natl Acad Sci U S A,* **86(24):** 9707-9711.

Cools, J., Bilhou-Nabera, C., Wlodarska, I., Cabrol, C., Talmant, P., Bernard, P., Hagemeijer, A. and Marynen, P. (1999). Fusion of a novel gene, BTL, to ETV6 in acute myeloid leukemias with a t(4;12)(q11-q12;p13). *Blood,* **94(5):** 1820-1824.

Cools, J., Mentens, N., Odero, M. D., Peeters, P., Wlodarska, I., Delforge, M., Hagemeijer, A. and Marynen, P. (2002). Evidence for position effects as a variant ETV6-mediated leukemogenic mechanism in myeloid leukemias with a t(4;12)(q11-q12;p13) or t(5;12)(q31;p13). *Blood,* **99(5):** 1776-1784.

Cory, S., Graham, M., Webb, E., Corcoran, L. and Adams, J. M. (1985). Variant (6;15) translocations in murine plasmacytomas involve a chromosome 15 locus at least 72 kb from the c-myc oncogene. *Embo J,* **4(3):** 675-681.

Crans, H. N. and Sakamoto, K. M. (2001). Transcription factors and translocations in lymphoid and myeloid leukemia. *Leukemia,* **15(3):** 313-331.

Crossley, M., Merika, M. and Orkin, S. H. (1995). Self-association of the erythroid transcription factor GATA-1 mediated by its zinc finger domains. *Mol Cell Biol,* **15(5):** 2448-2456.

Dang, C. V. (1999). c-Myc target genes involved in cell growth, apoptosis, and metabolism. *Mol Cell Biol,* **19(1):** 1-11.

Dave, B. J., Hess, M. M., Pickering, D. L., Zaleski, D. H., Pfeifer, A. L., Weisenburger, D. D., Armitage, J. O. and Sanger, W. G. (1999). Rearrangements of chromosome band 1p36 in non-Hodgkin's lymphoma. *Clin Cancer Res,* **5(6):** 1401-1409.

De Boer, C. J., Van Krieken, J. H., Schuuring, E. and Kluin, P. M. (1997). Bcl-1/cyclin D1 in malignant lymphoma. *Ann Oncol,* **8(Suppl 2):** 109-117.

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(5894):** 765-767.

Del Peso, L., Hernandez-Alcoceba, R., Embade, N., Carnero, A., Esteve, P., Paje, C. and Lacal, J. C. (1997). Rho proteins induce metastatic properties in vivo. *Oncogene,* **15(25):** 3047-3057.

Deloukas, P., Schuler, G. D., Gyapay, G., Beasley, E. M., Soderlund, C., Rodriguez-Tome, P., Hui, L., Matise, T. C., Mckusick, K. B., Beckmann, J. S., Bentolila, S., Bihoreau, M., Birren, B. B., Browne, J., Butler, A., Castle, A. B., Chiannilkulchai, N., Clee, C., Day, P. J., Dehejia, A., Dibling, T., Drouot, N., Duprat, S., Fizames, C., Bentley, D. R., *et al.* (1998). A physical map of 30,000 human genes. *Science,* **282(5389):** 744-746.

Delwel, R., Funabiki, T., Kreider, B. L., Morishita, K. and Ihle, J. N. (1993). Four of the seven zinc fingers of the Evi-1 myeloid-transforming gene are required for sequence-specific binding to GA(C/T)AAGA(T/C)AAGATAA. *Mol Cell Biol,* **13(7):** 4291-4300.

Dewald, G. W., Schad, C. R., Christensen, E. R., Tiede, A. L., Zinsmeister, A. R., Spurbeck, J. L., Thibodeau, S. N. and Jalal, S. M. (1993). The application of fluorescent in situ hybridization to detect Mbcr/abl fusion in variant Ph chromosomes in CML and ALL. *Cancer Genet Cytogenet.* **71(1):** 7-14.

Dimartino, J. F. and Cleary, M. L. (1999). Mll rearrangements in haematological malignancies: lessons from clinical and biological studies. *Br J Haematol,* **106(3):** 614-626.

Dobrovic, A., Peters, G. B. and Ford, J. H. (1991). Molecular analysis of the Philadelphia chromosome. *Chromosoma,* **100(8):** 479-486.

Dobson, C. L., Warren, A. J., Pannell, R., Forster, A., Lavenir, I., Corral, J., Smith, A. J. and Rabbitts, T. H. (1999). The mll-AF9 gene fusion in mice controls myeloproliferation and specifies acute myeloid leukaemogenesis. *Embo J,* **18(13):** 3564-3574.

Dorfman, D. M., Wilson, D. B., Bruns, G. A. and Orkin, S. H. (1992). Human transcription factor GATA-2. Evidence for regulation of preproendothelin-1 gene expression in endothelial cells. *J Biol Chem,* **267(2):** 1279-1285.

Dube, I., Dixon, J., Beckett, T., Grossman, A., Weinstein, M., Benn, P., Mckeithan, T., Norman, C. and Pinkerton, P. (1989). Location of breakpoints within the major breakpoint cluster region (bcr) in 33 patients with bcr rearrangement-positive chronic myeloid leukemia (CML) with complex or absent Philadelphia chromosomes. *Genes Chromosomes Cancer,* **1(1):** 106-111.

Eguchi, M., Eguchi-Ishimae, M., Tojo, A., Morishita, K., Suzuki, K., Sato, Y., Kudoh, S., Tanaka, K., Setoyama, M., Nagamura, F., Asano, S. and Kamada, N. (1999). Fusion of ETV6 to neurotrophin-3 receptor TRKC in acute myeloid leukemia with t(12;15)(p13;q25). *Blood,* **93(4):** 1355-1363.

Elghetany, M. T., Maccallum, J. M. and Davey, F. R. (1990). The use of cytochemical procedures in the diagnosis and management of acute and chronic myeloid leukemia. *Clin Lab Med,* **10(4):** 707-720.

Erlich, R., Gleeson, P. A., Campbell, P., Dietzsch, E. and Toh, B. H. (1996). Molecular characterization of trans-Golgi p230. A human peripheral membrane protein encoded by a gene on chromosome 6p12-22 contains extensive coiled-coil alpha-helical domains and a granin motif. *J Biol Chem,* **271(14):** 8328-8337.

Ernst, P., Wang, J. and Korsmeyer, S. J. (2002). The role of MLL in hematopoiesis and leukemia. *Curr Opin Hematol,* **9(4):** 282-287.

Esteve, P., Embade, N., Perona, R., Jimenez, B., Del Peso, L., Leon, J., Arends, M., Miki, T. and Lacal, J. C. (1998). Rho-regulated signals induce apoptosis in vitro and in vivo by a p53-independent, but Bcl2 dependent pathway. *Oncogene,* **17(14):** 1855-1869.

Evans, T., Reitman, M. and Felsenfeld, G. (1988). An erythrocyte-specific DNA-binding factor recognizes a regulatory sequence common to all chicken globin genes. *Proc Natl Acad Sci U S A,* **85(16):** 5976-5980.

Falini, B. and Mason, D. Y. (2002). Proteins encoded by genes involved in chromosomal alterations in lymphoma and leukemia: clinical value of their detection by immunocytochemistry. *Blood.* **99(2):** 409-26.

Fears, S., Mathieu, C., Zeleznik-Le, N., Huang, S., Rowley, J. D. and Nucifora, G. (1996). Intergenic splicing of MDS1 and EVI1 occurs in normal tissues as well as in myeloid leukemia and produces a new member of the PR domain family. *Proc Natl Acad Sci U S A,* **93(4):** 1642-1647.

Fears, S., Gavin, M., Zhang, D. E., Hetherington, C., Ben-David, Y., Rowley, J. D. and Nucifora, G. (1997). Functional characterization of ETV6 and ETV6/CBFA2 in the regulation of the MCSFR proximal promoter. *Proc Natl Acad Sci U S A,* **94(5):** 1949-1954.

Fenrick, R., Amann, J. M., Lutterbach, B., Wang, L., Westendorf, J. J., Downing, J. R. and Hiebert, S. W. (1999). Both TEL and AML-1 contribute repression domains to the t(12;21) fusion protein. *Mol Cell Biol,* **19(10):** 6566-6574.

Fonatsch, C., Gudat, H., Lengfelder, E., Wandt, H., Silling-Engelhardt, G., Ludwig, W. D., Thiel, E., Freund, M., Bodenstein, H., Schwieder, G., *et al.* (1994). Correlation of cytogenetic findings with clinical features in 18 patients with inv(3)(q21q26) or t(3;3)(q21;q26). *Leukemia,* **8(8):** 1318-1326.

Fujiwara, Y., Browne, C. P., Cunniff, K., Goff, S. C. and Orkin, S. H. (1996). Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *Proc Natl Acad Sci U S A,* **93(22):** 12355-12358.

Goddard, A. D., Borrow, J., Freemont, P. S. and Solomon, E. (1991). Characterization of a zinc finger gene disrupted by the t(15;17) in acute promyelocytic leukemia. *Science,* **254(5036):** 1371-1374.

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(2):** 307-316.

Golub, T. R., Barker, G. F., Bohlander, S. K., Hiebert, S. W., Ward, D. C., Bray-Ward, P., Morgan, E., Raimondi, S. C., Rowley, J. D. and Gilliland, D. G. (1995). Fusion of the TEL gene on 12p13 to the AML1 gene on 21q22 in acute lymphoblastic leukemia. *Proc Natl Acad Sci U S A,* **92(11):** 4917-4921.

Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D. and Lander, E. S. (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science,* **286(5439):** 531-537.

Graham, M. and Adams, J. M. (1986). Chromosome 8 breakpoint far 3' of the c-myc oncogene in a Burkitt's lymphoma 2;8 variant translocation is equivalent to the murine pvt-1 locus. *Embo J,* **5(11):** 2845-2851.

Griffiths, A.J.F., Miller, J.H., Suzuki, D., Lewontin, R.C. and Gelbart, W.M. (2000). *An Introduction to Genetic Analysis*. W. H. Freeman. New York.

Grigg, A. P., Gascoyne, R. D., Phillips, G. L. and Horsman, D. E. (1993). Clinical, haematological and cytogenetic features in 24 patients with structural rearrangements of the Q arm of chromosome 3. *Br J Haematol,* **83(1):** 158-165.

Harata, M., Nishimori, K. and Hatta, S. (2001). Identification of two cDNAs for human actin-related proteins (Arps) that have remarkable similarity to conventional actin. *Biochim Biophys Acta,* **1522(2):** 130-133.

He, L., Yu, J. X., Liu, L., Buyse, I. M., Wang, M. S., Yang, Q. C., Nakagawara, A., Brodeur, G. M., Shi, Y. E. and Huang, S. (1998). RIZ1, but not the alternative RIZ2 product of the same gene, is underexpressed in breast cancer, and forced RIZ1 expression causes G2-M cell cycle arrest and/or apoptosis. *Cancer Res,* **58(19):** 4238-4244.

Hecht, J. L. and Aster, J. C. (2000). Molecular biology of Burkitt's lymphoma. *J Clin Oncol.* **18(21):** 3707-21.

Heisterkamp, N., Stephenson, J. R., Groffen, J., Hansen, P. F., De Klein, A., Bartram, C. R. and Grosveld, G. (1983). Localization of the c-ab1 oncogene adjacent to a translocation break point in chronic myelocytic leukaemia. *Nature,* **306(5940):** 239-242.

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(6263):** 251-253.

Hiebert, S. W., Sun, W., Davis, J. N., Golub, T., Shurtleff, S., Buijs, A., Downing, J. R., Grosveld, G., Roussell, M. F., Gilliland, D. G., Lenny, N. and Meyers, S. (1996). The t(12;21) translocation converts AML-1B from an activator to a repressor of transcription. *Mol Cell Biol,* **16(4):** 1349-1355.

Himmel, K. L., Bi, F., Shen, H., Jenkins, N. A., Copeland, N. G., Zheng, Y. and Largaespada, D. A. (2002). Activation of clg, a novel dbl family guanine nucleotide exchange factor gene, by proviral insertion at evi24, a common integration site in B cell and myeloid leukemias. *J Biol Chem,* **277(16):** 13463-13472.

Hoglund, M., Johansson, B., Pedersen-Bjergaard, J., Marynen, P. and Mitelman, F. (1996). Molecular characterization of 12p abnormalities in hematologic malignancies: deletion of KIP1, rearrangement of TEL, and amplification of CCND2. *Blood,* **87(1):** 324-330.

Huang, S., Shao, G. and Liu, L. (1998). The PR domain of the Rb-binding zinc finger protein RIZ1 is a protein binding interface and is related to the SET domain functioning in chromatin-mediated gene expression. *J Biol Chem,* **273(26):** 15933-15939.

Hussey, D. J. and Dobrovic, A. (2002). Recurrent coiled-coil motifs in NUP98 fusion partners provide a clue to leukemogenesis. *Blood,* **99(3):** 1097-1098.

Ikonomi, P., Rivera, C. E., Riordan, M., Washington, G., Schechter, A. N. and Noguchi, C. T. (2000). Overexpression of GATA-2 inhibits erythroid and promotes megakaryocyte differentiation. *Exp Hematol,* **28(12):** 1423-1431.

Izutsu, K., Kurokawa, M., Imai, Y., Maki, K., Mitani, K. and Hirai, H. (2001). The corepressor CtBP interacts with Evi-1 to repress transforming growth factor beta signaling. *Blood,* **97(9):** 2815-2822.

Jaffe, A. B. and Hall, A. (2002). Rho GTPases in transformation and metastasis. *Adv Cancer Res,* **84**(57-80.

Jiang, G., Liu, L., Buyse, I. M., Simon, D. and Huang, S. (1999). Decreased RIZ1 expression but not RIZ2 in hepatoma and suppression of hepatoma tumorigenicity by RIZ1. *Int J Cancer,* **83(4):** 541-546.

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(1):** 764-776.

Kearney, L. (1999). The impact of the new fish technologies on the cytogenetics of haematological malignancies. *Br J Haematol,* **104(4):** 648-658.

Kjer-Nielsen, L., Van Vliet, C., Erlich, R., Toh, B. H. and Gleeson, P. A. (1999). The Golgi-targeting sequence of the peripheral membrane protein p230. *J Cell Sci,* **112(Pt 11):** 1645-1654.

Klugbauer, S., Demidchik, E. P., Lengfelder, E. and Rabes, H. M. (1998). Detection of a novel type of RET rearrangement (PTC5) in thyroid carcinomas after Chernobyl and analysis of the involved RET-fused gene RFG5. *Cancer Res,* **58(2):** 198-203.

Knuutila, S., Aalto, Y., Autio, K., Bjorkqvist, A. M., El-Rifai, W., Hemmer, S., Huhta, T., Kettunen, E., Kiuru-Kuhlefelt, S., Larramendy, M. L., Lushnikova, T., Monni, O., Pere, H., Tapper, J., Tarkkanen, M., Varis, A., Wasenius, V. M., Wolf, M. and Zhu, Y. (1999). DNA copy number losses in human neoplasms. *Am J Pathol,* **155(3):** 683-694.

Ko, L. J. and Engel, J. D. (1993). DNA-binding specificities of the GATA transcription factor family. *Mol Cell Biol,* **13(7):** 4011-4022.

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(5):** 2145-2150.

Kouzarides, T. (2002). Histone methylation in transcriptional control. *Curr Opin Genet Dev,* **12(2):** 198-209.

Kruger, D. H., Barcak, G. J., Reuter, M. and Smith, H. O. (1988). EcoRII can be activated to cleave refractory DNA recognition sites. *Nucleic Acids Res,* **16(9):** 3997-4008.

Kurokawa, M., Ogawa, S., Tanaka, T., Mitani, K., Yazaki, Y., Witte, O. N. and Hirai, H. (1995). The AML1/Evi-1 fusion protein in the t(3;21) translocation exhibits transforming activity on Rat1 fibroblasts with dependence on the Evi-1 sequence. *Oncogene,* **11(5):** 833-840.

Labbaye, C., Quaranta, M. T., Pagliuca, A., Militi, S., Licht, J. D., Testa, U. and Peschle, C. (2002). PLZF induces megakaryocytic development, activates Tpo receptor expression and interacts with GATA1 protein. *Oncogene,* **21(43):** 6669-6679.

Lampert, F., Rudolph, B., Christiansen, H. and Franke, F. (1988). Identical chromosome 1p breakpoint abnormality in both the tumor and the constitutional karyotype of a patient with neuroblastoma. *Cancer Genet Cytogenet,* **34(2):** 235-239.

Langabeer, S. E., Rogers, J. R., Harrison, G., Wheatley, K., Walker, H., Bain, B. J., Burnett, A. K., Goldstone, A. H., Linch, D. C. and Grimwade, D. (2001). EVI1 expression in acute myeloid leukaemia. *Br J Haematol,* **112(1):** 208-211.

Le Beau, M. M., Westbrook, C. A., Diaz, M. O. and Rowley, J. D. (1985). c-src is consistently conserved in the chromosomal deletion (20q) observed in myeloid disorders. *Proc Natl Acad Sci U S A,* **82(19):** 6692-6696.

Liu, L., Shao, G., Steele-Perkins, G. and Huang, S. (1997). The retinoblastoma interacting zinc finger gene RIZ produces a PR domain-lacking product through an internal promoter. *J Biol Chem,* **272(5):** 2984-2991.

Look, A. T. (1997). Oncogenic transcription factors in the human acute leukemias. *Science,* **278(5340):** 1059-1064.

Lopez, R. G., Carron, C., Oury, C., Gardellin, P., Bernard, O. and Ghysdael, J. (1999). TEL is a sequence-specific transcriptional repressor. *J Biol Chem,* **274(42):** 30132-30138.

Ma, G. T., Roth, M. E., Groskopf, J. C., Tsai, F. Y., Orkin, S. H., Grosveld, F., Engel, J. D. and Linzer, D. I. (1997). GATA-2 and GATA-3 regulate trophoblast-specific gene expression in vivo. *Development,* **124(4):** 907-914.

Marculescu, R., Le, T., Simon, P., Jaeger, U. and Nadel, B. (2002). V(D)J-mediated translocations in lymphoid neoplasms: a functional assessment of genomic instability by cryptic sites. *J Exp Med,* **195(1):** 85-98.

Marlton, P., Claxton, D. F., Liu, P., Estey, E. H., Beran, M., Lebeau, M., Testa, J. R., Collins, F. S., Rowley, J. D. and Siciliano, M. J. (1995). Molecular characterization of 16p deletions associated with inversion 16 defines the critical fusion for leukemogenesis. *Blood,* **85(3):** 772-779.

Marsden, K. A., Pearse, A. M., Collins, G. G., Ford, D. S., Heard, S. and Kimber, R. I. (1992). Acute leukemia with t(1;3)(p36;q21), evolution to t(1;3)(p36;q21), t(14;17)(q32;q21), and loss of red cell A and Le(b) antigens. *Cancer Genet Cytogenet,* **64(1):** 80-85.

Matsumura, I. and Kanakura, Y. (2002). Molecular control of megakaryopoiesis and thrombopoiesis. *Int J Hematol,* **75(5):** 473-483.

McPherson, J. D., Marra, M., Hillier, L., Waterston, R. H., Chinwalla, A., Wallis, J., Sekhon, M., Wylie, K., Mardis, E. R., Wilson, R. K., *et al.* (2001). A physical map of the human genome. *Nature,* **409(6822):** 934-941.

Miller, S. A., Dykes, D. D. and Polesky, H. F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res,* **16(3):** 1215.

Minegishi, N., Morita, S., Minegishi, M., Tsuchiya, S., Konno, T., Hayashi, N. and Yamamoto, M. (1997). Expression of GATA transcription factors in myelogenous and lymphoblastic leukemia cells. *Int J Hematol,* **65(3):** 239-249.

Minucci, S., Maccarana, M., Cioce, M., De Luca, P., Gelmetti, V., Segalla, S., Di Croce, L., Giavara, S., Matteucci, C., Gobbi, A., Bianchini, A., Colombo, E., Schiavoni, I., Badaracco, G., Hu, X., Lazar, M. A., Landsberger, N., Nervi, C. and Pelicci, P. G. (2000). Oligomerization of RAR and AML1 transcription factors as a novel mechanism of oncogenic activation. *Mol Cell,* **5(5):** 811-820.

Mitelman, F. (2000). Recurrent chromosome aberrations in cancer. *Mutat Res,* **462(2-3):** 247-253.

Mochizuki, N., Shimizu, S., Nagasawa, T., Tanaka, H., Taniwaki, M., Yokota, J. and Morishita, K. (2000). A novel gene, MEL1, mapped to 1p36.3 is highly homologous to the MDS1/EVI1 gene and is transcriptionally activated in t(1;3)(p36;q21)-positive leukemia cells. *Blood,* **96(9):** 3209-3214.

Moir, D. J., Jones, P. A., Pearson, J., Duncan, J. R., Cook, P. and Buckle, V. J. (1984). A new translocation, t(1;3) (p36;q21), in myelodysplastic disorders. *Blood,* **64(2):** 553-555.

Molkentin, J. D. (2000). The zinc finger-containing transcription factors GATA-4, -5, and -6. Ubiquitously expressed regulators of tissue-specific gene expression. *J Biol Chem,* **275(50):** 38949-38952.

Mori, N., Morosetti, R., Spira, S., Lee, S., Ben-Yehuda, D., Schiller, G., Landolfi, R., Mizoguchi, H. and Koeffler, H. P. (1998). Chromosome band 1p36 contains a putative tumor suppressor gene important in the evolution of chronic myelocytic leukemia. *Blood,* **92(9):** 3405-3409.

Morishita, K., Parker, D. S., Mucenski, M. L., Jenkins, N. A., Copeland, N. G. and Ihle, J. N. (1988). Retroviral activation of a novel gene encoding a zinc finger protein in IL-3-dependent myeloid leukemia cell lines. *Cell,* **54(6):** 831-840.

Morishita, K., Parganas, E., Parham, D. M., Matsugi, T. and Ihle, J. N. (1990). The Evi-1 zinc finger myeloid transforming gene is normally expressed in the kidney and in developing oocytes. *Oncogene,* **5(9):** 1419-1423.

Morishita, K., Parganas, E., William, C. L., Whittaker, M. H., Drabkin, H., Oval, J., Taetle, R., Valentine, M. B. and Ihle, J. N. (1992). Activation of EVI1 gene expression in human acute myelogenous leukemias by translocations spanning 300-400 kilobases on chromosome band 3q26. *Proc Natl Acad Sci U S A,* **89(9):** 3937-3941.

Morishita, K., Suzukawa, K., Taki, T., Ihle, J. N. and Yokota, J. (1995). EVI-1 zinc finger protein works as a transcriptional activator via binding to a consensus sequence of GACAAGATAAGATAAN1-28 CTCATCTTC. *Oncogene,* **10(10):** 1961-1967.

Mucenski, M. L., Taylor, B. A., Ihle, J. N., Hartley, J. W., Morse, H. C., 3rd, Jenkins, N. A. and Copeland, N. G. (1988). Identification of a common ecotropic viral integration site, Evi-1, in the DNA of AKXD murine myeloid tumors. *Mol Cell Biol,* **8(1):** 301-308.

Munro, S. and Nichols, B. J. (1999). The GRIP domain - a novel Golgi-targeting domain found in several coiled-coil proteins. *Curr Biol,* **9(7):** 377-380.

Nacheva, E. P., Gribble, S., Andrews, K., Wienberg, J. and Grace, C. D. (2000). Screening for specific chromosome involvement in hematological malignancies using a set of seven chromosome painting probes. An alternative approach for chromosome analysis using standard FISH instrumentation. *Cancer Genet Cytogenet,* **122(2):** 65-72.

Nagai, T., Harigae, H., Ishihara, H., Motohashi, H., Minegishi, N., Tsuchiya, S., Hayashi, N., Gu, L., Andres, B., Engel, J. D., *et al.* (1994). Transcription factor GATA-2 is expressed in erythroid, early myeloid, and CD34+ human leukemia-derived cell lines. *Blood,* **84(4):** 1074-1084.

Najfeld, V., Coyle, T. and Berk, P. D. (1988). Transformation of polycythemia vera to acute nonlymphocytic leukemia accompanied by t(1;3)(p36;q21) karyotype. *Cancer Genet Cytogenet,* **33(2):** 193-200.

Nakauchi, H., Takano, H., Ema, H. and Osawa, M. (1999). Further characterization of CD34-low/negative mouse hematopoietic stem cells. *Ann N Y Acad Sci.* **872**:57-66.

Nucifora, G., Begy, C. R., Kobayashi, H., Roulston, D., Claxton, D., Pedersen-Bjergaard, J., Parganas, E., Ihle, J. N. and Rowley, J. D. (1994). Consistent intergenic splicing and production of multiple transcripts between AML1 at 21q22 and unrelated genes at 3q26 in (3;21)(q26;q22) translocations. *Proc Natl Acad Sci U S A,* **91(9):** 4004-4008.

Nucifora, G. (1997). The EVI1 gene in myeloid leukemia. *Leukemia,* **11(12):** 2022-2031.

Odero, M. D., Carlson, K., Calasanz, M. J., Lahortiga, I., Chinwalla, V. and Rowley, J. D. (2001). Identification of new translocations involving ETV6 in hematologic malignancies by fluorescence in situ hybridization and spectral karyotyping. *Genes Chromosomes Cancer,* **31(2):** 134-142.

Odero, M. D., Vizmanos, J. L., Roman, J. P., Lahortiga, I., Panizo, C., Calasanz, M. J., Zeleznik-Le, N. J., Rowley, J. D. and Novo, F. J. (2002). A novel gene, MDS2, is fused to ETV6/TEL in a t(1;12)(p36.1;p13) in a patient with myelodysplastic syndrome. *Genes Chromosomes Cancer,* **35(1):** 11-19.

Okuda, T., Nishimura, M., Nakao, M. and Fujita, Y. (2001). RUNX1/AML1: a central player in hematopoiesis. *Int J Hematol,* **74(3):** 252-257.

Oller, A. R., Vanden Broek, W., Conrad, M. and Topal, M. D. (1991). Ability of DNA and spermidine to affect the activity of restriction endonucleases from several bacterial species. *Biochemistry,* **30(9):** 2543-2549.

Olney, H. J., Mitelman, F., Johansson, B., Mrozek, K., Berger, R. and Rowley, J. D. (2002). Unique balanced chromosome abnormalities in treatment-related myelodysplastic syndromes and acute myeloid leukemia: report from an international workshop. *Genes Chromosomes Cancer,* **33(4):** 413-423.

Orkin, S. H. (1992). GATA-binding transcription factors in hematopoietic cells. *Blood,* **80(3):** 575-581.

Orkin, S. H. (1998). Embryonic stem cells and transgenic mice in the study of hematopoiesis. *Int J Dev Biol,* **42(7 Spec No):** 927-934.

Overbeck, A. F., Brtva, T. R., Cox, A. D., Graham, S. M., Huff, S. Y., Khosravi-Far, R., Quilliam, L. A., Solski, P. A. and Der, C. J. (1995). Guanine nucleotide exchange factors: activators of Ras superfamily proteins. *Mol Reprod Dev,* **42(4):** 468-476.

Pan, J. Y. and Wessling-Resnick, M. (1998). GEF-mediated GDP/GTP exchange by monomeric GTPases: a regulatory role for Mg2+? *Bioessays,* **20(6):** 516-521.

Pandita, T. K. (2002). ATM function and telomere stability. *Oncogene,* **21(4):** 611-618.

Pandolfi, P. P. (2001). In vivo analysis of the molecular genetics of acute promyelocytic leukemia. *Oncogene,* **20(40):** 5726-5735.

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(1):** 34-38.

Peeters, P., Wlodarska, I., Baens, M., Criel, A., Selleslag, D., Hagemeijer, A., Van Den Berghe, H. and Marynen, P. (1997a). Fusion of ETV6 to MDS1/EVI1 as a result of t(3;12)(q26;p13) in myeloproliferative disorders. *Cancer Res,* **57(4):** 564-569.

Peeters, P., Raynaud, S. D., Cools, J., Wlodarska, I., Grosgeorge, J., Philip, P., Monpoux, F., Van Rompaey, L., Baens, M., Van Den Berghe, H. and Marynen, P. (1997b). Fusion of TEL, the ETS-variant gene 6 (ETV6), to the receptor-associated kinase JAK2 as a result of t(9;12) in a lymphoid and t(9;15;12) in a myeloid leukemia. *Blood,* **90(7):** 2535-2540.

Perkins, A. S. and Kim, J. H. (1996). Zinc fingers 1-7 of EVI1 fail to bind to the GATA motif by itself but require the core site GACAAGATA for binding. *J Biol Chem,* **271(2):** 1104-1110.

Persons, D. A., Allay, J. A., Allay, E. R., Ashmun, R. A., Orlic, D., Jane, S. M., Cunningham, J. M. and Nienhuis, A. W. (1999). Enforced expression of the GATA-2 transcription factor blocks normal hematopoiesis. *Blood,* **93(2):** 488-499.

Pevny, L., Simon, M. C., Robertson, E., Klein, W. H., Tsai, S. F., D'agati, V., Orkin, S. H. and Costantini, F. (1991). Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature,* **349(6306):** 257-260.

Pevny, L., Lin, C. S., D'agati, V., Simon, M. C., Orkin, S. H. and Costantini, F. (1995). Development of hematopoietic cells lacking transcription factor GATA-1. *Development,* **121(1):** 163-172.

Pintado, T., Ferro, M. T., San Roman, C., Mayayo, M. and Larana, J. G. (1985). Clinical correlations of the 3q21;q26 cytogenetic anomaly. A leukemic or myelodysplastic syndrome with preserved or increased platelet production and lack of response to cytotoxic drug therapy. *Cancer,* **55(3):** 535-541.

Rabes, H. M., Demidchik, E. P., Sidorow, J. D., Lengfelder, E., Beimfohr, C., Hoelzel, D. and Klugbauer, S. (2000). Pattern of radiation-induced RET and NTRK1 rearrangements in 191 post-chernobyl papillary thyroid carcinomas: biological, phenotypic, and clinical implications. *Clin Cancer Res,* **6(3):** 1093-1103.

Radich, J. P. (2002). The promise of gene expression analysis in hematopoetic malignancies. *Biochim Biophys Acta,* **1602(1):** 88-95.

Rea, S., Eisenhaber, F., O'carroll, D., Strahl, B. D., Sun, Z. W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C. P., Allis, C. D. and Jenuwein, T. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature,* **406(6796):** 593-599.

Rego, E. M. and Pandolfi, P. P. (2002). Reciprocal products of chromosomal translocations in human cancer pathogenesis: key players or innocent bystanders? *Trends Mol Med,* **8(8):** 396-405.

Richardson, C. and Jasin, M. (2000). Frequent chromosomal translocations induced by DNA double-strand breaks. *Nature,* **405(6787):** 697-700.

Richkind, K., Hromas, R., Lytle, C., Crenshaw, D., Velasco, J., Roherty, S., Srinivasiah, J. and Varella-Garcia, M. (2000). Identification of two new translocations that disrupt the AML1 gene. *Cancer Genet Cytogenet,* **122(2):** 141-143.

Ridley, A. J. (2001). Rho family proteins: coordinating cell responses. *Trends Cell Biol,* **11(12):** 471-477.

Ringertz, N., R. and Savage, R. E. (1976). *Cell Hybrids.* Academic Press. New York.

Romana, S. P., Mauchauffe, M., Le Coniat, M., Chumakov, I., Le Paslier, D., Berger, R. and Bernard, O. A. (1995). The t(12;21) of acute lymphoblastic leukemia results in a tel-AML1 gene fusion. *Blood,* **85(12):** 3662-3670.

Rowley, J. D. (1973a). Identificaton of a translocation with quinacrine fluorescence in a patient with acute leukemia. *Ann Genet,* **16(2):** 109-112.

Rowley, J. D. (1973b). Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature,* **243(5405):** 290-293.

Rowley, J. D. (1999). The role of chromosome translocations in leukemogenesis. *Semin Hematol,* **36(4 Suppl 7):** 59-72.

Rubnitz, J. E., Downing, J. R. and Pui, C. H. (1999). Significance of the TEL-AML fusion gene in childhood AML. *Leukemia,* **13(9):** 1470-1471.

Russell, M., List, A., Greenberg, P., Woodward, S., Glinsmann, B., Parganas, E., Ihle, J. and Taetle, R. (1994). Expression of EVI1 in myelodysplastic syndromes and other hematologic malignancies without 3q26 translocations. *Blood,* **84(4):** 1243-1248.

Sait, S. N., Qadir, M. U., Conroy, J. M., Matsui, S., Nowak, N. J. and Baer, M. R. (2002). Double minute chromosomes in acute myeloid leukemia and myelodysplastic syndrome: identification of new amplification regions by fluorescence in situ hybridization and spectral karyotyping. *Genes Chromosomes Cancer,* **34(1):** 42-47.

Salomon-Nguyen, F., Della-Valle, V., Mauchauffe, M., Busson-Le Coniat, M., Ghysdael, J., Berger, R. and Bernard, O. A. (2000). The t(1;12)(q21;p13) translocation of human acute myeloblastic leukemia results in a TEL-ARNT fusion. *Proc Natl Acad Sci U S A,* **97(12):** 6757-6762.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbour Laboratory. New York.

Sato, Y., Bohlander, S. K., Kobayashi, H., Reshmi, S., Suto, Y., Davis, E. M., Espinosa, R., Hoopes, R., Montgomery, K. T., Kucherlapati, R. S., Le Beau, M. M. and Rowley, J. D. (1997). Heterogeneity in the breakpoints in balanced rearrangements involving band 12p13 in hematologic malignancies identified by fluorescence in situ hybridization: TEL (ETV6) is involved in only one half. *Blood,* **90(12):** 4886-4893.

Sato, Y., Izumi, T., Kanamori, H., Davis, E. M., Miura, Y., Larson, R. A., Le Beau, M. M., Ozawa, K. and Rowley, J. D. (2002). t(1;3)(p36;p21) is a recurring therapy-related translocation. *Genes Chromosomes Cancer,* **34(2):** 186-192.

Sawyers, C. L., Denny, C. T. and Witte, O. N. (1991). Leukemia and the disruption of normal hematopoiesis. *Cell,* **64(2):** 337-350.

Schoch, C., Kohlmann, A., Schnittger, S., Brors, B., Dugas, M., Mergenthaler, S., Kern, W., Hiddemann, W., Eils, R. and Haferlach, T. (2002). Acute myeloid leukemias with reciprocal rearrangements can be distinguished by specific gene expression profiles. *Proc Natl Acad Sci U S A,* **99(15):** 10008-10013.

Schwartz, J. G., Clare, N., Hansen, K., Britton, H. and Manhoff, L. (1986). Acute promyelocytic leukemia: report of a variant translocation, t(1;17). *Cancer Genet Cytogenet,* **20(1-2):** 89-93.

Secker-Walker, L. M., Mehta, A. and Bain, B. (1995). Abnormalities of 3q21 and 3q26 in myeloid malignancy: a United Kingdom Cancer Cytogenetic Group study. *Br J Haematol,* **91(2):** 490-501.

Shapiro, V. S., Lee, P. and Winoto, A. (1995). Identification and cloning of the G3B cDNA encoding a 3' segment of a protein binding to GATA-3. *Gene,* **163(2):** 329-330.

Sharan, S. K., Morimatsu, M., Albrecht, U., Lim, D. S., Regel, E., Dinh, C., Sands, A., Eichele, G., Hasty, P. and Bradley, A. (1997). Embryonic lethality and radiation hypersensitivity mediated by Rad51 in mice lacking Brca2. *Nature,* **386(6627):** 804-810.

Sharrocks, A. D. (2001). The ETS-domain transcription factor family. *Nat Rev Mol Cell Biol,* **2(11):** 827-837.

Shimizu, S., Suzukawa, K., Kodera, T., Nagasawa, T., Abe, T., Taniwaki, M., Yagasaki, F., Tanaka, H., Fujisawa, S., Johansson, B., Ahlgren, T., Yokota, J. and Morishita, K. (2000). Identification of breakpoint cluster regions at 1p36.3 and 3q21 in hematologic malignancies with t(1;3)(p36;q21). *Genes Chromosomes Cancer,* **27(3):** 229-238.

Shivdasani, R. A. (1997). Stem cell transcription factors. *Hematol Oncol Clin North Am,* **11(6):** 1199-1206.

Shivdasani, R. A. (2001). Molecular and transcriptional regulation of megakaryocyte differentiation. *Stem Cells,* **19(5):** 397-407.

Sitailo, S., Sood, R., Barton, K. and Nucifora, G. (1999). Forced expression of the leukemia-associated gene EVI1 in ES cells: a model for myeloid leukemia with 3q26 rearrangements. *Leukemia,* **13(11):** 1639-1645.

Slovak, M. L., Bedell, V., Popplewell, L., Arber, D. A., Schoch, C. and Slater, R. (2002). 21q22 balanced chromosome aberrations in therapy-related hematopoietic disorders: report from an international workshop. *Genes Chromosomes Cancer,* **33(4):** 379-394.

Soderholm, J., Kobayashi, H., Mathieu, C., Rowley, J. D. and Nucifora, G. (1997). The leukemia-associated gene MDS1/EVI1 is a new type of GATA-binding transactivator. *Leukemia,* **11(3):** 352-358.

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(2):** 166-175.

South Australian Cancer Registry. (2001) *Epidemiology of Cancer in South Australia. Incidence, Mortality and Survival, 1977-2000. Incidence and Mortality, 2000.* Openbook Publishers. Adelaide.

Speaks, S. L., Sanger, W. G., Masih, A. S., Harrington, D. S., Hess, M. and Armitage, J. O. (1992). Recurrent abnormalities of chromosome bands 10q23-25 in non-Hodgkin's lymphoma. *Genes Chromosomes Cancer,* **5(3):** 239-243.

Specchia, G., Cuneo, A., Liso, V., Contino, R., Pastore, D., Gentile, E., Rocchi, M. and Castoldi, G. L. (1999). A novel translocation t(1;7)(p36;q34) in three patients with acute myeloid leukaemia. *Br J Haematol,* **105(1):** 208-214.

Speck, N. A. (2001). Core binding factor and its role in normal hematopoietic development. *Curr Opin Hematol,* **8(4):** 192-196.

Stanulla, M., Wang, J., Chervinsky, D. S. and Aplan, P. D. (1997). Topoisomerase II inhibitors induce DNA double-strand breaks at a specific site within the AML1 locus. *Leukemia,* **11(4):** 490-496.

Stanulla, M., Chhalliyil, P., Wang, J., Jani-Sait, S. N. and Aplan, P. D. (2001). Mechanisms of MLL gene rearrangement: site-specific DNA cleavage within the breakpoint cluster region is independent of chromosomal context. *Hum Mol Genet,* **10(22):** 2481-2491.

Steele-Perkins, G., Fang, W., Yang, X. H., Van Gele, M., Carling, T., Gu, J., Buyse, I. M., Fletcher, J. A., Liu, J., Bronson, R., Chadwick, R. B., De La Chapelle, A., Zhang, X., Speleman, F. and Huang, S. (2001). Tumor formation and inactivation of RIZ1, an Rb-binding member of a nuclear protein-methyltransferase superfamily. *Genes Dev,* **15(17):** 2250-2262.

Stefanescu, D. T., Colita, D., Nicoara, S. and Calin, G. (1994). t(1;7)(p36;q32): a new recurring abnormality in primary myelodysplastic syndrome. *Cancer Genet Cytogenet,* **75(2):** 103-105.

Suto, Y., Sato, Y., Smith, S. D., Rowley, J. D. and Bohlander, S. K. (1997). A t(6;12)(q23;p13) results in the fusion of ETV6 to a novel gene, STL, in a B-cell ALL cell line. *Genes Chromosomes Cancer,* **18(4):** 254-268.

Suzukawa, K., Parganas, E., Gajjar, A., Abe, T., Takahashi, S., Tani, K., Asano, S., Asou, H., Kamada, N., Yokota, J., *et al.* (1994). Identification of a breakpoint cluster region 3' of the ribophorin I gene at 3q21 associated with the transcriptional activation of the EVI1 gene in acute myelogenous leukemias with inv(3)(q21q26). *Blood,* **84(8):** 2681-2688.

Takahashi, S., Shimizu, R., Suwabe, N., Kuroha, T., Yoh, K., Ohta, J., Nishimura, S., Lim, K. C., Engel, J. D. and Yamamoto, M. (2000). GATA factor transgenes under GATA-1 locus control rescue germline GATA-1 mutant deficiencies. *Blood,* **96(3):** 910-916.

Takahashi, S. and Licht, J. D. (2002). The human promyelocytic leukemia zinc finger gene is regulated by the Evi-1 oncoprotein and a novel guanine-rich site binding protein. *Leukemia,* **16(9):** 1755-1762.

Testoni, N., Borsaru, G., Martinelli, G., Carboni, C., Ruggeri, D., Ottaviani, E., Pelliconi, S., Ricci, P., Pastano, R., Visani, G., Zaccaria, A. and Tura, S. (1999). 3q21 and 3q26 cytogenetic abnormalities in acute myeloblastic leukemia: biological and clinical features. *Haematologica,* **84(8):** 690-694.

Ting, C. N., Olson, M. C., Barton, K. P. and Leiden, J. M. (1996). Transcription factor GATA-3 is required for development of the T-cell lineage. *Nature,* **384(6608):** 474-478.

Tsai, F. Y., Keller, G., Kuo, F. C., Weiss, M., Chen, J., Rosenblatt, M., Alt, F. W. and Orkin, S. H. (1994). An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature,* **371(6494):** 221-226.

Tsai, F. Y. and Orkin, S. H. (1997). Transcription factor GATA-2 is required for proliferation/survival of early hematopoietic cells and mast cell formation, but not for erythroid and myeloid terminal differentiation. *Blood,* **89(10):** 3636-3643.

Tsai, F. Y., Browne, C. P. and Orkin, S. H. (1998). Knock-in mutation of transcription factor GATA-3 into the GATA-1 locus: partial rescue of GATA-1 loss of function in erythroid cells. *Dev Biol,* **196(2):** 218-227.

Tsuzuki, S., Towatari, M., Saito, H. and Enver, T. (2000). Potentiation of GATA-2 activity through interactions with the promyelocytic leukemia protein (PML) and the t(15;17)-generated PML-retinoic acid receptor alpha oncoprotein. *Mol Cell Biol,* **20(17):** 6276-6286.

Van Etten, R. A. (2001). Retroviral transduction models of Ph+ leukemia: advantages and limitations for modeling human hematological malignancies in mice. *Blood Cells Mol Dis,* **27(1):** 201-205.

Van Limbergen, H., Poppe, B., Michaux, L., Herens, C., Brown, J., Noens, L., Berneman, Z., De Bock, R., De Paepe, A. and Speleman, F. (2002). Identification of cytogenetic subclasses and recurring chromosomal aberrations in AML and MDS with complex karyotypes using M-FISH. *Genes Chromosomes Cancer,* **33(1):** 60-72.

Varga, A. E., Dobrovic, A., Webb, G. C. and Hutchinson, R. (2001). Clustering of 1p36 breakpoints distal to 1p36.2 in hematological malignancies. *Cancer Genet Cytogenet,* **125(1):** 78-79.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., *et al.* (2001). The sequence of the human genome. *Science,* **291(5507):** 1304-1351.

Visvader, J. and Adams, J. M. (1993). Megakaryocytic differentiation induced in 416B myeloid cells by GATA-2 and GATA-3 transgenes or 5-azacytidine is tightly coupled to GATA-1 expression. *Blood,* **82(5):** 1493-1501.

Wall, L., Deboer, E. and Grosveld, F. (1988). The human beta-globin gene 3' enhancer contains multiple binding sites for an erythroid-specific protein. *Genes Dev,* **2(9):** 1089-1100.

Wang, L. C., Swat, W., Fujiwara, Y., Davidson, L., Visvader, J., Kuo, F., Alt, F. W., Gilliland, D. G., Golub, T. R. and Orkin, S. H. (1998). The TEL/ETV6 gene is required specifically for hematopoiesis in the bone marrow. *Genes Dev,* **12(15):** 2392-2402.

Weiss, M. C. and Green, H. (1967). Human-mouse hybrid cell lines containing partial complements of human chromosomes and functioning human genes. *Proc Natl Acad Sci U S A,* **58(3):** 1104-1111.

Weiss, M. J. and Orkin, S. H. (1995). GATA transcription factors: key regulators of hematopoiesis. *Exp Hematol,* **23(2):** 99-107.

Weiss, M. J., Yu, C. and Orkin, S. H. (1997). Erythroid-cell-specific properties of transcription factor GATA-1 revealed by phenotypic rescue of a gene-targeted cell line. *Mol Cell Biol,* **17(3):** 1642-1651.

Welborn, J. L., Lewis, J. P., Jenks, H. and Walling, P. (1987). Diagnostic and prognostic significance of t(1;3)(p36;q21) in the disorders of hematopoiesis. *Cancer Genet Cytogenet,* **28(2):** 277-285.

Wherlock, M. and Mellor, H. (2002). The Rho GTPase family: a Racs to Wrchs story. *J Cell Sci,* **115(Pt 2):** 239-240.

Wieser, R. (2002). Rearrangements of chromosome band 3q21 in myeloid leukemia. *Leuk Lymphoma,* **43(1):** 59-65.

Wimmer, K., Vinatzer, U., Zwirn, P., Fonatsch, C. and Wieser, R. (1998). Comparative expression analysis of the antagonistic transcription factors EVI1 and MDS1-EVI1 in murine tissues and during in vitro hematopoietic differentiation. *Biochem Biophys Res Commun,* **252(3):** 691-696.

Yagasaki, F., Jinnai, I., Yoshida, S., Yokoyama, Y., Matsuda, A., Kusumoto, S., Kobayashi, H., Terasaki, H., Ohyashiki, K., Asou, N., Murohashi, I., Bessho, M. and Hirashima, K. (1999). Fusion of TEL/ETV6 to a novel ACS2 in myelodysplastic syndrome and acute myelogenous leukemia with t(5;12)(q31;p13). *Genes Chromosomes Cancer,* **26(3):** 192-202.

Yagasaki, F., Wakao, D., Yokoyama, Y., Uchida, Y., Murohashi, I., Kayano, H., Taniwaki, M., Matsuda, A. and Bessho, M. (2001). Fusion of ETV6 to fibroblast growth factor receptor 3 in peripheral T-cell lymphoma with a t(4;12)(p16;p13) chromosomal translocation. *Cancer Res,* **61(23):** 8371-8374.

Yamada, K., Sugimoto, E., Amano, M., Imamura, Y., Kubota, T. and Matsumoto, M. (1983). Two cases of acute promyelocytic leukemia with variant translocations: the importance of chromosome No. 17 abnormality. *Cancer Genet Cytogenet,* **9(2):** 93-99.

Yamaguchi, Y., Zon, L. I., Ackerman, S. J., Yamamoto, M. and Suda, T. (1998). Forced GATA-1 expression in the murine myeloid cell line M1: induction of c-Mpl expression and megakaryocytic/erythroid differentiation. *Blood,* **91(2):** 450-457.

Yehuda, O., Abeliovich, D., Ben-Neriah, S., Sverdlin, I., Cohen, R., Varadi, G., Orr, R., Ashkenazi, Y. J., Heyd, J., Lugassy, G. and Ben Yehuda, D. (1999). Clinical implications of fluorescence in situ hybridization analysis in 13 chronic myeloid leukemia cases: Ph-negative and variant Ph-positive. *Cancer Genet Cytogenet,* **114(2):** 100-107.

Yeyati, P. L., Shaknovich, R., Boterashvili, S., Li, J., Ball, H. J., Waxman, S., Nason-Burchenal, K., Dmitrovsky, E., Zelent, A. and Licht, J. D. (1999). Leukemia translocation protein PLZF inhibits cell growth and expression of cyclin A. *Oncogene,* **18(4):** 925-934.

Yu, V. P., Koehler, M., Steinlein, C., Schmid, M., Hanakahi, L. A., Van Gool, A. J., West, S. C. and Venkitaraman, A. R. (2000). Gross chromosomal rearrangements and genetic exchange between nonhomologous chromosomes following BRCA2 inactivation. *Genes Dev,* **14(11):** 1400-1406.

Yuan, Y., Zhou, L., Miyamoto, T., Iwasaki, H., Harakawa, N., Hetherington, C. J., Burel, S. A., Lagasse, E., Weissman, I. L., Akashi, K. and Zhang, D. E. (2001). AML1-ETO expression is directly involved in the development of acute myeloid leukemia in the presence of additional mutations. *Proc Natl Acad Sci U S A,* **98(18):** 10398-10403.

Zhou, Y., Lim, K. C., Onodera, K., Takahashi, S., Ohta, J., Minegishi, N., Tsai, F. Y., Orkin, S. H., Yamamoto, M. and Engel, J. D. (1998). Rescue of the embryonic lethal hematopoietic defect reveals a critical role for GATA-2 in urogenital development. *Embo J,* **17(22):** 6689-6700.